Université de Montréal

Study of Genetic Factors in Treatment-Related Complications in Patients with Childhood Acute Lymphoblastic Leukemia and Post Transplantation of Hematopoietic Stem Cells

Par

Kateryna Petrykey

Département de Pharmacologie et de Physiologie
Faculté de médecine

Thèse présentée en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.) en Pharmacologie,
Option Pharmacogénomique

Décembre 2021

Université de Montréal
Département de Pharmacologie et de Physiologie, Faculté de médicine

Cette thèse intitulée

Study of Genetic Factors in Treatment-Related Complications in Patients with Childhood Acute Lymphoblastic Leukemia and Post Transplantation of Hematopoietic Stem Cells

Présentée par
Kateryna Petrykey

A été évaluée par un jury composé des personnes suivantes:

Dr. Noël J-M Raynal
Président-rapporteur

Dre. Maja Krajinovic
Directrice de recherche

Dr. Denis DeBlois
Membre du jury

Dre. Francine Durocher
Examinatrice externe

Dr. Réné Cardinal
Représentant du doyen

SOMMAIRE

La leucémie lymphoblastique aiguë (LLA) est le cancer le plus fréquent chez les enfants. Malgré le fait que plus de 80% des enfants atteints de LLA sont aujourd'hui guéris de leur maladie, ce succès a toutefois un prix élevé, car l'exposition aux médicaments cytotoxique et/ou à l'irradiation pendant une période vulnérable du développement de l'enfant peut entraîner des conséquences à long terme. En effet, environ 60% des enfants ayant survécu à une LLA devront vivre avec des problèmes de santé liés au traitement, également appelés effets indésirables tardifs (late-adverse effects, LAEs). Parmi ces derniers, on notera des problèmes métaboliques, l'ostéoporose, une altération des fonctions cognitives ou cardiaques, ainsi que la dépression et l'anxiété. Si certains survivants ne présentent aucune de ces complications, d'autres peuvent en avoir plusieurs. Différents facteurs peuvent contribuer à cette variabilité, notamment le traitement reçu, les caractéristiques de la maladie, les habitudes de vie et, surtout, la constitution génétique du patient. Ce projet s'est concentré sur les biomarqueurs génétiques permettant d'identifier les individus les plus susceptibles de souffrir de LAEs. Récemment, une étude exhaustive (évaluations cliniques, psychosociales et biochimiques) s'est déroulée au CHU Sainte-Justine pour caractériser chacune de ces morbidités chez 250 survivants de la LLA de l'enfant (cohorte PETALE). De plus, on a obtenu le profil génétique de chaque participant. Nous avons utilisé cet ensemble de données et des outils statistiques et bio-informatiques pour réaliser des études d'association comparant la fréquence des variants génétiques chez les survivants ayant développé ou non des LAEs ; en particulier, les complications cardiovasculaires et neurocognitives, ainsi que les troubles de l'humeur tels que l'anxiété et la dépression. D'autres facteurs de risque tels que les caractéristiques de traitement et/ou de la leucémie ont été pris en compte lors de l'analyse pour dériver les meilleurs prédicteurs génétiques.

Ainsi, en utilisant l'approche des gènes candidats, nous avons identifié les variants communs des gènes *MTR, PPARA, ABCC3, CALML5, CACNB2* et *PCDHB10* qui étaient associés à des déficits de performance des tests neurocognitifs, tandis que les variants des gènes *SLCO1B1* et *EPHA5* étaient associés à l'anxiété et à la dépression. Deux variants, rs1805087 dans le gène *MTR* et rs58225473 dans le gène *CACNB2* sont particulièrement intéressants, car ces associations ont été validées dans la cohorte de réplication SJLIFE (St. Jude Children's Research Hospital, Memphis, USA).

Les analyses d'association ont été complémentées par une étude d'association à l'échelle de l'exome, qui a identifié plusieurs gènes supplémentaires comme des modulateurs potentiels du risque de développer des complications neurocognitives liées au traitement (gènes *AK8* et *ZNF382*), ainsi que l'anxiété et la dépression (gènes *PTPRZ1, MUC16, TNRC6C-AS1, APOL2, C6orf165, EXO5, CYP2W1* et *PCMTD1*). Le variant rs61732180 du gène *ZNF382* a ensuite été validé dans la cohorte de réplication SJLIFE.

Également, nous avons effectué des analyses d'association concernant les complications cardiaques liées au traitement qui ont identifié plusieurs nouveaux marqueurs associés à ces complications dans les gènes *TTN, NOS1, ABCG2, CBR1, ABCC5*, *AKR1C3*, *NOD2* et *ZNF267*.

De plus, nous avons résumé les connaissances actuelles sur les marqueurs pharmacogénomiques qui ont été associés aux effets de cardiotoxicités, induites par les anthracyclines, qui affectent les patients atteints de cancer pédiatrique. Nous avons également inclus un aperçu de l'applicabilité des résultats rapportés, notamment ceux qui ont été validés dans la cohorte PETALE.

Par ailleurs, nous nous sommes intéressés aux complications qui surviennent après une greffe de cellules souches hématopoïétiques. Nous avons appliqué des approches bio-informatiques et statistiques similaires pour obtenir un profil plus complet de la composante génétique derrière ces

complications potentiellement mortelles. Ainsi, une étude d'association à l'échelle de l'exome a été réalisée dans une cohorte de patients pédiatriques subissant une greffe de cellules souches hématopoïétiques après un régime de conditionnement contenant du busulfan. Nous avons identifié de nouvelles variations génétiques conférant un risque plus élevé de syndrome d'obstruction sinusoïdale (notamment dans les gènes *UGT2B10, BHLHE22,* et KIAA1715) et de maladie aiguë du greffon contre l'hôte (dans les gènes *ERC1, PLEK, NOP9* et *SPRED1*), qui pourraient être utiles pour des stratégies personnalisées de prévention et de traitement.

Ces travaux contribuent à la compréhension de l'influence des facteurs génétiques sur le risque de développer des complications liées au traitement, tant au cours du traitement qu'à long terme. De plus, les marqueurs génétiques signalés ainsi que d'autres facteurs de risque connus peuvent conduire à des modèles de prédiction identifiant les patients à risque accru de ces complications.

Mots-clés

Leucémie lymphoblastique aiguë de l'enfant, survivants du cancer, cancer infantile, effets indésirables tardifs, complications neurocognitives, anxiété, dépression, performances cognitives, troubles de l'humeur, cardiotoxicité induite par les anthracyclines, doxorubicine, syndrome d'obstruction sinusoïdale hépatique, maladie aiguë du greffon contre l'hôte, greffe de cellules souches hématopoïétiques, busulfan, facteurs génétiques, étude d'association, séquençage de l'exome entier.

ABSTRACT

Acute lymphoblastic leukemia (ALL) is the most common cancer in children. Even though more than 80% of children with ALL are now cured of their disease, this success comes at a high price as exposure to cytotoxic drugs and/or radiation during a vulnerable period of child development may have long-term consequences. In fact, approximately 60% of children who survive ALL will have to live with treatment-related health problems, also called late-adverse effects (LAEs). These include metabolic problems, osteoporosis, impaired cardiac or cognitive functions, as well as depression and anxiety. While some survivors do not have any of these complications, others may have more than one.

Different factors can contribute to this variability, in particular, the treatment received, the characteristics of the disease, the lifestyle, and, above all, the genetic makeup of the patient.

This project focused on genetic biomarkers capable of identifying the individuals most likely to suffer from LAEs. Recently, an exhaustive study (clinical, psychosocial, and biochemical evaluations) took place at Sainte-Justine University Health Center (Montreal, Canada), with the goal to characterize each of these morbidities in 250 survivors of childhood ALL (PETALE cohort). In addition, the genetic profile of each participant was obtained, and we used statistical and bioinformatics tools to perform association studies on this dataset in order to compare the frequency of genetic variants in survivors with or without LAEs. We evaluated cardiovascular and neurocognitive complications, as well as mood disorders such as anxiety and depression. Other risk factors, such as treatment and/or leukemia characteristics were also considered during the analysis to derive the best genetic predictors.

Thus, using the candidate gene approach, we identified common variants in the *MTR, PPARA, ABCC3, CALML5, CACNB2,* and *PCDHB10* genes that were associated with deficits in neurocognitive tests performance, whereas variants in the *SLCO1B1* and *EPHA5* genes were associated with anxiety and depression. Two variants, rs1805087 in the *MTR* gene and rs58225473 in the *CACNB2* gene, are of particular interest since these associations were validated in an independent SJLIFE replication cohort (St. Jude Children's Research Hospital, Memphis, USA).

The association analyses were complemented by an exome-wide association study, which identified several additional genes as potential modulators of the risk of developing treatment-related neurocognitive complications (genes *AK8* and *ZNF382*), as well as anxiety and depression (genes *PTPRZ1, MUC16, TNRC6C-AS1, APOL2, C6orf165, EXO5, CYP2W1,* and *PCMTD1*). Variant rs61732180 in the *ZNF382* gene was further validated in the replication SJLIFE cohort.

To a great extent, we performed association analyses regarding treatment-related cardiac complications which identified several novel markers associated with these toxicities in the *TTN, NOS1, ABCG2, CBR1, ABCC5, AKR1C3*, *NOD2,* and *ZNF267* genes in survivors of childhood ALL.

In addition, we summarized the current knowledge on pharmacogenomic markers related to anthracycline-induced cardiotoxicity affecting pediatric cancer patients. We also included a brief overview of the applicability of reported findings to the PETALE cohort, validating several of them.

Besides, we were interested in the complications that arise after a hematopoietic stem cell transplantation. We applied similar bioinformatics and statistical approaches to gain a more complete insight into the genetic component behind these life-threatening complications. Thus, an exome-wide association study was performed in a cohort of pediatric patients undergoing hematopoietic stem cell transplantation following a conditioning regimen containing busulfan. Our results identified new genetic variations conferring a higher risk of sinusoidal obstruction syndrome (notably in the *UGT2B10, BHLHE22*, and *KIAA1715* genes) and acute graft-versus-host disease (*ERC1, PLEK, NOP9,* and *SPRED1* genes), which could be useful for personalized prevention and treatment strategies.

This work contributes to the understanding of the influence of genetic factors on the risk of developing treatment-related complications, both during treatment and in the long term. Furthermore, the reported genetic markers along with other known risk factors can lead to prediction models identifying patients at increased risk for these complications.

Keywords

Childhood acute lymphoblastic leukemia, cancer survivors, childhood cancer, late adverse effects, neurocognitive complications, anxiety, depression, cognitive performance, mood disorders, anthracycline-induced cardiotoxicity, doxorubicin, sinusoidal obstruction syndrome, acute graft versus host disease, hematopoietic stem cell transplantation, busulfan, genetic factors, exome-wide association study, whole-exome sequencing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ABC: Adenosine triphosphate-binding cassette subfamily C
ABCC11*: ATP Binding Cassette Subfamily C Member 11*
ABCC3*: ATP Binding Cassette Subfamily C Member 3*
ABCC5*: ATP Binding Cassette Subfamily C Member 5*
ABCG2*: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group)*
ABL*: Tyrosine-protein kinase ABL1*
ACT: anthracycline-induced cardiotoxicity
AD: Alzheimer's disease
ADH: Alcohol dehydrogenase
ADH7*: Alcohol Dehydrogenase 7*
ADHD: attention deficit / hyperactivity disorder
ADR: adverse drug reaction
AGPAT3**:** 1-acylglycerol-3-phosphate O-acyltransferase 3
AK8*: Adenylate Kinase 8 ATP-AMP Transphosphorylase 8*
AKR1C3*: Aldo-Keto Reductase Family 1 Member C3;*
ALL: Acute Lymphoblastic Leukemia
AMPH: amphiphysin
APOL2*: Apolipoprotein L2*
ARL16*: ADP Ribosylation Factor Like GTPase 16*
ATP: adenosine triphosphate
ATP2B1*: ATPase Plasma Membrane Ca2+ Transporting 1*
BCR*: Breakpoint Cluster Region Protein*
bHLH: basic helix-loop-helix
BHLHE22: basic helix-loop-helix family member e22
BSI-18: Brief Symptom Inventory-18
Bu: Busulfan
BYI: Beck Youth Inventory
C6orf165*: Cilia And Flagella Associated Protein 206*
CACNB2*: Calcium Voltage-Gated Channel Auxiliary Subunit Beta2*
CALML5*: Calmodulin Like 5*
CASR: *calcium sensing receptor*
CAT : catalase
CBR1*: Carbonyl Reductase 1;*
CBRs: cytosolic carbonyl reductases
CCL8: *C-C motif chemokine ligand 8; PLEK, pleckstrin;*
CCSS: Childhood Cancer Survivors Study
CELF4*: CUGBP Elav-Like Family Member 4*
CG: candidate genes

CHU: Centre Hospitalier Universitaire
CI: confidence interval
CLPs: common lymphoid progenitors
CMPs: common myeloid progenitors
CNS: Central Nervous System
CPIC: Clinical Pharmacogenetics Implementation Consortium
CPNDS: Canadian Pharmacogenomics Network for Drug Safety
CRT: Cranial radiation therapy
CS: corticosteroids
CSF: Cerebrospinal Fluid
cumAUC: cumulative area under the curve
Cy: Cyclophosphamide
CYP: Cytochrome P450
DDR: DNA damage response
DFCI: Dana-Faber Cancer Institute
D-KEFS: Delis-Kaplan Executive Function System
DNA: deoxyribonucleic acid
DPWG: Dutch Pharmacogenetics Working Group
DUOX2*: Dual Oxidase 2; ZNF: Zinc Finger Protein 382*
DUOXA1*: Dual Oxidase Maturation Factor 1*
EBMT: European Society for Blood and Marrow Transplantation
EETs: epoxyeicosatrienoic acids
EPHA5*: EPH Receptor A5, Brain-Specific Kinase*
eQTL: expression quantitative trait locus
ER: endoplasmic reticulum
ERC1*: ELKS/RAB6-interacting/CAST family member 1*
EryPs: erythroid progenitors
FAT3**:** FAT atypical cadherin 3
FDR: False discovery rate
FDR-BH: Benjamini–Hochberg false discovery rate
FISH: Fluorescence In Situ Hybridization
FMO*: flavin-containing monooxygenases*
GMPs: granulocyte-monocyte progenitors
GPR35*: G Protein-Coupled Receptor 3*
GST: Glutathione S-transferase
GSTT1: *Glutathione S-Transferase Theta 1*
GTEx: Genotype-Tissue Expression
GWAS: Genome-wide association studies
HADH: hydroxyacyl-CoA dehydrogenase
HAS3*: Hyaluronic Acid Synthase 3*
HF: heart failure

***HFE**: hemochromatosis gene*
HIF: hypoxia induced factors
***HNMT**: Histamine N-methyltransferase*
HR: hazard ratio
HR: high risk
HSCs: hematopoietic stem cells
HSCT: Hematopoietic Stem Cell Transplantation
HSCT: Hematopoietic Stem Cell Transplantation
***HSPA4**: Heat Shock Protein Family A (Hsp70) Member 4*
***ISG20:** interferon stimulated exonuclease gene 20;*
***KIAA1715 (LNPK)*:** lunapark (endoplasmic reticulum junction formation factor)
***KIR3DL1**: Killer Cell Immunoglobulin Like Receptor*
KM: Kaplan–Meier
LAEs: Late adverse effects
LD: linkage disequilibrium
LLA: leucémie lymphoblastique aiguë
LP: Lumbar Puncture
LSCs: leukemic stem cells
LV: left ventricle

LVEF 2D (Simpson): Left ventricular ejection fraction method of discs biplane (decimal)

LVEF M-MODE: Left Ventricular Ejection Fraction (M-mode; Teichholz)
LVFS M-MODE: Left ventricular fractional shortening (M-mode, decimal)
MAF: Minor allele frequency
Mel: Melphalan
MEPs: megakaryocyte-erythroid progenitors
***MICB**: MHC Class I Polypeptide-Related Sequence B*
MKs: megakaryocytes
***MLL* (*KMT2A*):** *Lysine methyltransferase 2A*
MPPs: multipotent progenitors
MRD: Minimal Residual Disease
mRNA: messenger ribonucleic acid
***MTR**: 5-Methyltetrahydrofolate-Homocysteine Methyltransferase*
MTX: methotrexate
***MUC16**: Mucin 16, Cell Surface Associated*
***NOD2**: Nucleotide Binding Oligomerization Domain Containing 2*
***NOP9**: NOP9 nucleolar protein;*
NOS: Nitric oxide synthases
***NOS1**: Nitric Oxide Synthase 1 (Neuronal);*
OATP: organic anion transporting polypeptide
OR: odds ratio

PETALE: **P**révenir les **E**ffets **TA**rdifs des traitements de la **LE**ucémie lymphoblastique aigüe

PGx: Pharmacogenomics

PLAUR*: Urokinase Plasminogen Activator Surface Receptor*

PLCE1*: Phospholipase C Epsilon 1*

PLINK: open-source genome-wide association analysis toolkit

PolyPhen-2: **Poly**morphism **Phen**otyping **v2**

PPARA*: Peroxisome Proliferator Activated Receptor Alpha*

PPARG*: Peroxisome Proliferator Activated Receptor Gamma*

PROMIS: Patient-Reported Outcomes Measurement Information System

PTP: mitochondrial permeability transition pore

PTPRZ1*: Protein Tyrosine Phosphatase Receptor Type Z1*

RARG: Retinoic Acid Receptor Gamma

RBCs: red blood cells

RNPGx: French National Network (Réseau) of Pharmacogenetics (RNPGx)

ROS: reactive oxygen species

SIFT: **S**orting **I**ntolerant **F**rom **T**olerant

SJLIFE: St-Jude Lifetime cohort

SJUHC: Sainte-Justine University Health Center

SKAT-O test: Optimal Sequence Kernel Association Test

SLC: solute carriers

SLC10A2*: Solute Carrier Family 10 Member 2*

SLCO1B1*: Solute Carrier Organic Anion Transporter Family Member 1B1*

SLCO2B1*: Solute Carrier Organic Anion Transporter Family Member 2B1*

SNP: Single nucleotide polymorphism

SNV: Single nucleotide variation

SNV: single-nucleotide variations

SOS: Sinusoidal Obstruction Syndrome

SPG7*: Spastic Paraplegia 7 Protein*

SPRED1*: sprouty-related, EVH1 domain-containing protein 1*

SR: standard risk

STXBP3: *syntaxin binding protein 3;*

SULT: Sulfotransferase

TAGAP*: T cell activation RhoGTPase activating protein*

TNRC6C-AS1*: TNRC6C antisense RNA 1*

TTN: *Titin*

UGT: UDP-glucuronosyltransferase

UGT2B10*: UDP glucuronosyltransferase family 2 member B10*

UPR: unfolded protein response

VEGFA: vascular endothelial growth factor A

VHR: Very High Risk

VOD: Veno-Occlusive Disease
WAIS-IV: Wechsler Adult Intelligence Scale-Fourth Edition
WBC: White Blood Cell
WES: Whole exome sequencing
ZNF267*: Zinc Finger Protein 267*
ZNF608: *Zinc Finger Protein 608*

DEDICATION

I dedicate this thesis to my mother, Anna.
For her endless love, support, and encouragement.

ACKNOWLEDGEMENTS

It is impossible to adequately express the depth of my gratitude to all who have assisted me in my research and made a huge contribution in so many ways to my PhD.

First and foremost, I would like to thank my supervisor, **Dr. Maja Krajinovic**, for her invaluable guidance. As a brilliant and inspiring mentor over the past few years, Maja pushed me to achieve my best, but has also been an endless source of encouragement, patience, and support. She has offered me many unique opportunities including the opportunity to work on outstanding projects and take on teaching responsibilities, for which I will remain forever grateful. I will continue to strive to be as ambitious, dedicated, and passionate as Maja throughout my future professional journey and will always cherish my experience working in her lab.

I would also like to express my deepest gratitude to the members of my thesis committee, **Dr. Francine Durocher**, **Dr. Denis DeBlois**, and **Dr. Noël J-M Raynal**, for accepting their roles and taking the time to go through my work.

I would like to deeply thank **Dr. Réné Cardinal** for his guidance, encouragement, and extremely valuable recommendations. I also want to express my gratitude to the staff of the **Department of Pharmacology and Physiology** for their dedicated work and concern for the student's well-being.

I'm particularly grateful to all the collaborators I had the privilege to work with: **Dr. Daniel Sinnett**, **Dr. Gregor U. Andelfinger**, **Dr. Caroline Laverdière**, **Dr. Nathalie Alos**, **Dr. Philippe Robaey**, **Dr. Marc Ansari**, and **Dr. Tiago Nava**, and the others who were not named here, as well

as the collaborators from the **St. Jude Research Hospital**. I would especially like to thank **Drs Sarah Lippé** and **Serge Sultan** for their contributions, and for presenting explanations of complex ideas with the utmost clarity and patience, for their encouragement and support.

I would also like to acknowledge all the patients who accepted to participate in the studies and made this work possible, as well as the **Université de Montréal** and the **Cole Foundation** for providing the financial support.

The whole PhD process has definitely been made easier thanks to the atmosphere of support, collaboration and friendship fostered in Dr. Krajinovic's lab. I would like to thank all previous and current members of the lab, in particular **Vincent Gagné** and **Aziz Rezgui** for their friendly encouragement and valuable professional advice. My special thanks go to **Rachid Abaji**, with whom we first shared a common path of doctoral studies and then became close friends. He and his wonderful family have always been there for me, and their support has been so precious, especially during the most difficult last months of writing this dissertation.

Last but certainly not least, I would like to thank my amazing family, my dear **parents** and **sister**, for their belief in me, their endless love and support. My beloved daughter **Victoria** and husband **Sergey** have gone through various phases of the PhD journey with me. Their love, encouragement, and patience, as well as the chocolate supplies were priceless. All of you have been my best cheerleaders. Thank you for making me see this adventure through to the end!

PREFACE

The present thesis entitled "**Étude de facteurs génétiques dans les complications du traitement chez les patients atteints de la leucémie lymphoblastique aigue de l'enfant et post transplantation de cellules souches hématopoïétiques**" has been carried out by me under the guidance and supervision of Dr. Maja Krajinovic, and is submitted to the faculty of Medicine at the Université de Montréal in partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) in Pharmacology (Pharmacogenomics option). This work is presented in the by-article format.

During my five years of doctoral studies, I had the opportunity to participate in several projects of pharmacogenomics research, discover and validate new genetic markers, review, and summarize currently available data, as well as to apply my extensive prior experience in programming and undertake teaching responsibilities. I had a chance to not only deepen my knowledge in genetics and pharmacology, but also to further develop my observational and analytical skills.

In the body of this thesis, in the first two chapters of Section A, I will provide a brief introduction covering the basic information necessary for the understanding of the context of this work and describe the developed pipeline used for the genetic association studies. The following chapters of this dissertation (Sections B and C) represent the original research papers (except for Chapter 6 where the original research was combined with a literature review).

While I feel great pride in the totality of the work presented in this thesis, I remain bound by a professional responsibility to be transparent about the extent of my contribution to the different projects. For example, in the Busulfan project (described in Section C, Chapters 7 and 8), my part could be considered relatively modest compared to the other parts of the thesis, but it had

nevertheless an important role in the advancement and the fruitful completion of the work, owing to my participation in the conception, assembly, and implementation of the pipeline used. On the other hand, my contribution to the remaining sections of the thesis (work in the context of the PETALE project, a multidisciplinary research project with the goal to identify and to characterize predictive biomarkers associated with treatment-related late adverse effects in childhood ALL survivors) could be viewed as more central and significant in comparison with the other authors, taking into account that the unfolding of the different outputs from this project was directly and predominantly guided by my endeavors.

Section A

Chapter 1

General Introduction

This chapter contains important information related to the various topics covered in this work in order to make it easier for readers to understand the following chapters.

1. General Introduction

1.1. Pharmacogenomics

Advances in clinical research and the implementation of high throughput technologies (e.g., next-generation sequencing, etc.) significantly increased the clinical use of genomic information. With the recent improvements in the ability to sequence and analyze large amounts of genomic data, paired with continuous optimization of computational and bioinformatic tools, it became possible to obtain an unbiased view of the entire human genome. This allowed for a comprehensive interrogation of the genetic variations involved in human health and disease, and for the discovery of variants predictive of treatment responsiveness[1,2].

An important part of the diversity of treatment response between individuals can be explained by factors such as age, sex, body mass index, comorbidities, and lifestyle. However, it is also well understood that the genetic constitution of individuals can considerably contribute to this variability and could be an important cause of treatment resistance and/or serious adverse effects in certain patients[3-5]. Pharmacogenomics (PGx) explores the influence of the genetic component on the observed inter-individual differences in response to treatment; intending to identify the genetic predictors of variability in drug efficacy and toxicity[3,5,6]. More specifically, PGx focuses on the role of genetics in modulating the activity of metabolizing enzymes, drug transporters, and molecular targets, as well as their potential impact on specific drug-related phenotypes. These identified genetic markers could also indicate novel drug targets and/or modifiers that may affect treatment outcome[4,5,7]. Characterization of genetic factors that may predispose a patient to an adverse drug reaction (ADR) could be essential in preventing such reactions knowing that ADRs represent one of the leading causes of death[8].

For instance, cytochrome P450 (CYP) isoenzyme 2D6 (*CYP2D6*) is highly implicated in drug metabolism with an estimation that around 25% of all existing drugs are transformed, at least in

part, by the latter[9]. Patients with multiple copies of the *CYP2D6* gene may therefore not achieve therapeutic plasma levels at the usual drug dose as they will eliminate the drug from their system too quickly (rapid metabolizers); whereas, individuals with only few functional *CYP2D6* copies may be poor metabolizers, causing drug levels to exceed the therapeutic range[9,10].

An important example of a PGx finding that is making its way from bench to bedside is that of the topoisomerase I inhibitor *Irinotecan*. This drug exerts antitumor effect in a variety of malignancies, notably against metastatic colorectal cancer, for which, limited treatment options existed before its appearance on market. Clinical pharmacogenomics studies revealed that a fraction of the population, which is at a higher risk of severe toxicity following the administration of standard doses of *Irinotecan*, can be identified prospectively[11]. This was explained by the finding that patients with genetic polymorphisms in the gene encoding uridine diphosphate glucuronosyltransferase (UGT)1A1, have a lower-than-normal capacity to metabolize SN-38, the active metabolite of *Irinotecan*[12,13]. Consequently, the US Food and Drug Administration (FDA) and Japan`s Pharmaceuticals and Medical Devices Agency (PMDA), as well as other regulatory authorities, added pharmacogenomics recommendations to the monography of this drug, suggesting the use of a lower initial dose of *Irinotecan* in patients with *UGT1A1*28/*28*, *UGT1A1*6/*6* or *UGT1A1*6/*28* dysfunctional genotypes[14-16].

Over the past several decades, significant advances have been made in the field of PGx. Indeed, numerous clinical trials to date have demonstrated and validated the additional benefits of PGx-guided therapy over existing standard approaches in a variety of treatment areas, including pain management, cardiology, neurology, oncology, organ transplantation, and immunosuppression[17]. However, the application of PGx approaches seems to vary considerably across different healthcare fields. For example, in cancer therapy, the therapeutic agents are often administered at

high doses, and patients routinely receive multi-drug combinations; which, taken together, and coupled with inter-individual variability and narrow therapeutic indexes of some of them, can lead to a spectrum of outcomes ranging from underexposure to severe toxicities[3,18,19]. In this case, the ability to predict how a particular patient will respond to a specific treatment is highly valuable; especially given the fact that the consequences of treatment failure could be life-threatening.

PGx has a great potential to improve the use of existing medications in order to enhance efficacy and reduce toxicity by allowing for optimal treatment selection and dose personalization based on the genetic characteristics of individuals. However, the interpretation of pharmacogenetic test results also requires appropriate education of medical professionals, as well as clinical practice recommendations and algorithms.

1.1.1. Genetic structural alterations

The human genome contains a large number of genetic variations and these variations take many different forms ranging from large chromosome rearrangements to single nucleotide polymorphisms[20]. Common genetic variations include: single nucleotide substitutions – referred to as single nucleotide polymorphisms (SNPs) in populations (if they are present with a frequency of 1% or more) and/or single-nucleotide variations (SNVs) in individuals (SNV can be a common SNP or a rare mutation); genomic insertions and deletions (which are often called indels); and duplications, collectively known as copy number variations (CNVs); as well as less-frequent inversions and translocations[21]. The most frequent inherited sequence variations are SNPs; of those, the ones with a minor allele frequency of >1% occur every 100 to 300 bp[22,23]. It is now estimated that about 5 to 7 million SNPs exist in humans, and they account for most inherited phenotypes, including disease susceptibility[22]. Interestingly, only a small proportion (approximately 60,000) of the known SNPs is located within the coding regions of genes, and roughly half of them cause amino acid changes in expressed proteins (non-synonymous SNPs)[24]. Amino acid changes have the potential to affect the function of the encoded protein. It is estimated that each human individual carries approximately 250 to 300 loss of function variants in annotated genes, and around 50 to 100 variants previously identified in congenital disorders[25]. SNPs in the same region of DNA can be stably inherited and form haplotypes (usually <50 kb apart from each other). Therefore, the human genome is organized in haplotype blocks with high linkage disequilibrium (LD): regions representing high level of inheritance that are separated by regions of low LD[26]. Accordingly, if a particular SNP, that is in a strong LD with a set of other SNPs is found to be associated with a disease phenotype or a drug-response phenotype, it may not necessarily be the direct and only cause of the phenotype of interest, but can rather indicate the

position in the genome where a susceptibility genetic region is located[6,27]. The vast majority of genes in the human genome contains regulatory regions. The functional significance of genetic variants located at the level of genomic regulatory regions (including those located at the level of the promoter, enhancer, and 3′ regulatory regions) is difficult to determine[28]. SNPs located at the 5′ regulatory region may influence mRNA transcription. Correspondingly, those located at the enhancer regions can alter the expression of several genes. Likewise, variants located in the 3′ untranslated regions can affect the stability of the transcripts[28]. The variants could also alter the existing canonical splicing sites (or generate new sites), typically resulting in frame deletion of an exon and in generation of a shorter protein[28].

Although many polymorphisms don't have functionally significant effects, those that result in either altered expression or activity of the gene product are those that usually represent an interest for pharmacogenomic association studies[4,29]. The frequency of a genotype and the magnitude of the detected effect largely influence the proportion of variability in the outcome (i.e., treatment response) that can be explained by polymorphism(s)[3,30].

The clinical significance of genetic variants reflects their functional/biological effects, in other words, variants that convey large functional effects are likely to convey a clinically distinct phenotype[28]. This can be seen as a gradient. On one end of the spectrum are those variants that exhibit very large effect sizes and, therefore, have high penetrating power[28]. These variants are usually responsible for genetic diseases that follow Mendelian patterns of inheritance, such as hereditary cardiomyopathies and ion channel disorders[28]. Variants with moderate to large effect sizes are responsible for single-gene diseases detected in small families (have incomplete penetrance) and sporadic cases. On the other end of the spectrum are variants that demonstrate low

effect sizes, and these variants are responsible for complex phenotypes such as hypertension, atherosclerosis, type 2 diabetes, etc.[28].

Studies identifying a genetic predisposition to the disease generally attempt to demonstrate the relationship between the genotype (for one or more polymorphic markers) and the phenotype associated with the predisposition to the disease[31]. There are two main approaches: one is based on candidate genes and the other is based on genome-wide association testing[32].

The candidate genes (CG) approach starts with the selection of presumed candidate genes based on their relevance to the studied disease mechanism (trait)[33]. This is followed by the assessment and selection of polymorphisms that have functional consequences, i.e., affect either the regulation of the gene, or its protein product[33]. Then, the variants are tested for association with disease (trait) based on the frequency of their occurrence in affected test subjects (cases) and selected controls[33]. Although CG studies tend to have rather high statistical power, they are incapable of discovering new genes or gene combinations[32].

Genome-wide association studies (GWAS), based on testing the entire genome, represent one of the major advances in biomedical discoveries in recent decades[34]. Since 2005, after the publication of the first GWAS results[35], thousands of robust associations between the genetic component and complex traits or diseases have been identified. However, the clinical applicability of these findings remains unclear, mostly because of the modest effect of associated variants and their unclear functional implications (since GWAS can detect genes regardless of whether their function has been previously known or identified as associated with the studied outcome)[32,34].

Moreover, despite the extensive discovery of trait- and disease-associated common variants, even in diseases for which large GWAS meta-analyses have been undertaken, much of the genetic contribution to complex traits remains unexplained[36]. Therefore, rare variants (with the frequency

lower than 1% in a given population) can explain additional disease risk or trait variability[36]. Furthermore, evolutionary theory predicts that deleterious alleles are likely to be rare due to purging selection[36], and in fact, loss-of-function variants that prevent the formation of functional proteins are particularly rare[37]. However, for the analysis of the association of rare variants, tests for association with one locus in traditional GWAS are underpowered (due to the low frequency of rare variants); except in cases where causal variants have a very large effect size[36]. To increase power, region-based collapsing or binning approaches, which aggregate variants into single genetic units, have become the standard for the rare variants analysis[38]. These methods assess the association of the combined effect of several rare variants with the outcome of interest in a biologically relevant region[39].

The methods used for the analysis of common and rare variants in this work will be discussed in more details in the following sections.

1.1.2. Implementation of pharmacogenomics knowledge into clinical practice

Prevention and treatment strategies that consider individual variability (the main concept behind precision medicine) have been a part of healthcare for many years[40]. For example, blood typing has been used to determine the type of blood transfusion for decades already[40]. However, the prospect of widespread application of this concept has been greatly improved by the recent development of large-scale biological databases (including proteomics, metabolomics, genomics, etc.) and powerful computational approaches to analyze big datasets[40].

The ease of use of the genotyping technologies, as well as their improved costs, made it possible to proactively assess patients' pharmacogenomic information for variations that could affect their response to pharmacotherapy[41]. For instance, in 2007 the Food and Drug Administration (FDA) began including pharmacogenomic information in drug labeling, and since then, approximately 10% of approved drugs contain this information on their labels[42].

Although, PGx has a growing impact on healthcare and constitutes one of the main knowledge resources to support precision medicine, its implementation into clinical practice has been relatively slow[43]. One of the barriers to accepting genetic data to guide drug use is the lack of knowledge on how to translate genetic test results into clinical action based on currently available evidence[43]. On the other hand, given that large randomized controlled trials to demonstrate the usefulness of pharmacogenetic testing are not always possible or considered necessary to establish clinical utility, the selection of appropriate evidence thresholds for routine clinical use of pharmacogenomics data becomes controversial and challenging[43]. Therefore, evidence-based guidelines are essential for implementing pharmacogenomic knowledge in daily clinical practice[44]. Currently, there are at least four committees developing guidelines with recommendations published in English: the Dutch Pharmacogenetics Working Group (DPWG), the Clinical Pharmacogenetics Implementation Consortium (CPIC), the Canadian Pharmacogenomics

Network for Drug Safety (CPNDS), and the French National Network (Réseau) of Pharmacogenetics (RNPGx)[44]. All four committees have similar methodologies for developing recommendations. However, at the beginning of their projects, the goals were different, resulting in the unique characteristics and strengths of their guiding principles[44]. In an effort to overcome the existing disagreements on recommendations, the committees also initiated the joint harmonization projects[44]. The DPWG and CPIC focus on pharmacotherapy guidelines for a large number of drugs in combination with the patient's genotype or predicted phenotype[44]. The DPWG, CPNDS, and RNPGx also recommend genetic testing in routine clinical practice, with RNPGx even describing the specific clinical setting or medical conditions for which genotyping is recommended[44].

For example, these committees provide recommendations on anticancer agents. Both the DPWG and the RNPGx consider genotyping of *UGT1A1* to be necessary prior to initiation of *Irinotecan* treatment. In addition, the RNPGx recommends genotyping for a particular category of patients, those who will receive an increased dose (> 240 mg/m2)[44]. The CPNDS strongly recommends genetic testing for the associated functional *TPMT* variants (*3A, *3B, and *3C) in all patients before starting *Cisplatin* treatment, as well as functionally inactive *TPMT*2* variant in children to prevent cisplatin-induced hearing loss[44,45]. In order to prevent anthracycline-induced cardiotoxicity, the CPNDS recommends genotyping the *RARG* rs2229774, *SLC28A3* rs7853758 and *UGT1A6*4* rs17863783 variants in all pediatric cancer patients who begin *Doxorubicin* or *Daunorubicin* therapy[44,46].

It is also worth mentioning that accurate and systematic quantification of drug response phenotypes (e.g. toxicity, resistance, etc.) as well as the establishment of optimal treatment approaches using genetic information in conjunction with data on non-genetic causes of inter-patient variability in

drug response, could be much more challenging than the detection of genome variability itself [47]. Therefore, resources provided by organizations such as CPIC, DPWG, CPNDS, and RNPGx that use standardized approaches to evaluate the current knowledge and provide clinical guidance are indispensable for the implementation of PGx into routine clinical practice[43]. Accordingly, PGx is an excellent example of how interdisciplinary translational research that includes fundamental biology, pharmacology, medicine, genetics, and bioinformatics can be integrated into healthcare practice.

1.2. Normal hematopoiesis and leukemia

The formed components of blood include different types of cells or cell fragments, each of which has a different morphology and function[48]. These components are produced through a process known as hematopoiesis; during which hematopoietic stem cells (HSCs) proliferate and undergo self-renewal or differentiation into progenitor cells that continue to differentiate into different types of mature blood cells, as well as into a sufficient number of blood cells necessary to maintain homeostasis[48,49]. The ability to self-renew and the ability to differentiate are two main characteristics of HSCs required for normal hematopoiesis[48]. Self-renewal is the process by which stem cells enter the cell cycle in order to divide and give rise to more stem cells, thereby preserving the stem cell pool[50]. While the ability to differentiate (also known as multipotency) allows HSCs to develop into more mature cells with the gradual lineage determination[48].

The hematopoietic hierarchy is well organized and begins with HSCs developing into myeloid and lymphoid progenitor cells. Myeloid progenitor cells continue to develop into erythrocytes, platelets (by megakaryocyte fragmentation), neutrophils, monocytes, basophils, and eosinophils; whereas lymphoid progenitor cells give rise to B-lymphocytes, T-lymphocytes, natural killer (NK) cells, and dendritic cell populations[48].

Normal HSCs reside in a specialized bone marrow (BM) microenvironment, the so-called HSC niche, that critically regulates the survival and function of HSCs[51]. Various cell types, including osteoblasts, perivascular, endothelial, and mesenchymal cells, contribute to the HSC niche[51,52]. Additionally, the BM functions as a primary and secondary lymphoid organ and contains various mature immune cells, including T and B cells, dendritic cells, and macrophages, which also contribute to the HSC niche[51,52]. Signals originating from the HSC niche are required to control the responses of HSCs and progenitor cells that regulate hematopoiesis during homeostasis, after BM stress (e.g., cytotoxic chemotherapy, irradiation) or during infection[51].

Like normal hematopoiesis, leukemia is also organized hierarchically; so that the leukemic stem cells (LSC) are responsible for initiating and maintaining disease and producing more differentiated malignant cells[51]. Despite being genetically abnormal, the LSCs share many characteristics with normal HSCs, including quiescence, multipotency, and self-renewal[51]. The LSCs depend on signals from the BM microenvironment and occupy similar niches as HSCs. However, the BM microenvironment and LSC niche change considerably due to the infiltration of activated leukemia-specific immune cells into BM[51,53]. In addition, many of the immune mechanisms that have evolved to activate emergency hematopoiesis during infection may actually promote the expansion and differentiation of LSCs, contributing to the progression of leukemia[51]. **Figure 1** provides a brief overview of the hierarchical organization of normal hematopoiesis and leukemic transformation of hematopoietic cells.

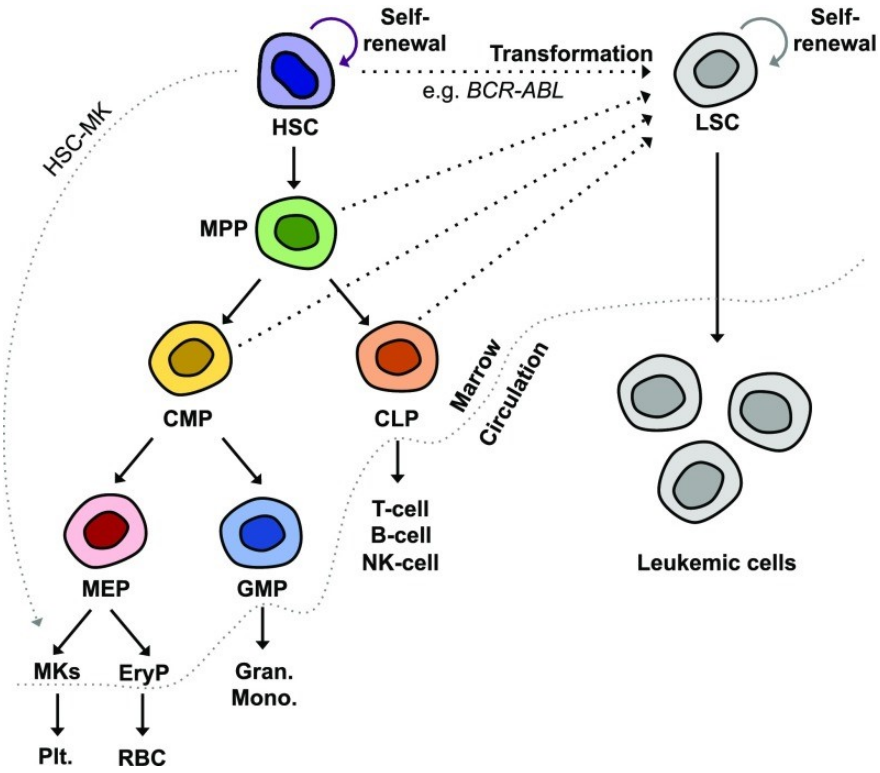


Figure 1. Hierarchical organization of normal hematopoiesis and leukemic transformation.

A conventional model of normal hematopoiesis is shown on the left where different blood lineages are derived from hematopoietic stem cells (HSCs). Leukemia stem cells (LSCs) are derived from the oncogenic transformation of HSCs. However, the transformation of progenitors can also turn them into LSCs depending on oncogenic mutations that define leukemia subtypes. An example is shown where oncogene *BCR-ABL* transforms HSCs but not progenitors to generate LSCs in chronic myeloid leukemia[54].

The *BCR-ABL* oncogene is generated by the Philadelphia chromosome (Ph) translocation, fusing the *BCR* (*Breakpoint Cluster Region Protein*) gene to the *ABL* (*Tyrosine-protein kinase ABL1*) gene.
MPPs: multipotent progenitors, CMPs: common myeloid progenitors, CLPs: common lymphoid progenitors, MEPs: megakaryocyte-erythroid progenitors, GMPs: granulocyte-monocyte progenitors, MKs: megakaryocytes, EryPs: erythroid progenitors, RBCs: red blood cells, platelets (Plt.), granulocytes (gran.), monocytes (mono.).

1.2.1. Acute Lymphoblastic Leukemia (ALL)

Acute lymphoblastic leukemia (ALL) is a rare hematological malignant neoplasm of the bone marrow in which progenitor lymphoblasts, blocked at an early stage of differentiation, rapidly proliferate, and displace normal hematopoietic cells in the bone marrow; hence, provoking a wide range of complications[55,56]. ALL may be of B-cell precursor or T-cell lineage[55].

The incidence of ALL reaches a sharp peak in children aged 1 to 4 and increases gradually in adults, from around age 50[56]. ALL accounts for less than 1% of adult cancers[57], but is the most common pediatric malignant tumor, accounting for around 25% of cancers and 80% of all leukemia in children[55,56].

Several genetic factors (mainly Down syndrome) are associated with an increased risk of ALL, but most patients do not have known hereditary factors[55]. Recent genome-wide association studies have reported polymorphic variants in several genes (including *ARID5B, CEBPE, GATA3,* and *IKZF1*) that are associated with an increased risk of ALL or certain subtypes of ALL; whereas rare germline mutations in the *PAX5* and *ETV6* genes are associated with familial ALL[55]. Chromosomal translocations and intrachromosomal rearrangements are early (and possibly initiating) leukemogenesis events[55]. These translocations and rearrangements are generally present in all leukemic cells and persist during relapses[55].

At the same time, few environmental risk factors were associated with ALL in children. For example, the increased incidence has been associated with exposure to radiation and some chemicals, however these associations represent only a very small number of cases[55].

1.2.2. Criteria for prognostic risk group classification

The application of risk stratification is one of the main characteristics of the treatment of ALL in children; prognostic factors that influence treatment outcome are used to divide patients into groups based on the risk of treatment failure. Thus, patients with favorable characteristics can be treated with less toxic regimens, while more aggressive treatment modalities are intended for patients with a high-risk disease[58].

Criteria for High risk (HR) stratification are mainly attributed based on age, white blood cell count, immunophenotype (presence of T-cell markers) and combination of these factors; as well as central nervous system (CNS) status and Minimal residual disease (MRD) at diagnosis[55,58,59].

Table 1 summarizes the criteria for prognostic risk group classification used for patients diagnosed and treated for childhood ALL according to Dana Farber Cancer Institute (DFCI) ALL protocols at Sainte-Justine University Health Center (SJUHC).

Table 1. Criteria for prognostic risk group classification, PETALE cohort.

-	Standard Risk	High Risk	Very High Risk
Age	**Protocol 87-91**: > 2 year - 8.99 years. **Protocol 95-00-05**: ≥ 1 year - 9.99 years.	**Protocol 91**: <1 year, WBC is based upon highest pre-Rx value obtained at admission. ≥ 1 year - ≤2 years or >9 years. **Protocol 95-00-05**: < 1 years or ≥10.00 to 17.99 years.	**Protocol 87**: WBC> 100,000/mm3 (100 x 10^9cells/L) and/or <1 year.
WBC count, pre-treatment	**Protocol 87-91**: <20,000/mm3 (20 x 10^9 cells/L); **Protocol 95-00-05**: <50,000/mm3 (50 x 10^9 cells/L).	**Protocol 87-91**: ≥20,000/mm3 (20 x 10^9 cells/L); **Protocol 95-00-05:** ≥50,000/mm3 (50 x 10^9 cells/L).	Presence of MLL gene translocations [such as t(4;11)] by karyotype or FISH or molecular analyses.
Diagnostic Lumbar Puncture	No blast cells in the CSF (CNS-1) or < 5 blasts in CSF (CNS-2) on Day 1.	≥ 5 blast cells in CSF (CNS-3) on Day 1.	Presence of hypodiploidy < 45 chromosomes by karyotype or FISH analysis.
CNS	CNS-1 CSF on Day 14 or 18 and at the end-of-induction.	<5 blast cells in CSF (CNS-2) on Day 14 or 18 and at the end-of-induction.	
Cranial Nerve Palsy and Mediastinal mass at Diagnosis	Absence	Presence	
Administration of the corticosteroid pre-treatment	No	Yes	
Immunophenotype	Absence of T-cell markers on lymphoblasts.	Predominance of T-cell markers on lymphoblasts	
MRD	A good response to the first phase of chemotherapy (induction), as measured by a bone marrow test at the end of induction phase with MRD less than 0.001%. And SR patients whose end-of-induction MRD status cannot be determined will still be considered SR.	B-lineage HR patients with MRD level < 0.001 on a marrow sample obtained at end of remission induction therapy will continue to be treated as HR. HR patients whose end-of-induction MRD status cannot be determined will still be considered HR. MRD will not be used to change risk status of T-lineage patients.	

Chromosomal Abnormalities	Absence of t(9;22), *MLL* gene translocations and hypodiploidy < 45 chromosomes by karyotype or FISH analysis.	Patients with t(9;22) will be continued on protocol as HR and a bone marrow donor will be sought for transplantation.	
Testicular at Diagnosis	The leukemia hasn't spread to the testicles.	The leukemia has spread to the testicles.	
Re-Classified from the Risk Group	**The SR patients with CNS-2 CSF on Days 14 or 18 or at the end-of-induction will be re-classified as high risk.**		
	The SR patients with CNS-3 CSF on Day 14 or 18 will be re-classified as high risk and at the end-of-induction will be considered Induction failure.		
	B-lineage SR and HR patients with MRD levels > 0.001 at the end of remission induction therapy will be re-classified as very HR.		

WBC: White Blood Cell; CSF: Cerebrospinal Fluid; CNS: Central Nervous System; MRD: Minimal Residual Disease; FISH: Fluorescence In Situ Hybridization; SR: Standard Risk; HR: High Risk; VHR: Very High Risk; LP: Lumbar Puncture; *MLL* (*KMT2A*): *Lysine methyltransferase 2A.*
CNS-1: WBC in CSF <=> 5 Without Blasts, LP is Negative;
CNS-2: WBC in CSF < 5 With Blasts, LP is Positive;
CNS-3: WBC in CSF ≥ 5 With Blasts, LP is Positive.
If a puncture contains Red Blood Cell: >100 is considered Traumatic; ≤100 is considered Non-Traumatic.

1.2.3. Treatment modalities

Since the early 1960s, remarkable therapeutic improvements have been made for pediatric ALL, which has become the most successful treatment paradigm in pediatric oncology[55]. The clinical success is achieved through improved risk-adapted therapy based on clinical presentation, cytogenetics, and minimal residual disease, as well as on effective central nervous system (CNS) prophylaxis, and enhanced supportive care measures[60]. Due to the modern treatment modalities, the survival rate of children with ALL is currently over 90%[55]. However, ALL is one of the leading causes of cancer-related death, as relapses continue in 15–20% of patients and post-relapse outcomes remain poor[60].

Treatment strategies for ALL have changed significantly over the past two decades[61]. For instance, prophylactic cranial irradiation (for prevention of CNS relapse) was gradually replaced by contemporary ALL protocols, which consist of intensified intravenous and intrathecal administration of chemotherapeutic drugs for standard risk patients[62-64]. Moreover, a recent clinical study demonstrated that with the withdrawal of CRT, ALL chemotherapy-only treatment protocols resulted in an unprecedented overall survival rate of 93.5%[65]. Nevertheless, patients receiving current therapies still exhibit many treatment-related side effects, such as osteonecrosis, cardiovascular and endocrine impairments, and other chronic medical conditions[66,67].

1.2.4. Targeted precision medicine opportunities in ALL

Understanding the genomics of ALL provides a compelling biological rationale for expanding the scope of precision medicine for treating ALL in children[68].

For example, treatment of Philadelphia chromosome (Ph +) positive ALL established the first molecular-targeted paradigm of childhood ALL[68]. Ph + ALL (accounting for approximately 3-5% of childhood ALL cases) is characterized by the canonical translocation t(9;22)(q34;q11), which fuses the *ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase (ABL1)* and *Breakpoint Cluster Region* (*BCR)* genes, in turn creating the BCR-ABL1 oncoprotein. Targeted BCR-ABL1 inhibition with ABL tyrosine kinase inhibitors (in combination with chemotherapy) has significantly improved survival for pediatric Ph + ALL. With this approach, approximately 70% of children can avoid hematopoietic stem cell transplantation (HSCT) in first complete remission[69].

The most exploitable pharmacogenomics targets in ALL are germline variants in the *Thiopurine S-methyltransferase* (*TPMT)* and *Nucleoside Diphosphate-Linked Moiety X Motif 15* (*NUDT15)* genes, which cause myelosuppression with thiopurine therapy[68]. Thiopurines are widely used as anticancer and immunosuppressive agents, but they also have narrow therapeutic indices due to frequent toxicity[70]. The latter is partly explained by the presence of *TPMT* variant alleles associated with low enzymatic activity[70], therefore leading to the intracellular accumulation of thiopurines in haematopoietic tissues[71]. In addition, recent studies have identified a common variant (rs116855232) in the *NUDT15* gene as another important determinant of thiopurine intolerance[72] in patients with inflammatory bowel disease[73] and in children with ALL[74]. *TPMT* and *NUDT15* genotyping are currently recommended at the diagnosis of ALL for individual dose adjustment of mercaptopurine, thereby demonstrating how pharmacogenomics can be implemented in clinical practice as part of the precision medicine approach[68,75].

1.2.5. Late-adverse effects of the treatment

Over the past decades, the survival rate of children and adolescents with cancer has steadily increased, while cancer-related mortality has fallen by more than 50%. Currently, eight out of every ten children and adolescents diagnosed with cancer will live five years or more after being diagnosed. The vast majority of those who reach the five-year survival mark will become long-term survivors[76].

There are approximately 30,000 cancer survivors estimated to be living in Canada today[77], while in the United States, approximately 1 in 750 individuals are childhood or adolescent cancer survivors[76].

The reported frequency of late effects among adults who are childhood cancer survivors ranges from 33% to 75%, depending on the type of cancer, its stage at the time of diagnosis and the type of treatment received[77]. It is estimated that two-thirds of survivors have at least 1 chronic or late-occurring effect from their cancer therapy; up to one-third of these late effects are considered major, serious or life-threatening[77]. Late-adverse effects can include cardiopulmonary, endocrine, renal, or pulmonary dysfunction, neurocognitive impairments and second cancers, among many others[76,77].

Figure 2 summarizes health-related and quality of life outcomes among long-term survivors of childhood and adolescent cancers.

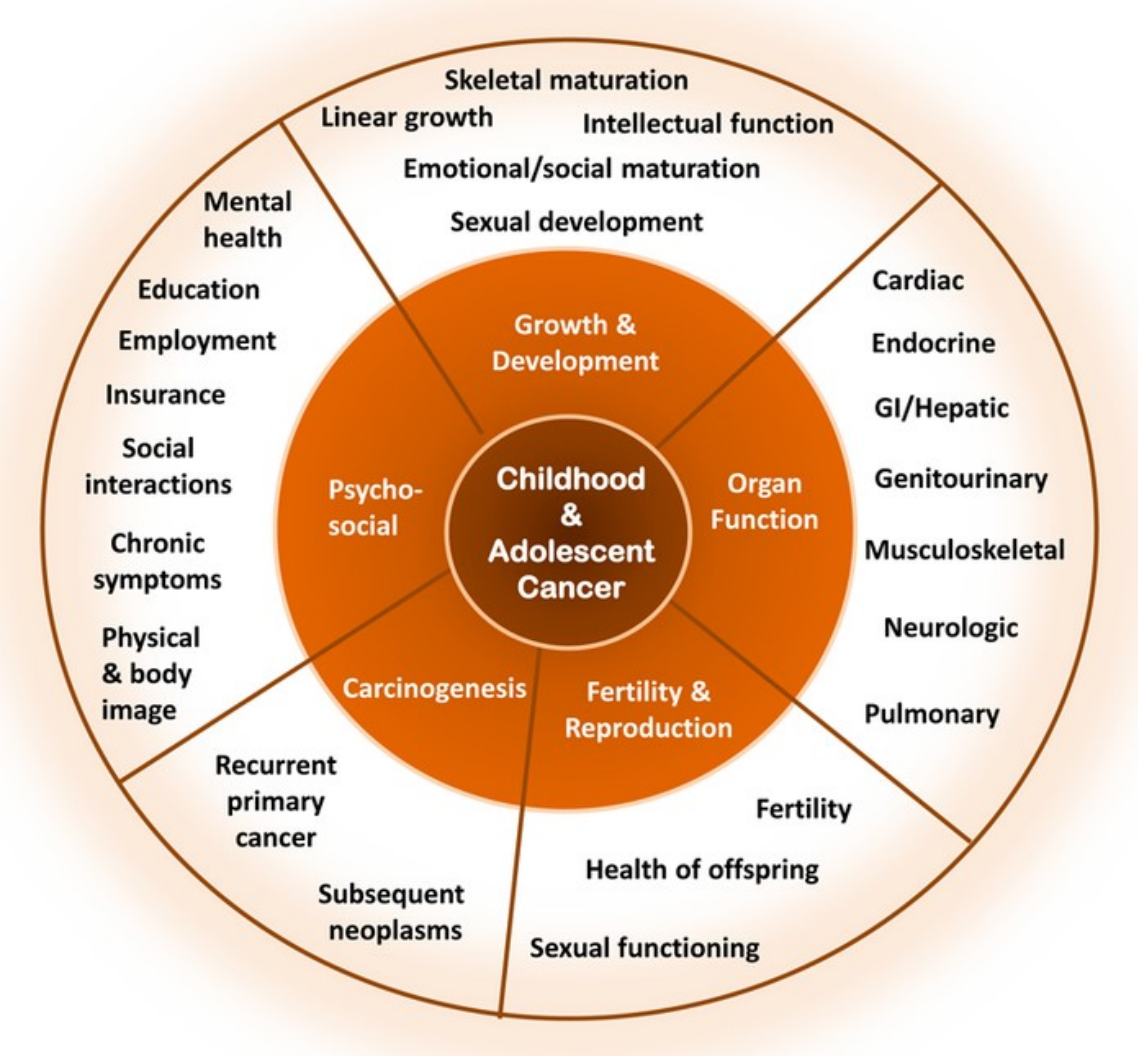


Figure 2 .Spectrum of health-related and quality of life outcomes among long-term survivors of childhood and adolescent cancers.

The long-term health consequences of childhood cancer may be substantial and can affect growth and development, organ function, reproductive capacity, as well as the risk of subsequent carcinogenesis. In addition to the adverse physical and chronic health effects associated with cancer therapy during childhood and adolescences, long-term survivors are at risk of various psychological and social outcomes[76].

Certain classes of chemotherapeutic agents used to successfully treat pediatric cancer patients are also associated with a wide range of potential long-term side effects. These medications include alkylating agents, anthracycline antibiotics, antimetabolites, corticosteroids, and vinca alkaloids[76]. In general, the risk of long-term side effects associated with chemotherapy depends on the cumulative dose, but can also be affected by other factors, such as the route of administration, the sex, and the age of the patient[76]. Selected examples of chemotherapy-related long-term effects are presented in **Table 2**[76].

Table 2. Selected examples of established chemotherapy-associated late effects.

Class of Chemotherapy	Chemotherapeutic Agents	Established Late Effects	References
Alkylating agents	*Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin, Cyclophosphamide, Ifosfamide, Lomustine, Mechlorethamine, Melphalan, Procarbazine, Thiotepa*; plus, the non-classical alkylators *Dacarbazine* and *Temozolomide*	Secondary myelodysplasia or acute myeloid leukemia Gonadal dysfunction and Infertility Pulmonary fibrosis (with exposure to *Busulfan*, *Carmustine* or *Lomustine*) Urinary tract abnormalities (with exposure to *Cyclophosphamide* or *Ifosfamide*)	*Kenney LB et al.*[78]; *Jones DP et al.*[79]; *Ritchey M et al.*[80]; *Huang TT et al.*[81]; *Metzger ML et al.*[82]; *Brock PR et al.*[83]; *Feldman DR et al.*[84]; *Hijiya N et al.*[85]; *Skinner R et al.*[86]

Class of Chemotherapy	Chemotherapeutic Agents	Established Late Effects	References
		Renal dysfunction (with exposure to *Cisplatin/Carboplatin* and *Ifosfamide*) Ototoxicity (with exposure to *Cisplatin* or very high dose *Carboplatin*) Dyslipidemia (with exposure to *Cisplatin*)	
Anthracyclines	*Daunorubicin, Doxorubicin, Epirubicin*, and *Idarubicin*	Left ventricular dysfunction Cardiomyopathy Dysrhythmias	*Van der Pal HJ et al.*[87]*; Mulrooney DA et al.*[88]*; Lipshultz SE et al.*[89]
Corticosteroids	*Dexamethasone, Prednisone*	Reduced bone mineral density Osteonecrosis Cataracts	*Wasilewski-Masker K et al.*[90]*; Kadan-Lottick NS et al.*[91]*; Whelan K et al.*[92]
Vinca Alkaloids	*Vincristine, Vinblastine*	Peripheral sensory and motor neuropathy	*Jain P et al.*[93]*; Ness KK et al.*[94]

Class of Chemotherapy	Chemotherapeutic Agents	Established Late Effects	References
Antimetabolites	*Methotrexate*	Neurocognitive impairment Leukoencephalopathy Liver dysfunction Renal toxicity Decreased bone mineral density	*Winick N et al.*[95]; *Jones DP et al.*[79], *Wasilewski-Masker K et al.*[90]; *Castellino S et al.*[96];
Epipodophyllotoxins	*Etoposide Teniposide*	Acute myeloid leukemia	

Reproduced with permission from *Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. Nat Rev Cancer. 2014;14(1):61-70. doi:10.1038/nrc3634* PMID: 24304873

The childhood cancer survivor population is increasing in size and lifespan, and this specific population needs an effective evaluation and targeted interventions[97]. Therefore, the healthcare providers should consider the impact of medical as well as psychosocial consequences of cancer treatment on general health, mental health, and function pertinent to the developmental age of the survivor[76]. Moreover, research that identifies treatment, genetic, demographic, and psychosocial/behavioral predictors of adverse outcomes is essential for screening and monitoring aging survivors[76].

1.3. Hematopoietic stem cell transplantation (HSCT)

Hematopoietic stem cell transplantation (HSCT) is a common procedure during which a suspension of cells is harvested from marrow cavities and administered intravenously to patients or animals whose hematopoietic system is impaired by radio- and/or chemotherapy. Therefore, donor cells colonize ablated bone marrow and restore hematopoietic capacity[98].

Indications for the HSCT include the following malignant diseases: Acute Lymphocytic Leukemia, Acute Myeloid Leukemia, Hodgkin and Non-Hodgkin Lymphoma, Multiple Myeloma, Myelodysplastic Syndrome, Myelofibrosis, Essential Thrombocytosis, and Polycythemia Vera and Solid Tumors; as well as non-malignant conditions such as Aplastic Anemia, Sickle Cell Anemia, Thalassemia, Severe Combined Immune Deficiency Syndrome (SCID) and other (Khaddour K, Hana CK, Mewawalla P. Hematopoietic Stem Cell Transplantation. [Updated 2021 Jul 25]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536951 /).

Acute leukemia is one the most frequent indication for unrelated donor HSCT. Despite major advances in chemotherapy, HSCT remains the best therapeutic option for a subset of patients with high-risk disease at presentation, as well as for the majority of patients who relapse[99].

Allogeneic HSCT is a well-established but comprehensive treatment option for malignant and non-malignant diseases in pediatric patients. The most commonly used myeloablative and non-myeloablative conditioning regimens in this category of patients include alkylating agents such as *Busulfan* (BU) and *Cyclophosphamide*[100].

1.3.1. Busulfan (BU) as a key component in conditioning protocols for HSCT

Initially, oral BU was used as palliative treatment of chronic myelogenous leukemia and other myeloproliferative disorders[101]. However, concerns about stunted growth and retarded intellectual development have been associated with the use of total body irradiation (TBI) in children; hence, this has influenced a gradual shift to chemotherapy-only conditioning in pediatric transplantation[101]. BU (in various combinations) has often been included in conditioning regimens used in pediatric HSCT since early 1980s, and is now widely used as an alternative to total body irradiation[101].

Intravenous (IV) BU shows large pharmacokinetic (PK) variability. Higher exposure (expressed as area-under-the-curve; AUC) is associated with an increased risk of toxicity: e.g., mucositis, graft-versus-host disease (GvHD), veno-occlusive disease/sinusoidal obstructive syndrome (VOD/SOS), and transplant-related mortality (TRM); whereas low busulfan-AUC has been linked with a higher probability to graft-rejection or disease relapse[102]. Moreover, inter and intra-individual variability in the kinetics of BU is more common in children compared to adults, and the toxicity of BU based regimens remains a concern[100]. It has been suggested that some of this variability in BU pharmacokinetics and treatment outcomes, especially toxicity, could be predicted by genetic variants of enzymes involved in the metabolism of BU[100]. BU is metabolized via conjugation with glutathione (GSH), which is predominantly catalyzed by glutathione-S-transferase alpha 1 (encoded by the *GSTA1* gene). Several studies have investigated the relationship between genetic variants of the *GSTA1* and *GSTM1* genes and the pharmacokinetics of Bu[103].

References

1. Krier JB, Kalia SS, Green RC. Genomic sequencing in clinical practice: applications, challenges, and opportunities. *Dialogues in clinical neuroscience.* 2016;18(3):299-312.
2. Green RC, Goddard KAB, Jarvik GP, et al. Clinical Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of Genomic Medicine. *American journal of human genetics.* 2016;98(6):1051-1066.
3. Ansari M, Krajinovic M. Pharmacogenomics in cancer treatment defining genetic bases for inter-individual differences in responses to chemotherapy. *Curr Opin Pediatr.* 2007;19(1):15-22.
4. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature.* 2004;429(6990):464-468.
5. Johnson JA. Pharmacogenetics in clinical practice: how far have we come and where are we going? *Pharmacogenomics.* 2013;14(7):835-843.
6. Roses AD. Pharmacogenetics and the practice of medicine. *Nature.* 2000;405(6788):857-865.
7. Daly AK. Pharmacogenetics and human genetic polymorphisms. *The Biochemical journal.* 2010;429(3):435-449.
8. Squassina A, Manchia M, Manolopoulos VG, et al. Realities and expectations of pharmacogenomics and personalized medicine: impact of translating genetic knowledge into clinical practice. *Pharmacogenomics.* 2010;11(8):1149-1167.
9. Gervasini G, Benitez J, Carrillo JA. Pharmacogenetic testing and therapeutic drug monitoring are complementary tools for optimal individualization of drug therapy. *European journal of clinical pharmacology.* 2010;66(8):755-774.
10. Vogenberg FR, Isaacson Barash C, Pursel M. Personalized medicine: part 1: evolution and development into theranostics. *P t.* 2010;35(10):560-576.
11. Iyer L, Das S, Janisch L, et al. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J.* 2002;2(1):43-47.
12. Innocenti F, Undevia SD, Iyer L, et al. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J Clin Oncol.* 2004;22(8):1382-1388.
13. Minami H, Sai K, Saeki M, et al. Irinotecan pharmacokinetics/pharmacodynamics and UGT1A genetic polymorphisms in Japanese: roles of UGT1A1*6 and *28. *Pharmacogenetics and genomics.* 2007;17(7):497-504.
14. O'Dwyer PJ, Catalano RB. Uridine Diphosphate Glucuronosyltransferase (UGT) 1A1 and Irinotecan: Practical Pharmacogenomics Arrives in Cancer Therapy. *Journal of Clinical Oncology.* 2006;24(28):4534-4538.
15. Fujita K, Kubota Y, Ishida H, Sasaki Y. Irinotecan, a key chemotherapeutic drug for metastatic colorectal cancer. *World journal of gastroenterology.* 2015;21(43):12234-12248.
16. Fujita D, Saito Y, Nakanishi T, Tamai I. Organic Anion Transporting Polypeptide (OATP) 2B1 Contributes to Gastrointestinal Toxicity of Anticancer Drug SN-38, Active Metabolite of Irinotecan Hydrochloride. *Drug Metab Dispos.* 2016;44(1):1-7.
17. Manolio TA. Bringing genome-wide association findings into clinical use. *Nature Reviews Genetics.* 2013;14:549.
18. Visscher H, Ross CJ, Rassekh SR, et al. Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J Clin Oncol.* 2012;30(13):1422-1428.
19. Ansari M, Rezgui MA, Theoret Y, et al. Glutathione S-transferase gene variations influence BU pharmacokinetics and outcome of hematopoietic SCT in pediatric patients. *Bone Marrow Transplant.* 2013;48(7):939-946.
20. Baccichet A, Qualman SK, Sinnett D. Allelic loss in childhood acute lymphoblastic leukemia. *Leuk Res.* 1997;21(9):817-823.

21. Eichler EE. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *N Engl J Med.* 2019;381(1):64-74.
22. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
23. Nelson MR, Marnellos G, Kammerer S, et al. Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res.* 2004;14(8):1664-1668.
24. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
25. Durbin RM, Altshuler D, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.
26. Takeuchi F, Yanai K, Morii T, et al. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. *Genetics.* 2005;170(1):291-304.
27. Sachidanandam R, Weissman D, Schmidt SC, et al. *The International SNP Map Working Group*
A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001;409:928.
28. Marian AJ. Clinical Interpretation and Management of Genetic Variants. *JACC Basic Transl Sci.* 2020;5(10):1029-1042.
29. Eichler EE, Nickerson DA, Altshuler D, et al. Completing the map of human genetic variation.
Human Genome Structural Variation Working Group. Nature. 2007;447(7141):161-165.
30. Cario G, Stanulla M, Fine BM, et al. Distinct gene expression profiles determine molecular treatment response in childhood acute lymphoblastic leukemia. *Blood.* 2005;105(2):821-826.
31. Schulze TG, McMahon FJ. Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. *Am J Med Genet.* 2002;114(1):1-11.
32. Amos W, Driscoll E, Hoffman JI. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc Biol Sci.* 2011;278(1709):1183-1188.
33. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silicotools. *BMC Genetics.* 2013;14(1):39.
34. Manolio TA. Bringing genome-wide association findings into clinical use. *Nature Reviews Genetics.* 2013;14(8):549-558.
35. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308(5720):385-389.
36. Lee S, Abecasis Gonçalo R, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. *The American Journal of Human Genetics.* 2014;95(1):5-23.
37. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science.* 2012;335(6070):823-828.
38. Li B, Leal SM. Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *The American Journal of Human Genetics.* 2008;83(3):311-321.
39. Zhang X, Basile AO, Pendergrass SA, Ritchie MD. Real world scenarios in rare variant association analysis: the impact of imbalance and sample size on the power in silico. *BMC Bioinformatics.* 2019;20(1):46.
40. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372(9):793-795.
41. Nichols D, Arnold S, Weiss HL, et al. Pharmacogenomic potential in advanced cancer patients. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists.* 2019;76(7):415-423.

42. Vivot A, Boutron I, Ravaud P, Porcher R. Guidance for pharmacogenomic biomarker testing in labels of FDA-approved drugs. *Genet Med.* 2015;17(9):733-738.
43. Caudle KE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Relling MV, Klein TE. Evidence and resources to implement pharmacogenetic knowledge for precision medicine. *Am J Health Syst Pharm.* 2016;73(23):1977-1985.
44. Abdullah-Koolmees H, van Keulen AM, Nijenhuis M, Deneer VHM. Pharmacogenetics Guidelines: Overview and Comparison of the DPWG, CPIC, CPNDS, and RNPGx Guidelines. *Front Pharmacol.* 2020;11:595219.
45. Lee JW, Pussegoda K, Rassekh SR, et al. Clinical Practice Recommendations for the Management and Prevention of Cisplatin-Induced Hearing Loss Using Pharmacogenetic Markers. *Ther Drug Monit.* 2016;38(4):423-431.
46. Aminkeng F, Ross CJ, Rassekh SR, et al. Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. *Br J Clin Pharmacol.* 2016;82(3):683-695.
47. Cheok MH, Pottier N, Kager L, Evans WE. Pharmacogenetics in acute lymphoblastic leukemia. *Semin Hematol.* 2009;46(1):39-51.
48. Gaballa M, Ramos CA. Overview of Normal Hematopoiesis. In: Mims MP, Miller-Chism C, Sosa IR, eds. New York: Springer Publishing Company:1-24.
49. Aqmasheh S, Shamsasanjan K, Akbarzadehlaleh P, Pashoutan Sarvar D, Timari H. Effects of Mesenchymal Stem Cell Derivatives on Hematopoiesis and Hematopoietic Stem Cells. *Adv Pharm Bull.* 2017;7(2):165-177.
50. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature.* 2001;414(6859):105-111.
51. Riether C, Schürch CM, Ochsenbein AF. Regulation of hematopoietic and leukemic stem cells by the immune system. *Cell Death & Differentiation.* 2015;22(2):187-198.
52. Zhang J, Niu C, Ye L, et al. Identification of the haematopoietic stem cell niche and control of the niche size. *Nature.* 2003;425(6960):836-841.
53. Colmone A, Amorim M, Pontier AL, Wang S, Jablonski E, Sipkins DA. Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science.* 2008;322(5909):1861-1865.
54. Wong SW, Lenzini S, Shin JW. Perspective: Biophysical regulation of cancerous and normal blood cell lineages in hematopoietic malignancies. *APL Bioeng.* 2018;2(3):031802.
55. Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. *New England Journal of Medicine.* 2015;373(16):1541-1552.
56. Katz AJ, Chia VM, Schoonen WM, Kelsh MA. Acute lymphoblastic leukemia: an assessment of international incidence, survival, and disease burden. *Cancer Causes & Control.* 2015;26(11):1627-1642.
57. Faderl S, O'Brien S, Pui CH, et al. Adult acute lymphoblastic leukemia: concepts and strategies. *Cancer.* 2010;116(5):1165-1176.
58. Cooper SL, Brown PA. Treatment of pediatric acute lymphoblastic leukemia. *Pediatr Clin North Am.* 2015;62(1):61-73.
59. Smith M, Bleyer A, Crist W, Murphy S, Sallan SE. Uniform criteria for childhood acute lymphoblastic leukemia risk classification. *J Clin Oncol.* 1996;14(2):680-681.
60. Santiago R, Tran TH. Genomics and precision medicine in pediatric acute lymphoblastic leukemia. *Journal of Translational Genetics and Genomics.* 2021;5(4):380-395.
61. Simone JV. History of the treatment of childhood ALL: a paradigm for cancer cure. *Best Pract Res Clin Haematol.* 2006;19(2):353-359.
62. The Genome Aggregation Database (gnomAD). In.

63. Pui CH, Mahmoud HH, Rivera GK, et al. Early intensification of intrathecal chemotherapy virtually eliminates central nervous system relapse in children with acute lymphoblastic leukemia. *Blood.* 1998;92(2):411-415.
64. Pui CH, Sandlund JT, Pei D, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood.* 2004;104(9):2690-2696.
65. Pui CH, Campana D, Pei D, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med.* 2009;360(26):2730-2741.
66. Cheung YT, Krull KR. Neurocognitive outcomes in long-term survivors of childhood acute lymphoblastic leukemia treated on contemporary treatment protocols: A systematic review. *Neurosci Biobehav Rev.* 2015;53:108-120.
67. Essig S, Li Q, Chen Y, et al. Risk of late effects of treatment in children newly diagnosed with standard-risk acute lymphoblastic leukaemia: a report from the Childhood Cancer Survivor Study cohort. *Lancet Oncol.* 2014;15(8):841-851.
68. Tran TH, Hunger SP. The genomic landscape of pediatric acute lymphoblastic leukemia and precision medicine opportunities. *Seminars in Cancer Biology.* 2020.
69. Schultz KR, Carroll A, Heerema NA, et al. Long-term follow-up of imatinib in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia: Children's Oncology Group study AALL0031. *Leukemia.* 2014;28(7):1467-1471.
70. Abaji R, Krajinovic M. Thiopurine S-methyltransferase polymorphisms in acute lymphoblastic leukemia, inflammatory bowel disease and autoimmune disorders: influence on treatment response. *Pharmgenomics Pers Med.* 2017;10:143-156.
71. Cheok MH, Evans WE. Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy. *Nature Reviews Cancer.* 2006;6(2):117-129.
72. Moriyama T, Nishii R, Perez-Andreu V, et al. NUDT15 polymorphisms alter thiopurine metabolism and hematopoietic toxicity. *Nat Genet.* 2016;48(4):367-373.
73. Yang SK, Hong M, Baek J, et al. A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nat Genet.* 2014;46(9):1017-1020.
74. Yang JJ, Landier W, Yang W, et al. Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia. *J Clin Oncol.* 2015;33(11):1235-1242.
75. Relling MV, Schwab M, Whirl-Carrillo M, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for Thiopurine Dosing Based on TPMT and NUDT15 Genotypes: 2018 Update. *Clin Pharmacol Ther.* 2019;105(5):1095-1105.
76. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. *Nat Rev Cancer.* 2014;14(1):61-70.
77. Ellison LF, De P, Mery LS, Grundy PE. Canadian cancer statistics at a glance: cancer in children. *Cmaj.* 2009;180(4):422-424.
78. Kenney LB, Cohen LE, Shnorhavorian M, et al. Male reproductive health after childhood, adolescent, and young adult cancers: a report from the Children's Oncology Group. *J Clin Oncol.* 2012;30(27):3408-3416.
79. Jones DP, Spunt SL, Green D, Springate JE. Renal late effects in patients treated for cancer in childhood: a report from the Children's Oncology Group. *Pediatr Blood Cancer.* 2008;51(6):724-731.
80. Ritchey M, Ferrer F, Shearer P, Spunt SL. Late effects on the urinary bladder in patients treated for cancer in childhood: a report from the Children's Oncology Group. *Pediatr Blood Cancer.* 2009;52(4):439-446.

81. Huang TT, Hudson MM, Stokes DC, Krasin MJ, Spunt SL, Ness KK. Pulmonary outcomes in survivors of childhood cancer: a systematic review. *Chest.* 2011;140(4):881-901.
82. Metzger ML, Meacham LR, Patterson B, et al. Female reproductive health after childhood, adolescent, and young adult cancers: guidelines for the assessment and management of female reproductive complications. *J Clin Oncol.* 2013;31(9):1239-1247.
83. Brock PR, Knight KR, Freyer DR, et al. Platinum-induced ototoxicity in children: a consensus review on mechanisms, predisposition, and protection, including a new International Society of Pediatric Oncology Boston ototoxicity scale. *J Clin Oncol.* 2012;30(19):2408-2417.
84. Feldman DR, Schaffer WL, Steingart RM. Late cardiovascular toxicity following chemotherapy for germ cell tumors. *J Natl Compr Canc Netw.* 2012;10(4):537-544.
85. Hijiya N, Ness KK, Ribeiro RC, Hudson MM. Acute leukemia as a secondary malignancy in children and adolescents: current findings and issues. *Cancer.* 2009;115(1):23-35.
86. Skinner R. Nephrotoxicity--what do we know and what don't we know? *J Pediatr Hematol Oncol.* 2011;33(2):128-134.
87. van der Pal HJ, van Dalen EC, van Delden E, et al. High risk of symptomatic cardiac events in childhood cancer survivors. *J Clin Oncol.* 2012;30(13):1429-1437.
88. Mulrooney DA, Yeazel MW, Kawashima T, et al. Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort. *Bmj.* 2009;339:b4606.
89. Lipshultz SE, Lipsitz SR, Sallan SE, et al. Chronic progressive cardiac dysfunction years after doxorubicin therapy for childhood acute lymphoblastic leukemia. *J Clin Oncol.* 2005;23(12):2629-2636.
90. Wasilewski-Masker K, Kaste SC, Hudson MM, Esiashvili N, Mattano LA, Meacham LR. Bone mineral density deficits in survivors of childhood cancer: long-term follow-up guidelines and review of the literature. *Pediatrics.* 2008;121(3):e705-713.
91. Kadan-Lottick NS, Dinu I, Wasilewski-Masker K, et al. Osteonecrosis in adult survivors of childhood cancer: a report from the childhood cancer survivor study. *J Clin Oncol.* 2008;26(18):3038-3045.
92. Whelan KF, Stratton K, Kawashima T, et al. Ocular late effects in childhood and adolescent cancer survivors: a report from the childhood cancer survivor study. *Pediatr Blood Cancer.* 2010;54(1):103-109.
93. Jain P, Gulati S, Seth R, Bakhshi S, Toteja GS, Pandey RM. Vincristine-induced neuropathy in childhood ALL (acute lymphoblastic leukemia) survivors: prevalence and electrophysiological characteristics. *J Child Neurol.* 2014;29(7):932-937.
94. Ness KK, Jones KE, Smith WA, et al. Chemotherapy-related neuropathic symptoms and functional impairment in adult survivors of extracranial solid tumors of childhood: results from the St. Jude Lifetime Cohort Study. *Arch Phys Med Rehabil.* 2013;94(8):1451-1457.
95. Winick N. Neurocognitive outcome in survivors of pediatric cancer. *Curr Opin Pediatr.* 2011;23(1):27-33.
96. Castellino S, Muir A, Shah A, et al. Hepato-biliary late effects in survivors of childhood and adolescent cancer: a report from the Children's Oncology Group. *Pediatr Blood Cancer.* 2010;54(5):663-669.
97. Bitsko MJ, Cohen D, Dillon R, Harvey J, Krull K, Klosky JL. Psychosocial Late Effects in Pediatric Cancer Survivors: A Report From the Children's Oncology Group. *Pediatr Blood Cancer.* 2016;63(2):337-343.
98. Wright DE, Wagers AJ, Gulati AP, Johnson FL, Weissman IL. Physiological Migration of Hematopoietic Stem and Progenitor Cells. *Science.* 2001;294(5548):1933-1936.

99. MacMillan ML, Davies SM, Nelson GO, et al. Twenty Years of Unrelated Donor Bone Marrow Transplantation for Pediatric Acute Leukemia Facilitated by the National Marrow Donor Program. *Biology of Blood and Marrow Transplantation.* 2008;14(9, Supplement):16-22.
100. Huezo-Diaz P, Uppugunduri CR, Tyagi AK, Krajinovic M, Ansari M. Pharmacogenetic aspects of drug metabolizing enzymes in busulfan based conditioning prior to allogenic hematopoietic stem cell transplantation in children. *Curr Drug Metab.* 2014;15(3):251-264.
101. Ciurea SO, Andersson BS. Busulfan in hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant.* 2009;15(5):523-536.
102. Bartelink IH, Lalmohamed A, van Reij EM, et al. Association of busulfan exposure with survival and toxicity after haemopoietic cell transplantation in children and young adults: a multicentre, retrospective cohort analysis. *Lancet Haematol.* 2016;3(11):e526-e536.
103. Uppugunduri CRS, Rezgui MA, Diaz PH, et al. The association of cytochrome P450 genetic polymorphisms with sulfolane formation and the efficacy of a busulfan-based conditioning regimen in pediatric patients undergoing hematopoietic stem cell transplantation. *The Pharmacogenomics Journal.* 2014;14(3):263-271.

Section A

Chapter 2

This chapter presents the working hypotheses that formed the basis of this research and identifies the goals for which the research was conducted. It also describes the main stages of the genetic association study pipeline.

2.1. Study Hypotheses

The presence of common and rare genetic polymorphisms in patients' genetic constitution can contribute to the substantial variability regarding the occurrence and severity of treatment related toxicities in a post hematopoietic stem cell transplantation period in pediatric patients, as well as of treatment-related late-adverse effects in survivors of childhood Acute Lymphoblastic Leukemia by modifying disease risk in conjunction with treatment.

The application of two different approaches – a candidate genes and an exome-wide association study, using common and rare genetic variants derived through whole-exome sequencing, may allow the discovery of new genetic markers associated with the risk of developing treatment-related complications both during treatment and in the long-term.

Assessing and understanding the individual contribution of each of the identified germline variants, as well as their combined effect, can provide valuable information about the molecular mechanisms that differently predispose patients to studied treatment outcomes. The reported genetic markers, along with other known risk factors, can further lead to prediction models identifying patients at increased risk for these treatment-related complications.

2.2. Research Objectives

- to build a pipeline of the genetic association analyses of common and rare variants, using bioinformatics tools such as PLINK, SKAT-O and programming language R.

- to use exome-wide sequencing data to perform candidate genes and exome-wide association studies in order to identify common and rare germline variations associated with:

  - long-term treatment-related cardiovascular and neurocognitive morbidities, as well as anxiety and depression in childhood ALL survivors.

  - acute complications post hematopoietic stem cell transplantation, including hepatic sinusoidal obstruction syndrome and acute graft versus host disease.

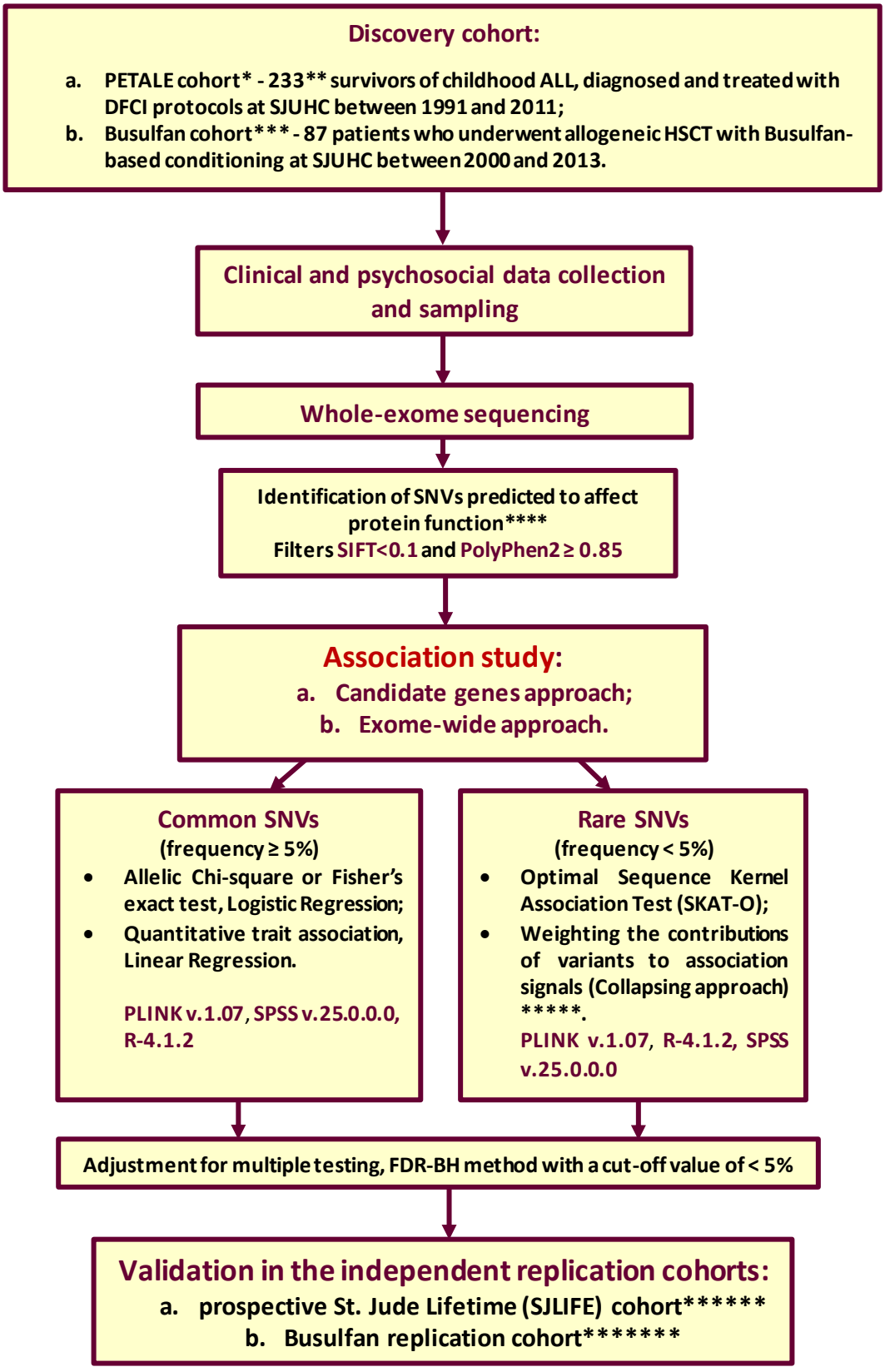
- to assess top-ranking associations through stratified and multivariate analyses.

- to validate the effect of the identified variants in independent replication cohorts.

2.3. Materials and methods

2.3.1. Genetic association study pipeline

Figure 1. Genetic association study pipeline.



PETALE: **P**révenir les **E**ffets **TA**rdifs des traitements de la **LE**ucémie lymphoblastique aigue; ALL: Acute Lymphoblastic Leukemia; DFCI: Dana-Farber Cancer Institute; SJUHC: Sainte-Justine University Health Center; HSCT: Hematopoietic Stem Cell Transplantation; SNV: Single nucleotide variation; FDR-BH: Benjamini–Hochberg false discovery rate.

*To test and confirm predominant European ancestry in the PETALE cohort the HapMap genotype reference data was used for Principal components analysis (a statistical method commonly used in population genetics to determine the pattern of distribution of genetic variation by geographic location and ethnicity).

** The total number of survivors included in different studies varies based on quality of the sequencing data, as well as on inclusion/exclusion criteria.

*** For the Busulfan cohort, only an exome-wide approach was applied, the set of the studied variants also included UTR variants.

**** The predicted functional impact of missense, nonsense and splicing common and rare variants was assessed in silico using SIFT and PolyPhen-2 filters[1-3].

***** Collapsing approach that combines several rare variants into a single variable[4,5], with iterative exclusion of each single variant, was additionally performed to allow weighting variant contributions to association signals.

****** Established in 2007 at St. Jude Children's Research Hospital, the prospective St. Jude Lifetime (SJLIFE) cohort, represents the cohort of survivors of pediatric, adolescent, and young adult cancers, as well as frequency-matched community controls. SJLIFE constitutes the collection of comprehensive treatment data on all participants, providing protocol-based medical assessments, patient-reported outcomes, validation of self-reported medical events, allowing the performance of periodic longitudinal evaluations and collection of biologic specimens[6].

******* Busulfan replication cohort consisted of 121 pediatric patients who underwent allogeneic HSCT from 2001-2015 in four different centers in Europe and Canada (Geneva University Hospital, University Medical Center Utrecht, Leiden University Medical Center, Robert Debré Hospital, Paris and Alberta Children's Hospital, Calgary).

SIFT (**S**orting **I**ntolerant **F**rom **T**olerant) is a bioinformatic tool that predicts whether an amino acid substitution affects protein function, substitutions with score less than 0.1 are predicted to affect protein function[1].

PolyPhen-2 (**Poly**morphism **Phen**otyping **v2**) is a bioinformatic tool which predicts possible impact of an amino acid substitution on the structure and function of a protein using straightforward physical and comparative evolutionary considerations, variants with score 0.85 and above are predicted to be damaging[2,3].

PLINK is an open-source genome-wide association analysis toolkit for performing large-scale analyses with high computational efficiency[7,8].

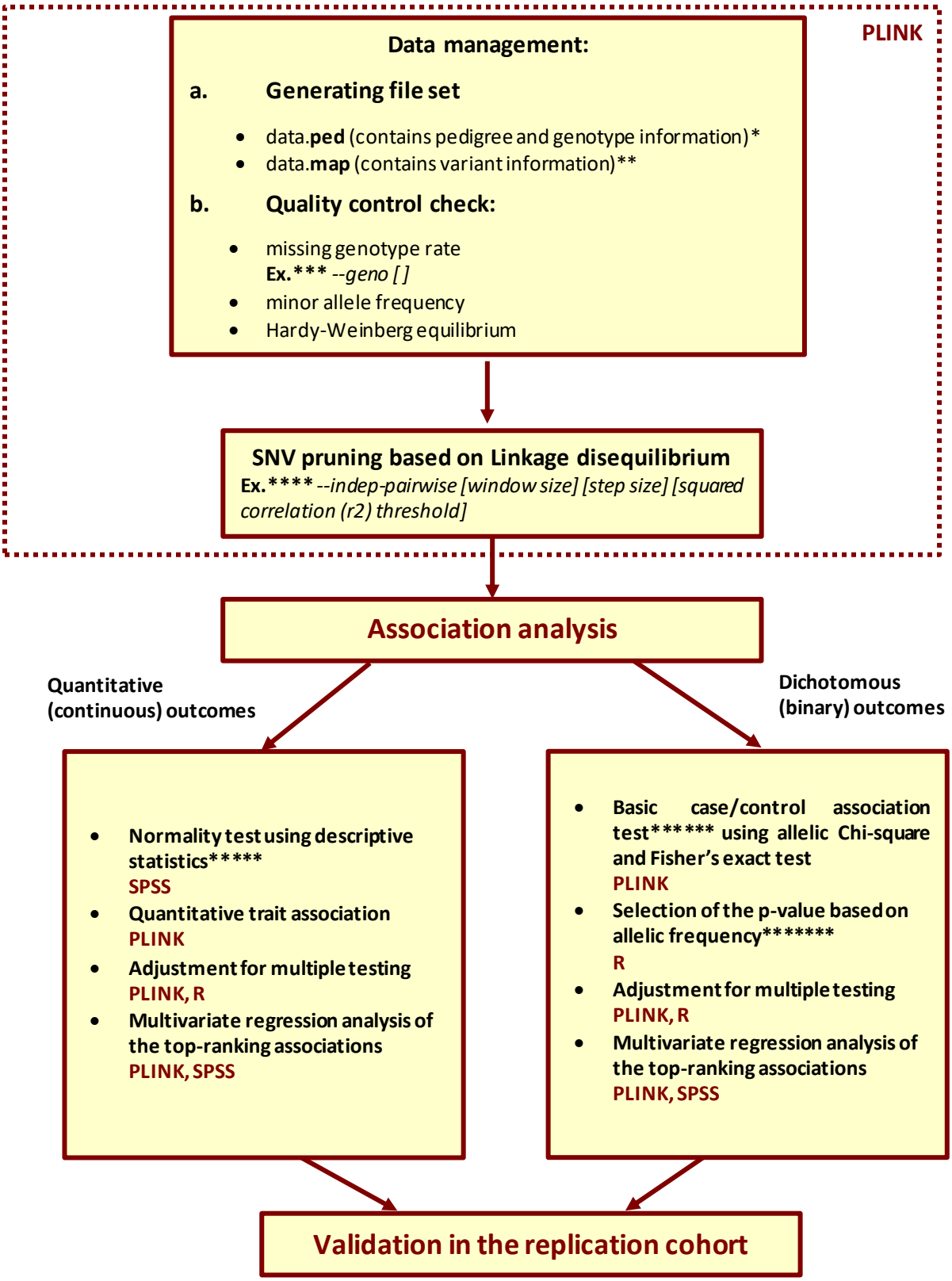
SPSS (**S**tatistical **P**ackage for the **S**ocial **S**ciences) is a software package used for statistical analysis.

R is a programming language and free software environment for statistical computing and graphics.

The Optimal Sequence Kernel Association Test (SKAT-O) will be discussed in one of the next sections.

2.3.1.1. Common variants

Figure 2. The summary of the common variants' analysis.



* each line of the **ped** file describes pedigree and genotype information for each sample, that contains mandatory 6 columns[7]**:**

Family ID
Individual ID
Paternal ID (0 if unplaced)
Maternal ID (0 if unplaced)
Sex (1=male; 2=female; other=unknown)
Phenotype (binary or continuous format)
Genotypes (column 7 onwards) can be represented by different characters (for example: **1,2,3,4** or **A,C,G,T**) and all markers should be biallelic.

** each line of the **map** file describes a single variant and contains mandatory 4 columns[7]**:**

Chromosome (1-22, X, Y or 0 if unplaced)
rs number or SNV identifier
Genetic distance (Morgans) or 0 if unplaced
Base-pair position (bp units)

*** ***--geno [ ]***

This command filters out all variants with missing call rates exceeding the provided value[7].

For example, ***--geno 0.2*** filters out variants exceeding a missing rate of 20%.

**** ***--indep-pairwise [window size] [step size] [squared correlation (r2) threshold]***

This command considers a specified window size in variant count (or kilobase, if this modifier is present in a dataset) and calculates Linkage disequilibrium (LD) between each pair of SNVs in the window, therefore at each step one of a pair of variants in the current window with squared correlation greater than the threshold is removed, after the window shifts forward, and procedure is repeated until no such pairs remain[7].

For example, ***--indep-pairwise 50 5 0.8***

Here, a window of **50** SNVs is considered, and LD is calculated between each pair of SNVs in this window, then one of a pair of SNVs if the LD is greater than **0.8** (80%) is removed, after the window shifts **5** SNVs forward and the procedure is repeated.

***** Normality tests are used to determine whether a dataset respects a normal distribution and to calculate the probability that the random variable underlying the dataset is normally distributed. The main tests for the assessment of normality are Kolmogorov-Smirnov (K-S) test, Shapiro-Wilk test and others[9].

****** Participants with and without indicated complications were defined as cases and controls, respectively. The basic association test is based on comparing allele frequencies between cases and controls.

******* P values were calculated by Chi-square and Fisher exact test. The Chi-square p-value was considered only for the variants with the minimum number of observations per cell =>5 in a 2 by 2 table:

	Allele a	Allele A	Total
Cases	$2n_2 + n_1$	$2n_0 + n_1$	2n
Controls	$2m_2 + m_1$	$2m_0 + m_1$	2m

A *reference allele*

a *variant allele*

If at least one of the cells has a frequency less than 5, then the p-value from the Fisher exact test was considered.

********R programming software was used to code the p-value selection algorithm as shown in **Figure 3**.

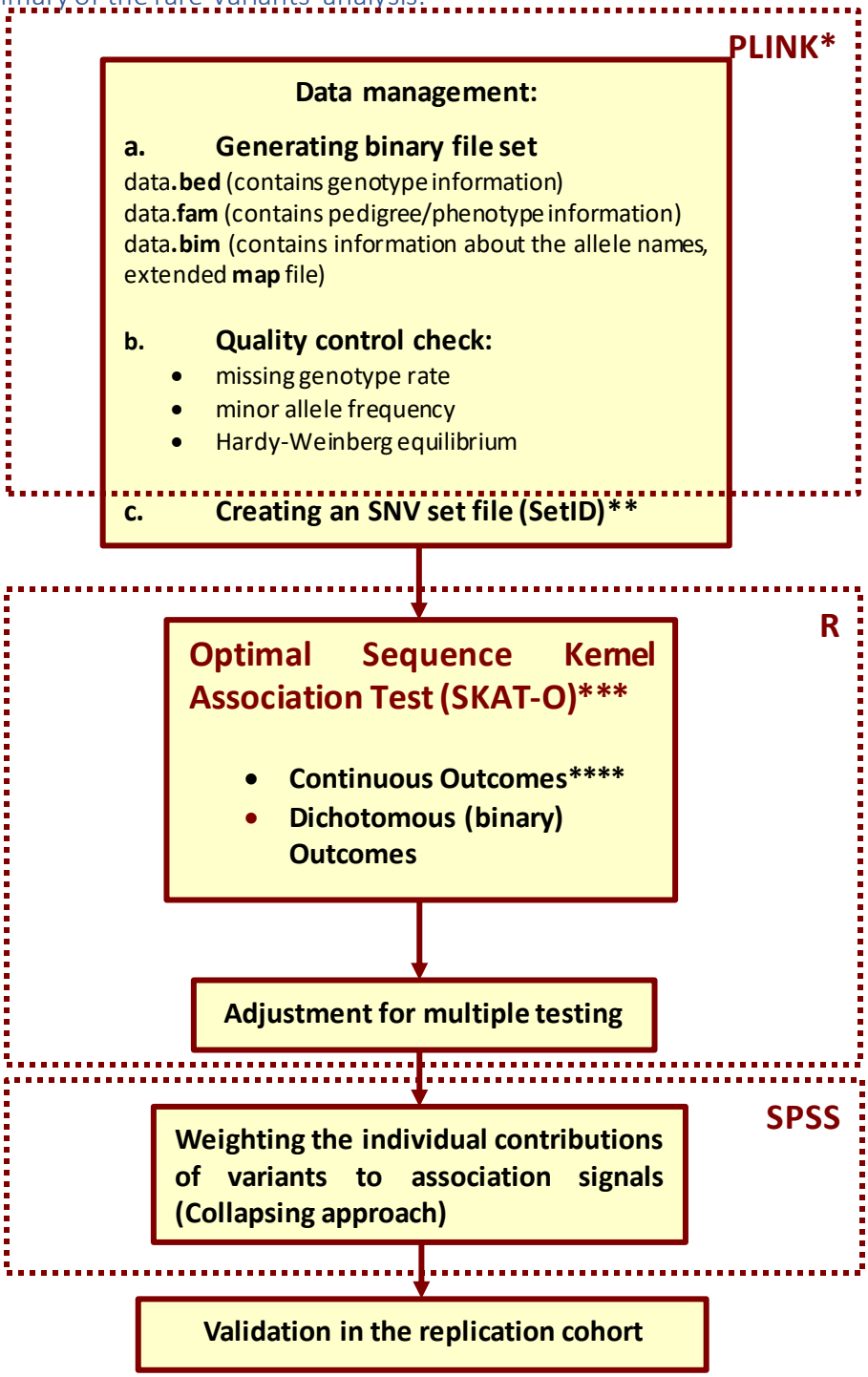
```
library(reshape)
library(reshape2)
library(data.table)
library(tidyr)
library(stringr)
library(dplyr)


df <- read.csv("kat_wes_2021_neuropsy_229p_R.model.csv",
               stringsAsFactors=FALSE, header=TRUE)

dftest <- subset(df, TEST == 'ALLELIC')

dfspiltaff <- data.frame(do.call('rbind', strsplit(as.character(dftest$AFF), '/', fixed = TRUE)))
dfspiltunaff <- data.frame(do.call('rbind', strsplit(as.character(dftest$UNAFF), '/', fixed = TRUE)))

dftest$AFF1 <- dfspiltaff$X1
dftest$AFF2 <- dfspiltaff$X2
dftest$UNAFF1 <- dfspiltunaff$X1
dftest$UNAFF2 <- dfspiltunaff$X2

dffisher <- read.csv("kat_wes_2021_neuropsy_229p_R.assoc.fisher.csv",
                     stringsAsFactors=FALSE, header=TRUE)

dftest$Pfisher <- dffisher$P

dftest$Presult <- ifelse(dftest$AFF1 >= 5 | dftest$AFF2 >= 5 | dftest$UNAFF1 >= 5 |
                           dftest$UNAFF2 >= 5, dftest$P, dftest$Pfisher)
```

Figure 3. Fragment of the code for the selection of the p-value from chi-square or Fisher exact test based on allelic frequency.

To count the alleles, a table consisting of 4 cells is created, in which, for each polymorphism, the frequency of carriers/non-carriers of the variant allele among cases and controls is indicated. If the number of observations in any of the categories (cells) is equal to or greater than 5, then the p-value of the chi-square test has been taken into account, similarly, if it is less than 5, then the p-value of Fisher's exact test is taken into account and added to the table of results.

2.3.1.2. Rare variants

Figure 4. The summary of the rare variants' analysis.



* The first two steps are performed using PLINK, similar to the common variants' analysis.

** SNV set file represents the complete list of variants assigned to specific genomic region (gene), the example of the list is shown in **Figure 5.**

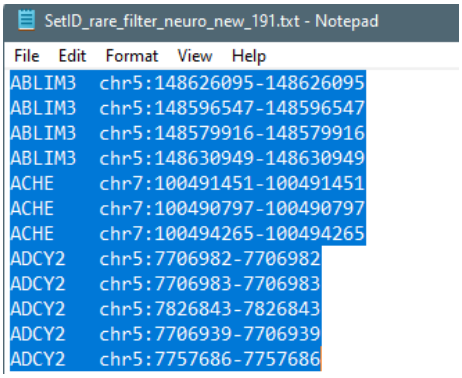


Figure 5. SNV set file structure.

The structure of the SNV set file consists of 2 columns, the first is the name of the gene, the second is the variant identifier (rs number) or its position (hg19 position is shown here).

The genomic regions (genes) that contained at least 2 rare variants were considered for analysis.

*** Optimal Sequence Kernel Association Test (SKAT-O)[10,11]

Common approaches for the powerful rare variant association analysis include burden tests[12] and sequence kernel association tests (SKAT)[13].

The Sequence kernel association test (SKAT) is a non-burden test. Instead of collapsing variants in a particular genetic region into a single burden variable and then to test for the cumulative effects of variants in that region[11] (defined as burden method), SKAT aggregates individual score test statistics of variants with weights (when variant effects are modeled linearly) and efficiently computes a region (a gene) level p value. Therefore, when a genetic region has variants with both protective and deleterious effect or many noncausal variants, SKAT is particularly powerful[11]. Nevertheless, it has several limitations. For example, SKAT can be less powerful than burden tests if a significant proportion of rare variants in a tested genomic region are truly causal and influence the phenotype of interest in the same direction[11,14]. Additionally, SKAT uses large-sample-based p value calculations, which can produce conservative type I errors for small-sample case-control sequencing association studies[11]. In order to address these limitations, the Optimal Sequence Kernel Association Test (SKAT-O) was proposed, which represents a unified test for rare-variant

effects by using the data to find the optimal linear combination of the burden test and SKAT[11]. Simply put, SKAT-O automatically functions like the burden test when the burden test is more powerful than SKAT, and functions like SKAT when the SKAT is more powerful than the burden test, which maintains the power in both scenarios[11].

**** R programming software was used to code the SKAT-O test as shown in **Figure 6**.

```
library(SKAT)
Generate_SSD_SetID("/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.bed",
                   "/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.bim",
                   "/Users/ketty/plink-1.07-dos/cardio_ph2_all_233p.fam",
                   "/Users/ketty/plink-1.07-dos/SetID_wes_petale_2020_rare_func_and_splicing_final.txt",
                   "/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.SSD",
                   "/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.Info",Is.FlipGenotype=TRUE)
SSD.INFO<-Open_SSD("/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.SSD",
                   "/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.Info")
SSD.INFO
options(max.print=1000000)
SetIndex<-SSD.INFO$SetInfo$SetIndex
SetIndex
SetID<-SSD.INFO$SetInfo$SetID
SetID
FAM<-Read_Plink_FAM("/Users/ketty/plink-1.07-dos/cardio_ph3_all_233p.fam", Is.binary=FALSE)
y<-FAM$Phenotype
obj<-SKAT_Null_Model(y ~ 1, out_type="C")
```

Figure 6. Fragment of the code for continuous outcome analysis with the SKAT-O test.

A set of binary input files containing the outcome of interest in continuous format, as well as an SNV set file, were used to perform the SKAT-O test.

2.4. References

1. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.
2. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 2006;7:61-80.
3. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7:Unit7.20.
4. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet Epidemiol.* 2011;35 Suppl 1(Suppl 1):S12-17.
5. Sung YJ, Korthauer KD, Swartz MD, Engelman CD. Methods for collapsing multiple rare variants in whole-genome sequence data. *Genet Epidemiol.* 2014;38 Suppl 1(0 1):S13-20.
6. Howell CR, Bjornard KL, Ness KK, et al. Cohort Profile: The St. Jude Lifetime Cohort Study (SJLIFE) for paediatric cancer survivors. *Int J Epidemiol.* 2021;50(1):39-49.
7. Purcell S. PLINK v1.07. https://zzz.bwh.harvard.edu/plink/index.shtml. Accessed.
8. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
9. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab.* 2012;10(2):486-489.
10. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics.* 2012;13(4):762-775.
11. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012;91(2):224-237.
12. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95(1):5-23.
13. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011;89(1):82-93.
14. Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011;7(3):e1001322.

Section B

Chapter 3

Influence of genetic factors on long-term treatment related neurocognitive complications, and on anxiety and depression in survivors of childhood acute lymphoblastic leukemia: The Petale study.

PLoS One. 2019 Jun 10;14(6): e0217314. doi: 10.1371/journal.pone.0217314. eCollection 2019.

PMID: 31181069 PMCID: PMC6557490 DOI: 10.1371/journal.pone.0217314

Influence of genetic factors on long-term treatment related neurocognitive complications, and on anxiety and depression in survivors of childhood acute lymphoblastic leukemia: The Petale study.

Running title: Genetics of Neurocognitive Complications in survivors of childhood ALL.

Kateryna Petrykey[1,2], Sarah Lippé[1,3], Philippe Robaey[1,5,6,7], Serge Sultan[1,3], Julie Laniel[1,3], Simon Drouin[1], Laurence Bertout[1], Patrick Beaulieu[1], Pascal St-Onge[1], Aubrée Boulet-Craig[1,3], Aziz Rezgui[1], Yutaka Yasui[8], Yadav Sapkota[8], Kevin R. Krull[8], Melissa M. Hudson[8,9], Caroline Laverdière[1,4], Daniel Sinnett[1,4] and Maja Krajinovic[1,2,4*]

1. Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada.
2. Department of Pharmacology and Physiology, Université de Montréal, Montreal (Quebec), Canada.
3. Department of Psychology, Université de Montréal, Montreal (Quebec), Canada.
4. Department of Pediatrics, Université de Montréal, Montreal (Quebec), Canada.
5. Children's Hospital of Eastern Ontario, Ottawa (Ontario), Canada.
6. Department of Psychiatry, Université de Montréal, Montreal (Quebec), Canada.
7. Department of Psychiatry, University of Ottawa, Ottawa (Ontario), Canada.
8. Epidemiology and Cancer Control Department, St. Jude Children's Research Hospital, Memphis (TN), USA.
9. Oncology Department, St. Jude Children's Research Hospital, Memphis (TN), USA.

***Corresponding author:**

Maja Krajinovic
E-mail: maja.krajinovic@umontreal.ca

Conflict of interest statement

The authors declare no potential conflicts of interest.

3.1. ABSTRACT

Background: A substantial number of survivors of childhood acute lymphoblastic leukemia suffer from treatment-related late adverse effects including neurocognitive impairment. While multiple studies have described neurocognitive outcomes in childhood acute lymphoblastic leukemia (ALL) survivors, relatively few have investigated their association with individual genetic constitution.

Methods: To further address this issue, genetic variants located in 99 genes relevant to the effects of anticancer drugs and in 360 genes implicated in nervous system function and predicted to affect protein function, were pooled from whole exome sequencing data of childhood ALL survivors (PETALE cohort) and analyzed for an association with neurocognitive complications, as well as with anxiety and depression. Variants that sustained correction for multiple testing were genotyped in entire cohort (n=236) and analyzed with same outcomes.

Results: Common variants in *MTR, PPARA, ABCC3, CALML5, CACNB2* and *PCDHB10* genes were associated with deficits in neurocognitive tests performance, whereas a variant in *SLCO1B1* and *EPHA5* genes was associated with anxiety and depression. Majority of associations were modulated by intensity of treatment. Associated variants were further analyzed in an independent SJLIFE cohort of 545 ALL survivors. Two variants, *rs1805087* in methionine synthase, *MTR* and *rs58225473* in voltage-dependent calcium channel protein encoding gene, *CACNB2* are of particular interest, since associations of borderline significance were found in replication cohort and remain significant in combined discovery and replication groups (OR=1.5, 95% CI, 1-2.3; p=0.04 and OR=3.7, 95% CI, 1.25-11; p=0.01, respectively). Variant *rs4149056* in *SLCO1B1* gene also deserves further attention since previously shown to affect methotrexate clearance and short-term toxicity in ALL patients.

Conclusions: Current findings can help understanding of the influence of genetic component on long-term neurocognitive impairment. Further studies are needed to confirm whether the identified variants may be useful in identifying survivors at increased risk of these complications.

Keywords

Childhood acute lymphoblastic leukemia, cancer survivors, late adverse effects, genetic factors, association study, whole exome sequencing, neurocognitive complications, anxiety, depression.

3.2. INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most frequent childhood cancer[1] accounting for approximately 25% of all cases[2]. The five-year survival rate of childhood ALL is currently greater than 85% due to the optimization of multi-agent risk-adapted treatment strategies[2,3]. However, the exposure to specific chemotherapeutic agents and/or cranial radiation therapy during a susceptible period of child development results in late-adverse effects (LAEs)[4-6] including neurocognitive impairments[2]. Clinically significant deficits among ALL survivors are most commonly found in attention[7-12], working memory[13], processing speed[9,14,15] and executive functions, such as verbal fluency and cognitive flexibility[16]. Neurocognitive impairment in childhood ALL survivors persist for many years after treatment[17,18]. Large survey studies like the Childhood Cancer Survivors Study (CCSS) among other studies conducted in childhood ALL survivors[7,9,19], have demonstrated higher risk of depression, anxiety, behavioural difficulties, distress, as well as post-traumatic symptoms compared to siblings[20-25]. Longitudinal follow-up in long-term survivors have indicated that frequency of distress evolves over time, with more than 10% of survivors experiencing significant increase in depression as well as in anxiety[26].

The childhood ALL survivor population is increasing in size and lifespan, and this specific population needs an effective evaluation and targeted interventions[27]. Thus, better understanding of LAEs and factors contributing to their development is important to guide survivorship health surveillance and strategies to prevent or remediate treatment-related toxicities[2,3]. Here we assessed the role of genetic factors in neurocognitive impairments along with anxiety and depression by interrogating the relationship between the above-mentioned complications and genotypic profiling of 459 candidate genes obtained through whole exome sequencing (WES) of childhood ALL survivors.

3.3. STUDY POPULATION AND METHODS

3.3.1. Discovery cohort

The discovery cohort included 236 patients diagnosed and treated for childhood ALL according to Dana Farber Cancer Institute (DFCI) ALL 87-01 to 05-01 protocols at Sainte-Justine University Health Center (SJUHC), Montreal, (Quebec), Canada. The participants were recruited during 2013-2015 in the context of the PETALE study, a multidisciplinary research project with the goal to identify and comprehensively characterize associated predictive biomarkers of long-term treatment related complications in childhood ALL survivors[6]. Eligible participants were younger than 19 years old at diagnosis, at least 5 years after diagnosis of ALL and older than 12 years at evaluation, without history of relapse or refractory ALL or neurological condition or Down syndrome and had not received a hematopoietic stem cell transplant. The time from end of treatment to evaluation ranged from 5-26 year with a median of 13 years. The patients were classified to standard (SR) and high risk (HR) groups based on prognostic factors, including age, white blood cell count, immunophenotype, and central nervous system (CNS) status at diagnosis[28,29]. The frequency of patients assigned to SR and HR groups during the treatment was 46.6% and 53.4%, respectively. They were almost exclusively of reported French Canadian descent (>95%). The HapMap genotype reference data[30,31] was used for Principal components analysis (PCA)[30] to test and confirm predominant European ancestry (**S1 Fig**).

3.3.2. Replication cohort

The replication cohort consisted of 545 ALL survivors (274 males and 271 females) of European ancestry (based on genotype data) enrolled in the St. Jude Lifetime Cohort (SJLIFE) study and were evaluated using the same testing procedures as in PETALE cohort. Participants were younger than 19 years old at diagnosis and younger than 40 years old at SJLIFE evaluation, with no diagnosis of neurologic condition or Down syndrome, no history of relapse and had not received a hematopoietic stem cell transplant. The time from primary cancer diagnosis to the most recent date of neurocognitive

evaluation ranged from 10.95 – 45.60 years with a median of 26.84 years. The risk group assignment during the treatment (SR and HR groups) was not available for this cohort.

3.3.3. Neuropsychological evaluation

A neurocognitive evaluation was performed using standardized testing procedures. It included three indices from two cognitive measures that reflect common impairments among childhood ALL survivors and are also good predictors of general neuropsychological outcomes[32]: Trail Making Test – Condition 4 - Letter-Number Sequencing score and Verbal Fluency – Condition 1 – Letter fluency score from the Delis-Kaplan Executive Function System (D-KEFS)[33]; and Digit Span from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)[34] total score. Trail Making Test (D-KEFS) score is a measure that reflects processing speed, psychomotor speed, and cognitive flexibility[35]. Verbal Fluency (D-KEFS) score is a measure of phonological fluency in verbal modality[36]. Digit span (WAIS-IV) total score is a measure of verbal working memory[37]. All raw scores were converted to age-adjusted scaled scores based on population means[38]. The neurocognitive outcomes were transformed into dichotomous variables and studied accordingly. For each of these variables, scores lower than one and a half standard deviations below the mean of the normative dataset were indicative of impairment[39], all other scores were considered non-impaired.

3.3.4. Emotional Distress: Anxiety and Depression

Participants were classified having emotional distress if they demonstrated elevated symptoms according to two measures referenced to age-specific norms. This was done in line with published recommendations[40,41] and previous use of the instruments[21,22,26]. For younger participants (<19 years), we used anxiety and depression modules of the Beck Youth Inventory (BYI), a self-report instrument to document psychological status in children from 7 to 18 years old[41]. For older participants (≥19 years) we used the Brief Symptom Inventory-18 (BSI-18 anxiety and depression score), an 18-item self-report questionnaire, assessing psychological distress in adults[40], previously also used in cohorts of young

and older adult survivors of childhood cancer[42,43]. Internal consistency coefficients measured by Cronbach's alphas were all satisfactory, >.80[44]. Age-adjusted scores one standard deviation above the population mean were considered as impaired.

3.3.5. Sequencing and quality control

Whole-exome sequencing (WES) was performed on germline DNA, extracted from peripheral blood samples from a subset of 191 participants of PETALE cohort, using standard protocols as described previously[6]. Whole exomes were captured in solution with Agilent's SureSelect Human All Exon 50Mb kits and sequenced on either Life Technologies SOLiD System 4.0 (mean coverage = 40X) or Illumina HiSeq 2500 platform (mean coverage = 113.1X) at SJUHC integrated clinical genomic centre in pediatrics. Reads were aligned to the hg19 reference genome using SOLiD LifeScope software[45] for the SOLiD samples and BWA-MEM[46] for the samples sequenced on the Illumina system. PICARD[47,48] was used to mark PCR duplicates and collect sequencing quality control metrics. Variant calling was performed using the Haplotype Caller and quality score recalibration was performed using Variant Recalibrator, both implemented in the Genome Analysis Tool Kit (GATK)[48]. Variants were selected based on the variant quality score (VQSR = PASS) and minimum depth of coverage (DP >= 10). The final germline variants were annotated by ANNOVAR[49] and the predicted functional impact of missense, nonsense and splicing common and rare variants was assessed in silico using Sift (<0.1) and PolyPhen2 (≥0.85) filters[50,51]. Variants were defined as rare (minor allele frequency, MAF<5%) and common (MAF≥5%) according to the reported frequency for European populations in the 1000 Genomes[52] and ESP6500 datasets[53]. These variants were considered as potentially damaging and were used for analyses. Variants exceeding missing rate of 20%, with minor allele count<2 and not in Hardy-Weinberg Equilibrium (P<0.001) were excluded.

3.3.6. Association analyses

Two sets of candidate genes were selected - the first consisted of 99 genes implicated in the metabolism of methotrexate (MTX) and corticosteroids (CS) which are known to impact neurocognitive outcomes[54,55] and the second consisted of 360 genes implicated in nervous system function, selected using the KEGG PATHWAY Database[56]. A total of 76 common variants (27 in MTX/CS pathway and 49 in nervous system function) and 1337 rare variants that satisfied all above filtering criteria were identified as functionally predicted and were used in association analyses. The analyses between common genetic variants and neurocognitive outcomes as well as with anxiety/depression were performed by the allelic chi-square or Fisher's exact test implemented in PLINK v.1.07[57,58]. Analyses were performed in 191 sequenced patients and stratified by sex, risk groups with different treatment intensity, and treatment with chemotherapy alone or chemotherapy and cranial radiation; since these factors have an established role in modulating neurocognitive outcomes[4,59]. The Benjamini-Hochberg procedure for false discovery rate (FDR)[60,61] was used to adjust for multiple testing with a cut-off value of < 5% considered statistically significant. Selective genotyping of top-ranking common SNPs (based additionally on Bonferroni p-value corrected for the number of variants tested, $p<0.001$ and $p<0.0019$ for the neural and MTX/CS pathways, respectively) was carried out on the Sequenom platform at the McGill University and Génome Québec Innovation Centre, Montreal, (Quebec), Canada, to confirm the results and extend the analysis to entire cohort (n=236) including one hundred ninety-one patients analysed above (**S1 Table**). Associations of genotyped variants with the outcomes were assessed using chi-square or Fisher exact test in SPSS v.24.0.0.0 and appropriate genetic models, which were presented relative to the minor allele. Genotype-outcome association was represented as an odds ratio (OR) with a 95% confidence interval (CI). For the variants of MTX pathway, for which the association showed similar trend in validation cohort, the modulation of the effect by cumulative drug dose was also analyzed. To that end, cumulative drug dose was dichotomized to above and below the median

and the association was analyzed in each subgroup. Additionally, logistic regression model was used in which main effect (genotype and drug dose) and interaction term were added. The detailed list of the studied polymorphisms **(DOI: 10.6084/m9.figshare.8051573)**, as well as the summary statistics for all polymorphisms analyzed from the sequencing data beyond those already presented in the regular and supplemental tables **(DOI: 10.6084/m9.figshare.8051825)** are provided.

For rare variants associations, we used the SKAT-O test (Optimal Sequence Kernel Association Test)[62,63] implemented in SKAT package v.1.3.2.1[64] with FDR < 5% considered statistically significant. Collapsing approach that combines several rare variants into a single variable[65,66], with iterative exclusion of each single variant, was additionally performed to allow weighting variant contributions to association signals. These analyses were performed as exploratory and associated variants were not further analyzed by genotyping.

3.3.7. Replication analysis

Genotype data for selected variants were obtained from a larger effort to sequence whole genomes of over three thousand long-term survivors participating in the SJLIFE cohort. For this replication analysis, we restricted inclusion to 545 ALL survivors of European ancestry. Associations of selected variants with respective neurocognitive outcomes were examined using chi-square or Fisher's exact tests, as appropriate, implemented in PLINK 1.9[57,58].

3.4. RESULTS

3.4.1. Neurocognitive and emotional disturbances

The median age of ALL survivors at the time of evaluation was 21 years, with almost equal sex distribution, their demographics and clinical characteristics are presented in **Table 1**. The most prevalent deficit in neurocognitive test performance was noted for digit span (19.5%) followed by verbal fluency (18.6%) and trail making test (9.3%). Moderate-severe anxiety was noted in 10.1% survivors, whereas 11.5% of survivors were affected by moderate-severe depression, which was comparable to published normative groups on anxiety and depression[40-42].

3.4.2. Common variants

Among common variants implicated in nervous system function obtained from WES data, significant associations were detected for four of them (*CALML5, CACNB2*, *PCDHB10* and *EPHA5)* either in all survivors or following stratification according to sex, risk groups or CRT (**S2 Table**). These variants were further analyzed by genotyping in the entire PETALE cohort, and the association was confirmed for all of them (**Table 2**). The analyses were performed for the same subgroups for which association was noted for WES data, and additionally in all participants. The neurocognitive deficit related to digit span task was associated, in an additive manner, with the minor allele of *rs58225473* in *CACNB2* gene either in all patients (p=0.02), or those who received chemotherapy only (p=0.004). Homozygotes for the minor C allele of *CALML5 rs10904516* were at higher risk of having deficit in verbal fluency score, whereas the neurocognitive deficit related to trail making test was associated with the minor allele of *rs2907323* in *PCDHB10* gene, both potentiated in HR participants (p=0.03 and p=0.01 respectively). The carriers of the minor C allele of *EPHA5 rs33932471* were at higher risk of both moderate-severe anxiety and depression, with the strongest effects seen in females (p=0.02 and p=0.003, respectively). Among common variants implicated in MTX/CS pathway obtained from WES data, the significant associations were detected for 6 of them (*MTR, PPARA*, *ABCC3, SHMT1* and *SLCO1B1,* (**S3 Table**).

The variants in *MTR, PPARA*, *ABCC3* and *SLCO1B1* genes were further analyzed by genotyping in entire PETALE cohort, and the association was confirmed for all of them (**Table 3**). The association between deficit in verbal fluency score and GG genotype of *MTR rs1805087* was seen for all survivors (p=0.01) and male participants (p=0.002). Deficits in verbal fluency performance were also associated with GG genotype of *ABCC3 rs12604031* among HR patients (p=0.001), as well as with *rs1800206* in *PPARA* gene in low-risk groups (p=0.008). The risk of moderate-severe depression was highest among carriers of the minor G allele of *SLCO1B1 rs4149056* who received chemotherapy only (p=0.002). All variants found significantly associated with a tested outcome (except those initially confined to risk subgroup such as those in *ABCC3* and *PCDHB10* genes) were further analyzed for an association with respective outcomes in an independent cohort of ALL survivors (SJLIFE cohort) **(Table 4 and 5)**. Two associations were noticeable. The association of borderline significance between deficit in verbal fluency score and the minor allele of *MTR rs1805087* was seen in all survivors (OR=1.7; 95% CI, 1.0-2.8; p=0.05). The association between deficit in digit span score and GG genotype of *CACNB2 rs58225473* showed similar trend as in PETALE cohort for all participants (OR=3.7, 95% CI, 1.0-13.9), as well as for patients who received chemotherapy only (OR=3.8, 95% CI, 0.9-16.5), however they did not reach significance (p=0.08 and 0.09, respectively). The associations for *rs58225473* and *rs1805087* variants in *CACNB2* and *MTR* genes were significant for combined PETALE and SJLIFE cohorts (**S4 Table**). The association between deficit in digit span score and GG genotype of *CACNB2 rs58225473* was significant for all participants (OR=3.7; 95% CI, 1.25-11; p=0.01) and for patients who received chemotherapy only (OR=7.2, 95% CI, 2.1-25; p=0.0004). The association between deficit in verbal fluency score and the minor allele of MTR rs1805087 was seen in all survivors (OR=1.5; 95% CI, 1-2.3; p=0.04) and in male participants (OR=1.8; 95% CI, 1-3.1; p=0.04).

Given that the *MTR* belongs to the MTX pathway, we further explored whether the effect of *rs1805087* was modulated by cumulative MTX doses, for which such data were available in the discovery group (**Table 1**). The relationship with the deficit in verbal fluency score was particularly obvious in patients who received higher overall cumulative doses (**Fig 1**, p=0.01 for patients with cumulative doses above median vs. p=0.3, for patients with cumulative doses below median).

3.4.3. Rare variants

The analysis of functionally predicted rare variants in PETALE cohort led to the detection of an association between the deficit in trail making test score and rare variants enrichment in *SLCO2B1, HSPA4* and *GSTT1* genes (p=0.0002, p=0.004 and p=0.003, respectively, **Table 6**). Using the collapsing approach, we explored variant combinations that contributed to the observed association signal, identifying two variants in *GSTT1*, three in *HSPA4* and four in *SLCO2B1* gene. Replication analyses were not performed for these findings because information regarding these variants was not available in the SJLIFE cohort.

3.5. DISCUSSION

Functionally predicted germline common variants in *MTR, PPARA, SLCO1B1, ABCC3, CALML5, CACNB2* and *PCDHB10* genes were found to be significantly associated with deficits in neurocognitive tests performance, whereas a variant in *EPHA5* gene was significantly associated with both anxiety and depression.

3.5.1. Neurocognitive performance

Among common variants associated with an impairment in neurocognitive function, *rs10904516* in the *MTR* gene, which was associated with a deficit in verbal fluency, seems particularly interesting given similar observation in the SJLIFE cohort. The *MTR* gene encodes a B12 dependent methionine synthase involved in remethylation of homocysteine (Hcy), which is the crucial step in methionine production in all types of cells[67]. Mutations in the *MTR* gene, as well as severe deficiency of vitamin B12, could result in elevated concentration of Hcy in plasma and cerebrospinal fluid. Studies have shown that Hcy exerts a neurotoxic action and may participate in the mechanisms of neurodegeneration, such as excitotoxicity, oxidative stress, calcium accumulation, and apoptosis[68-70]. *MTR* gene is involved in the metabolic pathway of MTX. Administration of MTX was associated with acute and subacute neurotoxic effects; these detrimental effects may accumulate over time[69]. The detected common functional polymorphism (*rs1805087*) leading to Asp919Gly amino acid replacement in the *MTR* gene could affect enzymatic activity, thus increasing the level of Hcy[68,69,71]. Indeed, we have shown an interaction between *MTR rs1805087* and cumulative MTX dose in survivors with the deficit in verbal fluency score. This confirms previous finding: this variant together with polymorphisms of other genes that are implicated in the Hcy pathway were already studied in the context of MTX long-term neurotoxicity and has been found to affect neurocognitive function in childhood ALL survivors[72,73].

CACNB2 rs58225473 variant was associated with the neurocognitive deficit as defined by the digit span test, which measures working memory. Similar risk values, although not significant, were noted in the SJLIFE cohort in all participants and in the group of survivors who received chemotherapy only. The *CACNB2* gene encodes an auxiliary voltage-dependent subunit of L-type calcium-channel that is mainly expressed in brain and heart tissue. Voltage-dependent calcium channels are crucial for neuronal differentiation and maturation. They induce large number of intracellular events such as neurotransmitter release, neuronal excitability, synaptic plasticity, and gene regulation[74]. Calcium influx mediated by those channels has both spatial and temporal components and encodes important signaling information[75]. Moreover, in a recent GWAS study *CACNB2* was identified among four significant risk loci underlying genetic effects shared between five major psychiatric disorders that included schizophrenia, autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar and major depressive disorders[76]. Additionally, the rare variants in this gene were found in affected members of families with autism spectrum disease[77]. Given the important role that *CACNB2* can play, it is not surprising that it was studied as a possible pharmacological target in treatment of mental disorders[78]. The *rs58225473* is a c.1803T>G substitution (NM_201590.2) leading to Asp601Glu replacement, which is predicted to affect protein function and possibly calcium channel function. Common variants in several other genes influenced neurocognitive decline in PETALE cohort. *PPARA* gene belongs to PPARs receptor family of ligand-activated transcription factors involved in the regulation of inflammation[79]. Effects of glucocorticoids can be reinforced by PPAR ligands[80]. The enhanced heterodimer formations of PPARA could be associated with increased expression of brain and glial cell-derived neurotrophic factors[81]. The *rs1800206* variant in *PPARA* gene was associated with lower verbal fluency score in females and survivors assigned to SR group or chemotherapy only. Similar association was noted for *ABCC3 rs12604031*. *ABCC3* is a member of the superfamily of ATP-

binding cassette (ABC) transporters, and the bioavailability of MTX may be affected by this transporter[82].

CALML5 gene is related to the calmodulin family of calcium binding proteins highly implicated in CNS function[83]. Its protective role and implication in the inhibition of neuronal death was described in Alzheimer disease[84,85]. We observed the significant association of variant *rs10904516* and deficit in verbal fluency score in all survivors, with the stronger effect seen in the HR group. Similar association mostly confined to HR group was noted between lower score on trial making test and *rs2907323* in the *PCDHB10* gene. *PCDH* (protocadherin) genes, are expressed in the central and peripheral nervous systems and are required for their normal development. They mediate a variety of processes, including neuronal survival, morphogenesis and connectivity, synaptic maintenance, and spatial patterning of axons and dendrites[86]. The variants in *PCDH* genes have been reported to be associated with dyslexia and bipolar disorder[87,88].

The neurotoxic effects of treatments in childhood ALL have been the subject of multiple investigations. These effects consist of central neurotoxicity clearly noticeable by encephalopathy and/or neurodevelopmental cognitive deficits[89-91], particularly in survivors exposed to a highly intensified treatment protocols with CNS-directed chemotherapy, even in the absence of CRT[4,19,92,93]. Cognitive impairment and information processing have been associated with intensity and duration of CS treatment[94,95]. Female survivors were reported to have more severe short-term memory impairment[59] and lower scores on attentional indices, cognitive flexibility[96,97] and visuomotor control[4]. Female childhood ALL survivors are more likely to present cerebral white matter damage[98] that may affect cognitive functioning. Congruent with these previous observations, several associations detected in the present study were modulated by sex and treatment intensity (reflected by the presence or not of CRT or risk groups).

3.5.2. Anxiety and depression

We studied dimensions of internalized symptoms which are frequent in normative populations, namely anxiety and depression[99], highlighting impairment in mental quality of life of childhood ALL survivors[21,22,100]. Although anxiety and depression measures are not equivalent to clinical diagnosis derived from the gold standard systematic interview[101], moderate-severe levels are generally interpreted as a risk for having clinically relevant anxiety or depression.

The *rs11556218* in the *EPHA5* gene was associated with higher risk of both anxiety and depression that was further potentiated in female patients. The *EPHA5* gene codes for a brain-specific kinase that is selectively expressed in a subset of serotonin neurons during embryonic and postnatal development[102]. Receptors in the EPH subfamily modify the strength of existing synapses in the adult brain[103]. Divergent vulnerabilities between females and males could be explained by gender differences in brain maturation[104,105], which might make females more vulnerable to the neurotoxic effects of chemotherapy. Other assumptions, such as endocrine factors, have also been hypothesized to explain sex differences in the susceptibility[59].

Moreover, we identified the association between depression and the presence of variant *rs4149056* in *SLCO1B1* gene. This association was detected in the group of survivors that received chemotherapy without CRT. *SLCO1B1* gene encodes a liver-specific member of the organic anion transporter family involved in hepatic uptake of MTX. This association deserved further attention given that the same variant was detected through genome wide association studies to contribute to inter-individual variability in the clearance of high-dose MTX[106]. It was subsequently replicated in independent cohorts and shown also as a predictor of short-term toxicity following MTX treatment[106-111]. MTX treatment has been associated with adverse emotional or behavioral outcomes[20]; thus, these results could justify further studies of *SLCO1B1* gene in related contexts.

3.5.3. Rare variants' analysis

The association between deficits in the trail making test score was identified in relation to rare variants enrichment in *HSPA4, SLCO2B1* and *GSTT1* genes, with a very strong individual contribution of *rs61745470* in *HSPA4* gene. This variant was recently associated with familial genetic risk for suicide (as well as with risk for psychiatric or substance abuse conditions)[112]. The *SLCO2B1* and *GSTT1* genes are highly implicated in physiological and pharmacological distribution of drugs and endogenous molecules. The *SLCO2B1,* a member of the organic anion transporting polypeptide (OATP) family, is involved in steroid hormone uptake and transport of steroid conjugates[113,114]. *GSTT1* was recently associated with higher risk for early onset of severe mental and bipolar disorders[115]. We also evaluated the association of deletion polymorphisms of *GSTT1*[116] (found with a frequency of 23.7% in discovery cohort) with the deficits in the trail making test score. There was no association of *GSTT1* null genotype with the deficits in the trail making test score.

The impact of here identified rare functional variants requires further investigation.

3.6. CONCLUDING REMARKS

Our study has certain limitations. Its limited sample size may affect the accuracy of the results, particularly in the context of the stratified analysis. Other unmeasured factors in this study, for example, inflammation and oxidative stress, could modulate or potentiate associations with genetic factors. The candidate gene approach may have missed genetic markers potentially involved in neurocognitive decline and mood disturbances that could have been detected through unbiased approaches. Among associations detected in the PETALE cohort only two showed a similar trend in SJLIFE cohort. Despite matching both outcomes and patients' characteristics between the two cohorts, it is possible that small sample size, differences in treatment protocols or time of ALL diagnosis [117-120] contributed to the observed discrepancies. Likewise, stratification by risk group designation was not available for the SJLIFE cohort, precluding replication of the risk-based stratified analyses. Although the analyses in PETALE cohort were corrected for multiple testing, and confounding was reduced due to homogeneous population and uniform treatment, we cannot exclude that some of the associations have been obtained by chance.

In conclusion, using a comprehensive candidate gene approach and whole exome sequencing data we identified a panel of functionally predicted genetic variants significantly associated with neurocognitive deficits, anxiety, and depression in childhood ALL survivors. Additional exome wide analysis might lead to the discovery of novel genes and genetic variants associated with neurocognitive LAEs as well as with the mood disorders.

While we acknowledge that the identified germline variants still need to be evaluated and validated through replication and functional studies, the current findings can help further understanding of the influence of genetic component on long-term complications related to cancer therapy.

3.7. LIST OF ABBREVIATIONS

ALL: Acute lymphoblastic leukemia
PETALE: **P**révenir les **E**ffets tardifs des **T**raitements de la leucémie **A**iguë **L**ymphoblastique chez l'**E**nfant
SJUHC: Sainte-Justine University Health Center
LAEs: Late adverse effects
WES: Whole exome sequencing
DFCI: Dana-Faber Cancer Institute
CCSS: Childhood Cancer Survivors Study
SJLIFE: St-Jude Lifetime cohort
CRT: Cranial radiation therapy
D-KEFS: Delis-Kaplan Executive Function System
WAIS-IV: Wechsler Adult Intelligence Scale-Fourth Edition
BYI: Beck Youth Inventory
BSI-18: Brief Symptom Inventory-18
MAF: Minor allele frequency
MTX: methotrexate
CS: corticosteroids
SKAT-O test: Optimal Sequence Kernel Association Test
FDR: False discovery rate
SNP: Single nucleotide polymorphism
SR: standard risk
HR: high risk
OR: odds ratio
CI: confidence interval
CNS: central nervous system
OATP: organic anion transporting polypeptide

3.8. DECLARATIONS

Ethics approval and consent to participate

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of SJUHC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon request and revision of the projects for which the data might be used.

Competing interests

The authors declare no competing financial interests.

Funding

Institute of Cancer Research (ICR) of the Canadian Institutes of Health Research (CIHR) **grant number: 118694** awarded to Dr Daniel Sinnett, Dr Maja Krajinovic, Dr Caroline Laverdière, Dr Philippe Robaey, Dr Sarah Lippé, Dr Serge Sultan (http://www.cihr-irsc.gc.ca); C17 Council, Cancer Research Society (CRS), Canadian Cancer Society Research Institute (CCSRI), Ontario Institute for Cancer Research (OICR), Pediatric Oncology Group of Ontario (POGO), Garron Family Cancer Centre at SickKids Hospital (Ontario), The Terry Fox foundation, **grant number: TFF - 105266** awarded to Dr Daniel Sinnett and Dr Maja Krajinovic (https://www.terryfox.org/),

FRQS Applied Medical Genetics Network and Sainte-Justine Hospital Foundation supported the PETALE study.

Acknowledgements

The authors would like to thank all childhood ALL survivors and their parents for the participation in the PETALE study, as well as all study collaborators for their valuable contribution.

3.9. TABLES AND FIGURES

3.9.1. Table 1. Patient demographics and clinical characteristics (N=236).

	N	%
Sex		
Male	115	48.7
Female	121	51.3
Neuropsychological outcomes (affected patients)		
Trial making test*	22	9.3
Verbal fluency*	44	18.6
Digit span*	46	19.5
Moderate-severe anxiety	21	10.1
Moderate-severe depression	24	11.5
DFCI protocol		
87-01	18	7.6
91-01	48	20.3
95-01	71	30.1
00-01	75	31.8
05-01	24	10.2
Prognostic risk group		
Standard risk (SR)	110	46.6
High risk (HR)	126	53.4
Cranial radiation therapy		
Yes**	131	55.5
No	105	44.5
Cumulative doses, mg/m2 - median (range)		
Parenteral/PO MTX (853.6-12750.5)	6576.5	
IT MTX (0-279)	150.4	
IV/PO corticosteroids*** (4425.7-24930.1)	8826.6	
IT corticosteroids*** (2.05-87.19)	31.3	
Age at diagnosis - median (range)		
Male (1-18)	5.0	
Female (0-18)	4.0	
Age at follow - up median (range)		
Male (12-36)	21.0	
Female (12-38)	21.0	

DFCI, Dana-Farber Cancer Institute; IV, intravenous; PO, per os; IT, intrathecal; MTX, methotrexate
*Score at least 1.5 standard deviation below the norm was considered as impaired in all neuropsychological tests
** Median (range), 12 Gy (12-18Gy)
*** Cumulative corticosteroid doses are calculated as prednisone equivalents

3.9.2. Table 2. Frequency of associated genotypes in patients with and without neurocognitive or emotional distress, genes of relevance for nervous system function, PETALE cohort (N=236).

Outcome	Genotype	Case*, N (%)	Control*, N (%)	Model	Case*, N (%)	Control*, N (%)	P**	OR (95%-CI)
				CACNB2 rs58225473				
				All patients				
Digit span	**TT**	25 (58.1)	128 (71.5)	**TT**	25 (58.1)	128 (71.5)	0.02[a]	2.0 (1.1-3.9)
	TG	16 (37.2)	51 (28.5)	**TG**	16 (37.2)	51 (28.5)		
	GG	2 (4.7)	0 (0.0)	**GG**	2 (4.7)	0 (0.0)		
				Chemotherapy only* **				
	TT	4 (40.0)	63 (70.8)	**TT**	4 (40.0)	63 (70.8)	0.004[a]	5.0 (1.5-16.4)
	TG	4 (40.0)	26 (29.2)	**TG**	4 (40.0)	26 (29.2)		
	GG	2 (20.0)	0 (0.0)	**GG**	2 (20.0)	0 (0.0)		
				PCDHB10 rs2907323				
				All patients				
Trial making test	**GG**	11 (55.0)	155 (73.1)	**GG**	11 (55.0)	155 (73.1)	0.02[a]	2.5 (1.1-6.2)
	GC	8 (40.0)	57 (26.9)	**GC**	8 (40.0)	57 (26.9)		
	CC	1 (5.0)	0 (0.0)	**CC**	1 (5.0)	0 (0.0)		
				High risk				
	GG	6 (46.2)	85 (76.6)	**GG**	6 (46.2)	85 (76.6)	0.01[a]	4.3 (1.4 -12.7)
	GC	6 (46.2)	26 (23.4)	**GC**	6 (46.2)	26 (23.4)		
	CC	1 (7.7)	0 (0.0)	**CC**	1 (7.7)	0 (0.0)		
				CALML5 rs10904516				
				All patients				
Verbal fluency	**TT**	18 (40.9)	93 (48.7)	**TT+TC**	37 (84.1)	178 (93.2)	0.05[r]	2.6 (1.0-6.9)
	TC	19 (43.2)	85 (44.5)					
	CC	7 (15.9)	13 (6.8)	**CC**	7 (15.9)	13 (6.8)		
				High risk				
	TT	8 (34.8)	52 (50.5)	**TT+TC**	18 (78.3)	96 (93.2)	0.03[r]	3.8 (1.1-13.3)
	TC	10 (43.5)	44 (42.7)					
	CC	5 (21.7)	7 (6.8)	**CC**	5 (21.7)	7 (6.8)		
				EPHA5 rs33932471				
Moderate-severe anxiety				**Females**				
	AA	5 (55.6)	82 (89.1)	**AA**	5 (55.6)	82 (89.1)	0.02[d]	6.6 (1.5-28.5)
	AC	4 (44.4)	8 (8.7)	**AC+CC**	4 (44.4)	10 (10.9)		
	CC	0 (0.0)	2 (2.2)					
				EPHA5 rs33932471				
Moderate-severe depression				**All patients**				
	AA	15 (78.9)	151 (88.8)	**AA**	15 (71.4)	151 (88.8)	0.03[d]	3.2 (1.1-9.2)
	AC	6 (28.6)	17 (10.0)	**AC+CC**	6 (28.6)	19 (11.2)		
	CC	0 (0.0)	2 (1.2)					

Females							
AA	7 (58.3)	80 (89.9)	**AA**	7 (58.3)	80 (89.9)	0.003[d]	6.3 (1.7-24.2)
AC	5 (41.7)	7 (7.9)	**AC+CC**	5 (41.7)	9 (10.1)		
CC	0 (0.0)	2 (2.2)					

CACNB2: Calcium Voltage-Gated Channel Auxiliary Subunit Beta2; *CALML5*: Calmodulin Like 5, *EPHA5*: EPH Receptor A5, Brain-Specific Kinase; OR, odds ratio; CI, confidence interval.
*Participants with and without indicated complications are defined as cases and controls, respectively.
**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (a: Additive; d: Dominant, r: Recessive).
***Chemotherapy without cranial radiation therapy.

3.9.3. Table 3. Frequency of associated genotypes in patients with and without neurocognitive or emotional distress, genes implicated in methotrexate and corticosteroids pathways, PETALE cohort (N=236).

Outcome	Genotype	Case*, N (%)	Control*, N (%)	Model	Case*, N (%)	Control*, N (%)	P**	OR (95%-CI)
				MTR rs1805087				
				All patients				
	AA	25 (61.0)	120 (66.7)	**AA+AG**	37 (90.2)	178 (98.9)	0.01[r]	9.6 (1.7-54.5)
	AG	12 (29.2)	58 (32.2)					
	GG	4 (9.8)	2 (1.1)	**GG**	4 (9.8)	2 (1.1)		
				Males				
	AA	12 (52.2)	57 (66.3)	**AA+AG**	19 (82.6)	86 (100.0)	0.002[r]	
	AG	7 (30.4)	29 (33.7)					
	GG	4 (17.4)	0 (0.0)	**GG**	4 (17.4)	0 (0.0)		
				PPARA rs1800206				
Verbal fluency				**Standard risk**				
	CC	13 (72.7)	76 (92.7)	**CC**	13 (72.7)	76 (92.7)	0.008[a]	4.6 (1.5-14.5)
	CG	3 (16.7)	6 (7.3)	**CG**	3 (16.7)	6 (7.3)		
	GG	2 (11.1)	0 (0.0)	**GG**	2 (11.1)	0 (0.0)		
				Chemotherapy only				
	CC	12 (75.0)	77 (92.8)	**CC**	12 (75.0)	77 (92.8)	0.02[a]	4.3 (1.3-13.5)
	CG	2 (12.5)	6 (7.2)	**CG**	2 (12.5)	6 (7.2)		
	GG	2 (12.5)	0 (0.0)	**GG**	2 (12.5)	0 (0.0)		
				ABCC3 rs12604031				
				High risk				
	AA	3 (13.0)	33 (33.3)	**AA+AG**	14 (60.9)	89 (89.9)	0.001[r]	5.7 (2.0-16.5)
	AG	11 (47.8)	56 (56.6)					
	GG	9 (39.1)	10 (10.1)	**GG**	9 (39.1)	10 (10.1)		
				SLCO1B1 rs4149056				
Moderate-severe depression				**Chemotherapy only**				
	AA	1 (14.3)	59 (75.6	**AA**	1 (14.3)	59 (76.5)	0.002[d]	18.6 (2.1-164.7)
	AG	5 (71.4)	19 (24.4)	**AG+GG**	6 (85.7)	19 (24.4)		
	GG	1 (14.3)	0 (0.0)					

MTR: 5-Methyltetrahydrofolate-Homocysteine Methyltransferase, *PPARA*: Peroxisome Proliferator Activated Receptor Alpha, *ABCC3*: ATP Binding Cassette Subfamily C Member 3, *SLCO1B1*: Solute Carrier Organic Anion Transporter Family Member 1B1.
*Participants with and without indicated complications are defined as cases and controls, respectively.
**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (a: Additive; d: Dominant, r: Recessive).

3.9.4. Table 4. Frequency of genotypes in patients with and without neurocognitive or emotional distress, genes of relevance for nervous system function, replication cohort of SJLIFE (N=545).

Outcome	Genotype	Case*N (%)	Control* N (%)	Model	Case* N (%)	Control* N (%)	P**	OR (95%-CI)
				CACNB2 rs58225473				
				All patients				
Digit span	**TT**	23 (71.8)	279 (69.1)	**TT+TG**	29 (90.6)	393 (97.3)	0.08[r]	3.7 (1.0-13.9)
	TG	6 (18.8)	114 (28.2)					
	GG	3 (9.4)	11 (2.7)	**GG**	3 (9.4)	11 (2.7)		
				Chemotherapy only				
	TT	16 (66.7)	112 (67.1)	**TT+TG**	21 (87.5)	161 (96.4)	0.09[r]	3.8 (0.9-16.5)
	TG	5 (20.8)	49 (29.3)					
	GG	3 (12.5)	6 (3.6)	**GG**	3 (12.5)	6 (3.6)		
				CALML5 rs10904516				
				All patients				
Verbal fluency	**TT**	32 (45.7)	180 (40.0)	**TT+TC**	62 (88.6)	387 (86.0)	0.6[r]	0.8 (0.4-1.7)
	TC	30 (42.9)	207 (46.0)					
	CC	8 (11.4)	63 (14.0)	**CC**	8 (11.4)	63 (14.0)		
				EPHA5 rs33932471				
				Females				
Moderate-severe anxiety	**AA**	27 (84.4)	205 (87.2)	**AA**	27 (84.4)	205 (87.2)	0.7[d]	1.3 (0.5-3.5)
	AC	5 (15.6)	27 (11.5)	**AC+CC**	5 (15.6)	30 (12.8)		
	CC	0 (0.0)	3 (1.3)					
				EPHA5 rs33932471				
				All patients				
Moderate-severe depression	**AA**	58 (79.5)	399 (86.0)	**AA**	58 (79.5)	399 (86.0)	0.1[d]	1.6 (0.8-3.0)
	AC	15 (20.5)	60 (12.9)	**AC+CC**	15 (20.5)	65 (14.0)		
	CC	0 (0.0)	5 (1.1)					
				Females				
	AA	30 (90.9)	202 (86.3)	**AA**	30 (9.9)	202 (86.3)	0.5[d]	0.6 (0.2-2.2)
	AC	3 (9.1)	29 (12.4)	**AC+CC**	3 (9.1)	32 (13.7)		
	CC	0 (0.0)	3 (1.3)					

*Participants with and without indicated complications are defined as cases and controls, respectively;
**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (d: Dominant, r: Recessive).

3.9.5. Table 5. Frequency of genotypes in patients with and without neurocognitive or emotional distress, genes implicated in methotrexate and corticosteroids pathways, replication cohort of SJLIFE (N=545).

Outcome	Genotype	Case* N (%)	Control* N (%)	Model	Case* N (%)	Control* N (%)	P**	OR (95%-CI)
				MTR rs1805087				
				All patients				
	AA	42 (59.2)	323 (70.7)					
	AG	25(35.2)	121 (26.5)	**AG+GG**	29 (40.8)	134 (29.3)	0.05[d]	1.7(1.0-2.8)
	GG	4 (5.6)	13 (2.8)	**AA**	42 (59.2)	323 (70.7)		
				Males				
Verbal fluency	**AA**	23 (60.5)	164 (72.2)					
	AG	13 (34.2)	56 (24.7)	**AG+GG**	15 (39.5)	63 (27.8)	0.1[d]	1.7 (0.8-3.5)
	GG	2 (5.3)	7 (3.1)	**AA**	23 (60.5)	164 (72.2)		
				PPARA rs1800206				
				Chemotherapy only				
	CC	39 (81.2)	155 (85.6)	**CC**	39 (73.8)	155 (85.6)		
	CG	9 (18.8)	26 (14.4)	**CG+GG**	9 (23.8)	26 (14.4)	0.5[d]	1.4 (0.6-3.2)
	GG	0 (0)	0 (0)					
				SLCO1B1 rs4149056				
Moderate-severe depression				**Chemotherapy only**				
	AA	31 (73.8)	142 (72.4)	**AA**	31 (73.8)	142 (72.4)		
	AG	10 (23.8)	48 (24.5)	**AG+GG**	11 (26.2)	54 (27.6)	0.9[d]	0.9 (0.4-2.0)
	GG	1 (2.4)	6 (3.1)					

*Participants with and without indicated complications are defined as cases and controls, respectively;
**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (d: Dominant).

3.9.6. Table 6. SKAT-O analysis of the rare functional variants, PETALE cohort, WES data, demonstrated for the deficit in trail making test scores (N=191).

Outcome	Gene	SNPs tested		MAF	P value	FDR	OR[1] [95% CI]
		position	rs number				
Trail making test	***SLCO2B1***	chr11:74875089-74875089		0.009	0.0002	0.004	8.7 [1.3-57.0]
		chr11:74880370-74880370	*rs35199625*	0.008			
		chr11:74899276-74899276	*rs377133671*	0.003			
	HSPA4	***chr5:132387979-132387979**	***rs61745470***	0.017	0.004	0.027	*7.8 [1.7-36.4]
		chr5:132408967-132408967	*rs61755724*	0.042			
		*chr5:132412511-132412511		0.003			
		*chr5:132437499-132437499	*rs61749631*	0.005			
	GSTT1	chr22:24379402-24379402	*rs11550605*	0.003	0.003	0.027	19.7 [1.7-230.6]
		chr22:24381742-24381742		0.006			

SNP: single nucleotide polymorphism; OR: odds ratio; CI: confidence interval; *SLCO2B1*: Solute Carrier Organic Anion Transporter Family Member 2B1; *HSPA4*: Heat Shock Protein Family A (Hsp70) Member 4; *GSTT1*: Glutathione S-Transferase Theta 1.

[1] OR of significant combination, combination include all variants or variants noted with asterisk*.

Individual contribution of variant *rs61745470* (noted in bold) was identified with OR=19.5, 95% CI [2.97-128], p=0.005.

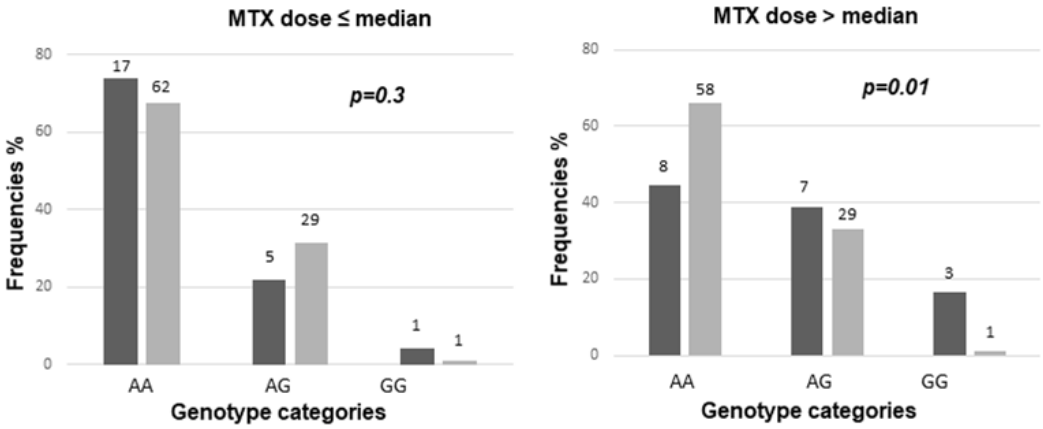


Figure 1. Interaction between MTR rs1805087 and cumulative methotrexate dose.

The frequency of each genotype in affected and non-affected group defined by verbal fluency score is presented by black and gray bars, respectively, in the groups that received cumulative MTX below median (left panel) or above median (right panel). The number of individuals represented by each bar and p values are indicted on the plot. OR for interaction is 3.3, 95% CI 0.9-11.5, p=0.07, as derived from logistic regression model including MTR genotype, cumulative MTX dose and interaction term.

3.10. REFERENCES

1. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatr Clin North Am.* 2015;62(1):47-60.
2. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin.* 2014;64(2):83-103.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66(1):7-30.
4. Buizer AI, de Sonneville LM, van den Heuvel-Eibrink MM, Veerman AJ. Chemotherapy and attentional dysfunction in survivors of childhood acute lymphoblastic leukemia: effect of treatment intensity. *Pediatr Blood Cancer.* 2005;45(3):281-290.
5. Hudson MM, Ness KK, Gurney JG, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. *JAMA.* 2013;309(22):2371-2381.
6. Marcoux S, Drouin S, Laverdiere C, et al. The PETALE study: Late adverse effects and biomarkers in childhood acute lymphoblastic leukemia survivors. *Pediatr Blood Cancer.* 2017;64(6).
7. Campbell LK, Scaduto M, Sharp W, et al. A meta-analysis of the neurocognitive sequelae of treatment for childhood acute lymphocytic leukemia. *Pediatr Blood Cancer.* 2007;49(1):65-73.
8. Ashford J, Schoffstall C, Reddick WE, et al. Attention and working memory abilities in children treated for acute lymphoblastic leukemia. *Cancer.* 2010;116(19):4638-4645.
9. Iyer NS, Balsamo LM, Bracken MB, Kadan-Lottick NS. Chemotherapy-only treatment effects on long-term neurocognitive functioning in childhood ALL survivors: a review and meta-analysis. *Blood.* 2015;126(3):346-353.
10. Kingma A, van Dommelen RI, Mooyaart EL, Wilmink JT, Deelman BG, Kamps WA. Slight cognitive impairment and magnetic resonance imaging abnormalities but normal school levels in children treated for acute lymphoblastic leukemia with chemotherapy only. *J Pediatr.* 2001;139(3):413-420.
11. Lofstad GE, Reinfjell T, Hestad K, Diseth TH. Cognitive outcome in children and adolescents treated for acute lymphoblastic leukaemia with chemotherapy only. *Acta Paediatr.* 2009;98(1):180-186.
12. Rodgers J, Marckus R, Kearns P, Windebank K. Attentional ability among survivors of leukaemia treated without cranial irradiation. *Arch Dis Child.* 2003;88(2):147-150.
13. Daams M, Schuitema I, van Dijk BW, et al. Long-term effects of cranial irradiation and intrathecal chemotherapy in treatment of childhood leukemia: a MEG study of power spectrum and correlated cognitive dysfunction. *BMC Neurol.* 2012;12:84.
14. Edelmann MN, Krull KR, Liu W, et al. Diffusion tensor imaging and neurocognition in survivors of childhood acute lymphoblastic leukaemia. *Brain.* 2014;137(Pt 11):2973-2983.
15. Mennes M, Stiers P, Vandenbussche E, et al. Attention and information processing in survivors of childhood acute lymphoblastic leukemia treated with chemotherapy only. *Pediatr Blood Cancer.* 2005;44(5):478-486.
16. Carey ME, Haut MW, Reminger SL, Hutter JJ, Theilmann R, Kaemingk KL. Reduced frontal white matter volume in long-term childhood leukemia survivors: a voxel-based morphometry study. *AJNR Am J Neuroradiol.* 2008;29(4):792-797.
17. Cole PD, Kamen BA. Delayed neurotoxicity associated with therapy for children with acute lymphoblastic leukemia. *Ment Retard Dev Disabil Res Rev.* 2006;12(3):174-183.

18. Krull KR, Brinkman TM, Li C, et al. Neurocognitive outcomes decades after treatment for childhood acute lymphoblastic leukemia: a report from the St Jude lifetime cohort study. *J Clin Oncol.* 2013;31(35):4407-4415.
19. Cheung YT, Krull KR. Neurocognitive outcomes in long-term survivors of childhood acute lymphoblastic leukemia treated on contemporary treatment protocols: A systematic review. *Neurosci Biobehav Rev.* 2015;53:108-120.
20. Schultz KA, Ness KK, Whitton J, et al. Behavioral and social outcomes in adolescent survivors of childhood cancer: a report from the childhood cancer survivor study. *J Clin Oncol.* 2007;25(24):3649-3656.
21. Zeltzer LK, Recklitis C, Buchbinder D, et al. Psychological status in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. *J Clin Oncol.* 2009;27(14):2396-2404.
22. Michel G, Rebholz CE, von der Weid NX, Bergstraesser E, Kuehni CE. Psychological distress in adult survivors of childhood cancer: the Swiss Childhood Cancer Survivor study. *J Clin Oncol.* 2010;28(10):1740-1748.
23. Stuber ML, Meeske KA, Krull KR, et al. Prevalence and predictors of posttraumatic stress disorder in adult survivors of childhood cancer. *Pediatrics.* 2010;125(5):e1124-1134.
24. Baytan B, Asut C, Cirpan Kantarcioglu A, Sezgin Evim M, Gunes AM. Health-Related Quality of Life, Depression, Anxiety, and Self-Image in Acute Lymphocytic Leukemia Survivors. *Turk J Haematol.* 2016;33(4):326-330.
25. Pepin AJ, Lippe S, Krajinovic M, et al. How to interpret high levels of distress when using the Distress Thermometer in the long-term follow-up clinic? A study with Acute Lymphoblastic Leukemia survivors. *Pediatr Hematol Oncol.* 2017;34(3):133-137.
26. Brinkman TM, Zhu L, Zeltzer LK, et al. Longitudinal patterns of psychological distress in adult survivors of childhood cancer. *Br J Cancer.* 2013;109(5):1373-1381.
27. Bitsko MJ, Cohen D, Dillon R, Harvey J, Krull K, Klosky JL. Psychosocial Late Effects in Pediatric Cancer Survivors: A Report From the Children's Oncology Group. *Pediatr Blood Cancer.* 2016;63(2):337-343.
28. Moghrabi A, Levy DE, Asselin B, et al. Results of the Dana-Farber Cancer Institute ALL Consortium Protocol 95-01 for children with acute lymphoblastic leukemia. *Blood.* 2007;109(3):896-904.
29. Silverman LB, Stevenson KE, O'Brien JE, et al. Long-term results of Dana-Farber Cancer Institute ALL Consortium protocols for children with newly diagnosed acute lymphoblastic leukemia (1985-2000). *Leukemia.* 2010;24(2):320-334.
30. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 2012;28(24):3326-3328.
31. PLINK: Whole genome data analysis toolset. http://zzz.bwh.harvard.edu/plink/index.shtml. Published 2018. Accessed.
32. Krull KR, Okcu MF, Potter B, et al. Screening for neurocognitive impairment in pediatric cancer long-term survivors. *J Clin Oncol.* 2008;26(25):4138-4143.
33. Delis DC KE, Kramer JH. . Delis-Kaplan executive function system (D-KEFS). *Psychological Corporation.* 2001.
34. D W. *Wechsler adult intelligence scale - Fourth Edition (WAIS-IV).* San Antonio, TX2008.

35. Yochim B, Baldo J, Nelson A, Delis DC. D-KEFS Trail Making Test performance in patients with lateral prefrontal cortex lesions. *J Int Neuropsychol Soc.* 2007;13(4):704-709.
36. Delis DC, Kramer JH, Kaplan E, Holdnack J. Reliability and validity of the Delis-Kaplan Executive Function System: an update. *J Int Neuropsychol Soc.* 2004;10(2):301-303.
37. Erdodi LA, Abeare CA, Lichtenstein JD, et al. Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychol Assess.* 2017;29(2):148-157.
38. Harrison AG, Armstrong IT, Harrison LE, Lange RT, Iverson GL. Comparing Canadian and American normative scores on the Wechsler Adult Intelligence Scale-Fourth Edition. *Arch Clin Neuropsychol.* 2014;29(8):737-746.
39. Boulet-Craig A, Robaey P, Laniel J, et al. DIVERGT screening procedure predicts general cognitive functioning in adult long-term survivors of pediatric acute lymphoblastic leukemia: A PETALE study. *Pediatr Blood Cancer.* 2018;65(9):e27259.
40. Derogatis LR. *Brief Symptom Inventory (BSI) 18: Administration, scoring, and procedures manual.* Minneapolis, MN: NCS Pearson; 2000.
41. Beck JS BA, Jolly JB, Steer RA. . *Beck Youth Inventories-Second Edition for children and adolescent's manual.* San Antonio, TX: PsychCorp; 2005.
42. Recklitis CJ, Parsons SK, Shih MC, Mertens A, Robison LL, Zeltzer L. Factor structure of the brief symptom inventory--18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. *Psychol Assess.* 2006;18(1):22-32.
43. Gianinazzi ME, Rueegg CS, Wengenroth L, et al. Adolescent survivors of childhood cancer: are they vulnerable for psychological distress? *Psychooncology.* 2013;22(9):2051-2058.
44. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80(1):99-103.
45. Spinella JF, Healy J, Saillour V, et al. Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. *BMC Cancer.* 2015;15:539.
46. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
47. Picard Tools - By Broad Institute. Code ZIP File. https://broadinstitute.github.io/picard/. Published 2018. Accessed.
48. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
49. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
50. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.
51. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 2006;7:61-80.
52. The 1000 Genomes Browsers | 1000 Genomes. http://www.internationalgenome.org/1000-genomes-browsers/. Published 2018. Accessed.
53. Exome Variant Server. http://evs.gs.washington.edu/EVS/. Published 2018. Accessed.
54. Mikkelsen TS, Thorn CF, Yang JJ, et al. PharmGKB summary: methotrexate pathway. *Pharmacogenetics and genomics.* 2011;21(10):679-686.

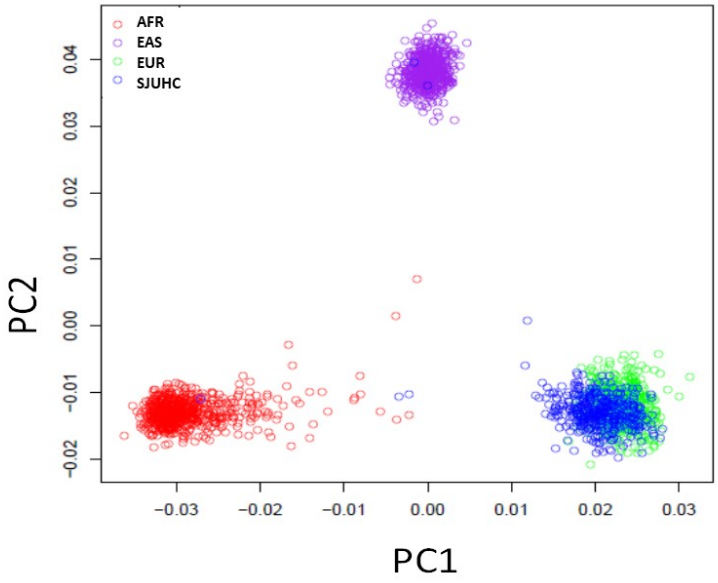
55. Schlossmacher G, Stevens A, White A. Glucocorticoid receptor-mediated apoptosis: mechanisms of resistance in cancer cells. *J Endocrinol.* 2011;211(1):17-25.
56. KEGG: Kyoto Encyclopedia of Genes and Genomes. https://www.genome.jp/kegg/. Published 2018. Accessed.
57. Whole genome association analysis toolset. http://zzz.bwh.harvard.edu/plink/index.shtml. Accessed.
58. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
59. Waber DP, Tarbell NJ, Kahn CM, Gelber RD, Sallan SE. The relationship of sex and treatment modality to neuropsychologic outcome in childhood acute lymphoblastic leukemia. *J Clin Oncol.* 1992;10(5):810-817.
60. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1-2):279-284.
61. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. *Genetics.* 2005;171(2):783-790.
62. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics.* 2012;91(2):224-237.
63. Wu MC, Kraft P, Epstein MP, et al. Powerful SNP-set analysis for case-control genome-wide association studies. *American journal of human genetics.* 2010;86(6):929-942.
64. CRAN - Package SKAT. https://cran.r-project.org/web/packages/SKAT/index.html. Published 2019. Accessed.
65. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet Epidemiol.* 2011;35 Suppl 1:S12-17.
66. Sung YJ, Korthauer KD, Swartz MD, Engelman CD. Methods for collapsing multiple rare variants in whole-genome sequence data. *Genet Epidemiol.* 2014;38 Suppl 1:S13-20.
67. Bellampalli R, Vohra M, Sharma K, et al. Acute lymphoblastic leukemia and genetic variations in BHMT gene: Case-control study and computational characterization. *Cancer Biomark.* 2017.
68. Rozycka A, Jagodzinski PP, Kozubski W, Lianeri M, Dorszewska J. Homocysteine Level and Mechanisms of Injury in Parkinson's Disease as Related to MTHFR, MTR, and MTHFD1 Genes Polymorphisms and L-Dopa Treatment. *Curr Genomics.* 2013;14(8):534-542.
69. Robaey P, Krajinovic M, Marcoux S, Moghrabi A. Pharmacogenetics of the neurodevelopmental impact of anticancer chemotherapy. *Dev Disabil Res Rev.* 2008;14(3):211-220.
70. Cole PD, Finkelstein Y, Stevenson KE, et al. Polymorphisms in Genes Related to Oxidative Stress Are Associated With Inferior Cognitive Function After Therapy for Childhood Acute Lymphoblastic Leukemia. *J Clin Oncol.* 2015;33(19):2205-2211.
71. Linnebank M, Malessa S, Moskau S, et al. Acute methotrexate-induced encephalopathy--causal relation to homozygous allelic state for MTR c.2756A>G (D919G)? *J Chemother.* 2007;19(4):455-457.
72. Krull KR, Bhojwani D, Conklin HM, et al. Genetic mediators of neurocognitive outcomes in survivors of childhood acute lymphoblastic leukemia. *J Clin Oncol.* 2013;31(17):2182-2188.
73. Kamdar KY, Krull KR, El-Zein RA, et al. Folate pathway polymorphisms predict deficits in attention and processing speed after childhood leukemia therapy. *Pediatr Blood Cancer.* 2011;57(3):454-460.

74. Nakao A, Miki T, Shoji H, et al. Comprehensive behavioral analysis of voltage-gated calcium channel beta-anchoring and -regulatory protein knockout mice. *Front Behav Neurosci.* 2015;9:141.
75. McEnery MW, Vance CL, Begg CM, Lee WL, Choi Y, Dubel SJ. Differential expression and association of calcium channel subunits in development and disease. *J Bioenerg Biomembr.* 1998;30(4):409-418.
76. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013;381(9875):1371-1379.
77. Breitenkamp AF, Matthes J, Nass RD, et al. Rare mutations of CACNB2 found in autism spectrum disease-affected families alter calcium channel function. *PLoS One.* 2014;9(4):e95579.
78. Soldatov NM. CACNB2: An Emerging Pharmacological Target for Hypertension, Heart Failure, Arrhythmia and Mental Disorders. *Curr Mol Pharmacol.* 2015;8(1):32-42.
79. Klotz L, Schmidt S, Heun R, Klockgether T, Kolsch H. Association of the PPARgamma gene polymorphism Pro12Ala with delayed onset of multiple sclerosis. *Neurosci Lett.* 2009;449(1):81-83.
80. Heberden C, Meffray E, Goustard-Langelier B, Maximin E, Lavialle M. Dexamethasone inhibits the maturation of newly formed neurons and glia supplemented with polyunsaturated fatty acids. *J Steroid Biochem Mol Biol.* 2013;138:395-402.
81. Fu Y, Zhen J, Lu Z. Synergetic Neuroprotective Effect of Docosahexaenoic Acid and Aspirin in SH-Y5Y by Inhibiting miR-21 and Activating RXRalpha and PPARalpha. *DNA Cell Biol.* 2017;36(6):482-489.
82. Steinbach D, Legrand O. ABC transporters and drug resistance in leukemia: was P-gp nothing but the first head of the Hydra? *Leukemia.* 2007;21:1172.
83. Mehul B, Bernard D, Simonetti L, Bernard MA, Schmidt R. Identification and cloning of a new calmodulin-like protein from human epidermis. *J Biol Chem.* 2000;275(17):12841-12847.
84. Matsuoka M. Protective effects of Humanin and calmodulin-like skin protein in Alzheimer's disease and broad range of abnormalities. *Mol Neurobiol.* 2015;51(3):1232-1239.
85. Hashimoto Y, Nawa M, Kurita M, Tokizawa M, Iwamatsu A, Matsuoka M. Secreted calmodulin-like skin protein inhibits neuronal death in cell-based Alzheimer's disease models via the heterotrimeric Humanin receptor. *Cell Death Dis.* 2013;4:e555.
86. Hasegawa S, Kumagai M, Hagihara M, et al. Distinct and Cooperative Functions for the Protocadherin-alpha, -beta and -gamma Clusters in Neuronal Survival and Axon Targeting. *Front Mol Neurosci.* 2016;9:155.
87. Pedrosa E, Stefanescu R, Margolis B, et al. Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia. *Schizophr Res.* 2008;102(1-3):210-219.
88. Naskar T, Faruq M, Banerjee P, et al. Ancestral Variations of the PCDHG Gene Cluster Predispose to Dyslexia in a Multiplex Family. *EBioMedicine.* 2018;28:168-179.
89. Kadan-Lottick NS, Zeltzer LK, Liu Q, et al. Neurocognitive functioning in adult survivors of childhood non-central nervous system cancers. *Journal of the National Cancer Institute.* 2010;102(12):881-893.
90. Buizer AI, de Sonneville LM, Veerman AJ. Effects of chemotherapy on neurocognitive function in children with acute lymphoblastic leukemia: a critical review of the literature. *Pediatr Blood Cancer.* 2009;52(4):447-454.
91. Janzen LA, Spiegler BJ. Neurodevelopmental sequelae of pediatric acute lymphoblastic leukemia and its treatment. *Dev Disabil Res Rev.* 2008;14(3):185-195.

92. Edelstein K, D'Agostino N, Bernstein LJ, et al. Long-term neurocognitive outcomes in young adult survivors of childhood acute lymphoblastic leukemia. *J Pediatr Hematol Oncol.* 2011;33(6):450-458.
93. Waber DP, Queally JT, Catania L, et al. Neuropsychological outcomes of standard risk and high risk patients treated for acute lymphoblastic leukemia on Dana-Farber ALL consortium protocol 95-01 at 5 years post-diagnosis. *Pediatr Blood Cancer.* 2012;58(5):758-765.
94. How J, Blattner M, Fowler S, Wang-Gillam A, Schindler SE. Chemotherapy-associated Posterior Reversible Encephalopathy Syndrome: A Case Report and Review of the Literature. *Neurologist.* 2016;21(6):112-117.
95. Stiefel FC, Breitbart WS, Holland JC. Corticosteroids in cancer: neuropsychiatric complications. *Cancer Invest.* 1989;7(5):479-491.
96. Brown RT, Madan-Swain A, Walco GA, et al. Cognitive and academic late effects among children previously treated for acute lymphocytic leukemia receiving chemotherapy as CNS prophylaxis. *J Pediatr Psychol.* 1998;23(5):333-340.
97. Jain N, Brouwers P, Okcu MF, Cirino PT, Krull KR. Sex-specific attention problems in long-term survivors of pediatric acute lymphoblastic leukemia. *Cancer.* 2009;115(18):4238-4245.
98. Reddick WE, Taghipour DJ, Glass JO, et al. Prognostic factors that increase the risk for reduced white matter volumes and deficits in attention and learning for survivors of childhood cancers. *Pediatr Blood Cancer.* 2014;61(6):1074-1079.
99. *WHO | Depression and Other Common Mental Disorders.* World Health Organization; 2017-02-23 13:46:21 2017.
100. Furlong W, Rae C, Feeny D, et al. Health-related quality of life among children with acute lymphoblastic leukemia. *Pediatr Blood Cancer.* 2012;59(4):717-724.
101. Recklitis CJ, Blackmon JE, Chang G. Validity of the Brief Symptom Inventory-18 (BSI-18) for identifying depression and anxiety in young adult cancer survivors: Comparison with a Structured Clinical Diagnostic Interview. *Psychol Assess.* 2017;29(10):1189-1200.
102. Teng T, Gaillard A, Muzerelle A, Gaspar P. EphrinA5 Signaling Is Required for the Distinctive Targeting of Raphe Serotonin Neurons in the Forebrain. *eNeuro.* 2017;4(1).
103. Murai KK, Pasquale EB. Can Eph receptors stimulate the mind? *Neuron.* 2002;33(2):159-162.
104. De Bellis MD, Keshavan MS, Beers SR, et al. Sex differences in brain maturation during childhood and adolescence. *Cereb Cortex.* 2001;11(6):552-557.
105. Schmithorst VJ, Holland SK, Dardzinski BJ. Developmental differences in white matter architecture between boys and girls. *Hum Brain Mapp.* 2008;29(6):696-710.
106. Ramsey LB, Panetta JC, Smith C, et al. Genome-wide study of methotrexate clearance replicates SLCO1B1. *Blood.* 2013;121(6):898-904.
107. Trevino LR, Shimasaki N, Yang W, et al. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol.* 2009;27(35):5972-5978.
108. Abe T, Kakyo M, Tokui T, et al. Identification of a novel gene family encoding human liver-specific organic anion transporter LST-1. *J Biol Chem.* 1999;274(24):17159-17163.
109. Liu SG, Gao C, Zhang RD, et al. Polymorphisms in methotrexate transporters and their relationship to plasma methotrexate levels, toxicity of high-dose methotrexate, and outcome of pediatric acute lymphoblastic leukemia. *Oncotarget.* 2017;8(23):37761-37772.

110. Yang L, Wu H, Gelder TV, et al. SLCO1B1 rs4149056 genetic polymorphism predicting methotrexate toxicity in Chinese patients with non-Hodgkin lymphoma. *Pharmacogenomics.* 2017;18(17):1557-1562.
111. Radtke S, Zolk O, Renner B, et al. Germline genetic variations in methotrexate candidate genes are associated with pharmacokinetics, toxicity, and outcome in childhood acute lymphoblastic leukemia. *Blood.* 2013;121(26):5145-5153.
112. Coon H, Darlington T, Pimentel R, et al. Genetic risk factors in two Utah pedigrees at high risk for suicide. *Transl Psychiatry.* 2013;3:e325.
113. Kock K, Koenen A, Giese B, et al. Rapid modulation of the organic anion transporting polypeptide 2B1 (OATP2B1, SLCO2B1) function by protein kinase C-mediated internalization. *J Biol Chem.* 2010;285(15):11336-11347.
114. Yang M, Xie W, Mostaghel E, et al. SLCO2B1 and SLCO1B3 may determine time to progression for patients receiving androgen deprivation therapy for prostate cancer. *J Clin Oncol.* 2011;29(18):2565-2573.
115. Pejovic-Milovancevic MM, Mandic-Maravic VD, Coric VM, et al. Glutathione S-Transferase Deletion Polymorphisms in Early-Onset Psychotic and Bipolar Disorders: A Case-Control Study. *Lab Med.* 2016;47(3):195-204.
116. Krajinovic M, Labuda D, Sinnett D. Glutathione S-transferase P1 genetic polymorphisms and susceptibility to childhood acute lymphoblastic leukaemia. *Pharmacogenetics.* 2002;12(8):655-658.
117. Krull KR, Zhang N, Santucci A, et al. Long-term decline in intelligence among adult survivors of childhood acute lymphoblastic leukemia treated with cranial radiation. *Blood.* 2013;122(4):550-553.
118. Edelmann MN, Ogg RJ, Scoggins MA, et al. Dexamethasone exposure and memory function in adult survivors of childhood acute lymphoblastic leukemia: A report from the SJLIFE cohort. *Pediatr Blood Cancer.* 2013;60(11):1778-1784.
119. Armstrong GT, Reddick WE, Petersen RC, et al. Evaluation of memory impairment in aging adult survivors of childhood acute lymphoblastic leukemia treated with cranial radiotherapy. *J Natl Cancer Inst.* 2013;105(12):899-907.
120. Hudson MM, Ness KK, Nolan VG, et al. Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. *Pediatr Blood Cancer.* 2011;56(5):825-836.

3.11. SUPPORTING INFORMATION



3.11.1. S1 Figure. Principal component analyses (PCA).

PCA analysis comparing sequencing data of 400 leukemia patients (including PETALE cohort) from Sainte-Justine University Health Center (SJUHC) to the HapMap genotype reference data (release 23) for Europeans (EUR), East Asians (EAS) and Africans (AFR).

PC1, Principal Component 1; PC2, Principal Component 2.

3.11.2. S1 Table. Genotyping: Identity of polymorphisms, details of PCR and ASO hybridization.

Polymorphisms				PCR	ASO	ASO Method
gene	**dbSNP**	**position**	**variation**	**primers**	**probes**	
ABCC3	rs12604031	Intron	**A**/G	F: TGGGTGAGTCGGCTCCAT R: AGCAGGTGCTCTGGATGC	CAGCCGC**G**GGTTC CAGCCGC**A**GGTTC	ASO
CACNB2	rs58225473	Exon	**T**/G(Asp600Glu)	Génome Québec		Sequenom
CALML5	rs10904516	Exon	**T**/C(Lys74Arg)	F:CAGGCCGGCCCTGGCGTTC**T** F:CAGGCCGGCCCTGGCGTTC**C** R:TGGAAACGGCACCATCAATG		Allele-specific PCR
EPHA5	rs33932471	Exon	**T**/G(Asn81Thr)	Génome Québec		Sequenom
GSTT1			**GSTT1/ GSTT1 null**	F:TTCCTTACTGGTCCTCACATCTC R:TCACCGGATCATGGCCAGCA		PCR
MTR	rs1805087	Exon	**A**/G(Asp919Gly)	Génome Québec		Sequenom
PCDHB10	rs2907323	Exon	**C**/G (Thr213Arg))	F:AAACTGTGGGGCATTGTCAT R:ATTAGTGGCGGTGATGAAGG	CAGCGCT**G**TGAGGG CAGCGCT**C**TGAGGG	ASO
PPARA	rs1800206	Exon	**C**/G(Leu162Val)	Génome Québec		Sequenom
SLCO1B1	rs4149056	Exon	**T**/C(Val174Ala)	Génome Québec		Sequenom

R, reverse, F, forward. The base substitution that distinguishes the two variants of each polymorphism is given in bold for ASO probes. dbSNP number is provided. Ancestral allele is given in bold and minor allele is underlined. The polymorphisms are presented as a change from ancestral to derived allele, unless ancestral allele is not known, when the change is given from major to minor allele. SNPs in coding region leading or not to amino-acid substitutions are indicated.

3.11.3. S2 Table. Significant results of association study of common variants from the candidate genes of relevance for nervous system function, PETALE cohort, WES data (n=191).

Outcome	Gene	SNP	Ref allele	Var allele	MAF	P value [a]	Allelic ratio frequencies	
							AFF	UNAFF
Trial making test	**High risk[b], N=108**							
	PCDHB10	rs2907323	C	G	0.13	0.0006	9/17	19/167
Verbal fluency	**All cohort, N=191**							
	CALML5	rs10904516	T	C	0.24	0.0002	29/43	56/230
	High risk[b], N=108							
	CALML5	rs10904516	T	C	0.24	0.0002	20/24	28/128
Digit span	**Males, N=87**							
	CALML5	rs10904516	T	C	0.22	0.0003	16/20	21/111
	Chemotherapy only, N=79							
	CACNB2	rs58225473	T	G	0.17	0.0004	9/11	18/120
Moderate-severe anxiety	**Females, N=104**							
	EPHA5	rs33932471	T	G	0.08	0.00003	6/12	9/163
Moderate-severe depression	**All cohort, N=191**							
	EPHA5	rs33932471	T	G	0.06	0.0004	8/36	13/283
	Females, N=104							
	EPHA5	rs33932471	T	G	0.08	0.0003	7/21	8/154

[a]Association test based on comparing allele frequencies between cases and controls. All associations have FDR-BH (Benjamini-Hochberg false discovery rate) lower than 5%. All also have p value lower than 0.001, which is Bonferroni cut-off value for the number of variants tested in nervous system function pathway.

[b]Stratified analyses according to sex and treatment intensity (standard vs high risk); chemotherapy only vs chemotherapy and cranial radiation therapy (CRT).

Ref: reference allele; Var: variant allele; MAF: minor allele frequency, *PCDHB10*, protocadherin beta 10, *CALML5*: Calmodulin Like 5, *CACNB2*: Calcium Voltage-Gated Channel Auxiliary Subunit Beta2, *EPHA5*: EPH Receptor A5, Brain-Specific Kinase.

3.11.4. S3 Table. Results of association study of common variants from methotrexate and corticosteroids pathways, PETALE cohort, WES data (n=191).

Outcome	Gene	SNP	Ref allele	Var allele	MAF	P value [a]	Allelic ratio frequencies	
							AFF	UNAFF
Verbal fluency	**Males, N=87 [b]**							
	MTR	rs1805087	A	G	0.19	0.0006	14/22	19/119
	Standard risk, N=83							
	PPARA	rs1800206	C	G	0.07	0.0001	7/21	5/133
	Females, N=104							
	PPARA	rs1800206	C	G	0.07	0.0003	6/30	4/168
	High risk, N=108							
	ABCC3	rs12604031	G	A	0.44	0.0006	25/11	50/84
	Chemotherapy only, N=79							
	PPARA	rs1800206	C	G	0.06	0.00005	6/18	4/130
	Chemotherapy and CRT, N=112							
	SHMT1[c]	rs1979277	G	A	0.27	0.002	21/25	35/123
	ABCC3	rs12604031	G	A	0.47	0.003	27/13	55/81
Moderate-severe depression	**Standard risk, N=83**							
	ADORA3[c]	rs35511654	T	G	0.12	0.002	6/12	10/112
	Chemotherapy only, N=79							
	SLCO1B1	rs4149056	T	C	0.15	0.0001	8/10	12/108
	ADORA3[c]	rs35511654	T	G	0.10	0.004	5/13	8/112
	ABCC3[c]	rs11568591	G	A	0.07	0.004	4/14	5/115

[a] Association test based on comparing allele frequencies between cases and controls. All associations have FDR-BH (Benjamini-Hochberg false discovery rate) lower than 5%.

[b] Stratified analyses according to sex and treatment intensity (standard vs high risk); chemotherapy only vs chemotherapy and cranial radiation therapy (CRT).

[c] SNPs or associations that did not qualify for genotyping with p value higher than 0.0019 (Bonferroni cut-off value for the number of genes tested in MTX/CS pathway).

Ref: reference allele; Var: variant allele; MAF: minor allele frequency; *MTR*: 5-Methyltetrahydrofolate-Homocysteine Methyltransferase, *PPARA*: Peroxisome Proliferator Activated Receptor Alpha, *ABCC3*: ATP Binding Cassette Subfamily C Member 3, *SHMT1*: Serine Hydroxymethyltransferase 1, *ADORA3*: Adenosine Receptor A3, *SLCO1B1*: Solute Carrier Organic Anion Transporter.

3.11.5. S4 Table. The combined cohort represents the pooled samples from the discovery PETALE cohort and replication SJLIFE cohort (N=781). Combined cohort analysis was performed for the variants in *CACNB2* and *MTR* genes.

Outcome	Genotype	Case* N (%)	Control* N (%)	Model	Case* N (%)	Control* N (%)	P**	OR (95%-CI)
				CACNB2 rs58225473				
				All patients				
Digit span	TT	48 (64.0)	407 (69.8)	TT+TG	70 (93.3)	572 (98.11)	0.01[r]	3.7 (1.25-11)
	TG	22 (29.3)	165 (28.3)					
	GG	5 (6.7)	11 (1.9)	GG	5 (6.7)	11 (1.89)		
				Chemotherapy only* **				
	TT	20 (58.8)	175 (68.4)	TT+TG	29 (85.3)	250 (97.7)	0.0004[r]	7.2 (2.1-25)
	TG	9 (26.5)	75 (29.3)					
	GG	5 (14.7)	6 (2.3)	GG	5 (14.7)	6 (2.3)		
				MTR rs1805087				
				All patients				
Verbal fluency	AA	67(59.8)	443 (69.6)	AA	67 (59.8)	443 (69.6)	0.04[d]	1.5 (1-2.3)
	AG	37 (33.0)	179 (28.1)	AG+GG	45 (40.2)	194 (30.5)		
	GG	8 (7.2)	15 (2.4)					
	AA	35 (57.4)	221 (70.6)	AA	35 (57.4)	221 (70.6)	0.04[d]	1.8 (1-3.1)
	AG	20 (32.8)	85 (27.2)	AG+GG	26 (42.6)	92 (29.4)		
	GG	6 (9.8)	7 (2.2)					

*Participants with and without indicated complications are defined as cases and controls, respectively
**P values are calculated by Chi-square. The most representative genetic model used is indicated (d: Dominant, r: Recessive).
***Chemotherapy without cranial radiation therapy.

Section B

Chapter 4

An exome-wide association study between genetic factors and long-term treatment related neurocognitive deficits, and anxiety and depression in survivors of childhood acute lymphoblastic leukemia.

The following article is currently in preparation; the top-ranking association results are under investigation in the SJLIFE replication cohort.

An exome-wide association study between genetic factors and long-term treatment related neurocognitive deficits, and anxiety and depression in survivors of childhood acute lymphoblastic leukemia.

Running title: Genetics of Neurocognitive complications and Mood disorders in survivors of childhood ALL.

Kateryna Petrykey[1,2], Sarah Lippé[1,3], Serge Sultan[1,3], Philippe Robaey[1,5,6,7], Simon Drouin[1], Laurence Bertout[1], Patrick Beaulieu[1], Pascal St-Onge[1], Jessica L. Baedke[8], Yutaka Yasui[8], Melissa M. Hudson[8,9], Caroline Laverdière[1,4], Daniel Sinnett[1,4], Maja Krajinovic[1,2,4*]

1. Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada.
2. Department of Pharmacology and Physiology, Université de Montréal (Quebec), Canada.
3. Department of Psychology, Université de Montréal (Quebec), Canada.
4. Department of Pediatrics, Université de Montréal (Quebec), Canada.
5. Children's Hospital of Eastern Ontario, Ottawa (Ontario), Canada.
6. Department of Psychiatry, Université de Montréal (Quebec), Canada.
7. Department of Psychiatry, University of Ottawa (Ontario), Canada.
8. Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis (TN), USA.
9. Department of Oncology, St. Jude Children's Research Hospital, Memphis (TN), USA.

***Corresponding author:**

Maja Krajinovic
Research Center, Sainte-Justine University Health Center
3175 chemin de la Côte-Sainte-Catherine
Montreal, Quebec, Canada, H3T 1C5
E-mail : maja.krajinovic@umontreal.ca
Fax : 514-345-4731
Phone : 514-345-4931 #6259

4.1. ABSTRACT

Background: An increased risk of neurocognitive deficits, anxiety, and depression, has been reported in childhood cancer survivors exposed to cytotoxic chemotherapeutic drugs and/or cranial radiation therapy during a sensitive period of child development.

Patients and Methods: To further address this problem, we analyzed associations of neurocognitive deficits, as well as anxiety and depression, with common and rare genetic variants derived from whole-exome sequencing data of acute lymphoblastic leukemia survivors (PETALE cohort). In addition, first rank associations were assessed using stratified and multivariable analyses.

Results: Significant associations were identified in the entire discovery cohort between the *AK8* gene and changes in neurocognitive function, whereas *PTPRZ1, MUC16, TNRC6C-AS1* were associated with anxiety. Following stratification according to sex, the *ZNF382* gene was linked to neurocognitive deficit in males, whereas *APOL2* and *C6orf165* were associated with anxiety, and *EXO5* gene with depression. Following stratification according to relapse risk groups, the modulatory effect of rare variants on depression was additionally found in the *CYP2W1* and *PCMTD1* genes.
Next, top-ranking common associations were analyzed in an independent SJLIFE replication cohort of ALL survivors. Among those, the male-specific association between neurocognitive deficit and common variant rs61732180 in the *ZNF382* gene was not significant, however a p-value<0.05 was observed when the entire SJLIFE cohort was analyzed. *ZNF382* was significant in males in the combined cohorts as shown by meta-analyses as well as the depression-associated gene *EXO5*.

Conclusions: Our results suggest that specific genes may be related to increased neuropsychological consequences. Further research is needed to confirm whether the current findings, along with other known risk factors, may be of value in identifying patients at increased risk of these long-term complications.

KEYWORDS

Childhood acute lymphoblastic leukemia, cancer survivors, late adverse effects, genetic factors, association study, whole exome sequencing, pharmacogenomic markers, cognitive performance, anxiety, depression, neurocognitive complications, mood disorders.

4.2. INTRODUCTION

The survival rates in children diagnosed with acute lymphoblastic leukemia (ALL)[1,2], the most frequent childhood cancer[3], have dramatically increased over the past decades due to the introduction of multi-agent risk-adapted treatment regimens and outstanding improvements in care delivery. However, exposure to cytotoxic therapy during a vulnerable period of child development can have long-term consequences, including impaired neurocognitive functions[4,5], and mood disorders[6,7]. Furthermore, childhood and adolescence are periods characterized by intensive development of the central nervous system[8,9], which is pertinent in the context of the impact of cancer treatment on the integrity of the white matter[10-14]. Indeed, numerous studies conducted in childhood ALL survivors[4,15,16] have reported an increased risk of neurocognitive deficits[6] in attention[15-20], working memory[21], processing speed[16,22,23], and executive functions, such as verbal fluency and cognitive flexibility[24]; as well as depression, anxiety, behavioral difficulties, distress, and post-traumatic symptoms compared with siblings[25-30].

Varying degrees of neurocognitive dysfunction and levels of emotional distress associated with cancer treatment have been observed that differ by patient characteristics, such as age and sex, and possibly reflecting different underlying mechanisms[5,31,32]. Moreover, while some survivors may not experience any of these complications, others may have more than one. Factors contributing to this variability, include the type of treatment, the characteristics of the malignancy, the lifestyle, and, the genetic makeup of the patient[33].

We examined whether common and rare genetic polymorphisms contribute to this variability by altering the risk of treatment-related neurocognitive deficits, as well as anxiety and depression in combination with non-genetic factors.

We previously analyzed these complications in a well-described cohort of ALL survivors (PETALE)[34] using a candidate gene approach[33]; two associations between the *MTR* and *CACNB2* genes and

neurocognitive deficit were validated in an independent SJLIFE replication cohort (St. Jude Children's Research Hospital, Memphis, USA)[33].

Here, the association analyses are extended to a hypothesis-free approach – an exome-wide association study, which could identify additional genes as potential modulators of the risk of these complications.

4.3. STUDY POPULATION AND METHODS

4.3.1. Discovery cohort

The discovery cohort included 229 patients diagnosed and treated for childhood ALL according to Dana Farber Cancer Institute (DFCI) ALL 87-01 to 05-01 protocols at Sainte-Justine University Health Center (SJUHC), Montreal, (Quebec), Canada. The participants were recruited during 2013-2015 in the context of the PETALE study[34]. Eligible participants were of European descent, younger than 19 years old at diagnosis and older than 12 years at evaluation, at least 5 years after diagnosis of ALL, without a history of relapse or refractory ALL or Down syndrome, and had not received a hematopoietic stem cell transplant. The median age of patients at the time of diagnosis was 4 years, the time from the end of treatment to evaluation ranged from 3-24 year with a median of 13 years (for 76.0% of participants, it was ≥10 years), both sexes were equally represented (51.1% of females). The patients were classified to standard (SR) and high relapse risk (HR) groups based on prognostic factors, including age, white blood cell count, immunophenotype, and central nervous system (CNS) status at diagnosis[35,36]. The frequency of patients assigned to SR and HR groups during the treatment was 45.9% and 54.1%, respectively.

4.3.2. Neuropsychological evaluation

A neurocognitive evaluation was performed using standardized testing procedures. It included three indices from two neurocognitive measures that reflect common impairments among childhood ALL survivors and are also good predictors of general neuropsychological outcomes[37]: Trail Making Test – Condition 4 - Letter-Number Sequencing score and Verbal Fluency – Condition 1 – Letter fluency score from the Delis-Kaplan Executive Function System (D-KEFS)[38]; and Digit Span from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)[39] total score. Trail Making Test (D-KEFS) score is a measure that reflects processing speed, psychomotor speed, and cognitive flexibility[40]. Verbal Fluency (D-KEFS) score is a measure of phonological fluency in verbal modality[41]. Digit span (WAIS-IV) total score is a measure of verbal working memory[42]. Raw scores were converted to age-adjusted scaled scores

based on population means[43]. Neurocognitive outcomes were transformed into dichotomous variables and studied accordingly. For each of these variables, scores lower than one and a half standard deviations below the mean of the normative dataset were indicative of impairment[44], all other scores were considered non-impaired.

4.3.3. Anxiety and Depression

Participants were classified as having anxiety or depression if they demonstrated elevated symptoms according to two measures referenced to age-specific norms. This was done in line with published recommendations[45,46] and previous use of the instruments modality [26,27,47]. For younger participants (<19 years), we used anxiety and depression modules of the Beck Youth Inventories -Second Edition (BYI), a self-report instrument to document psychological status in children from 7 to 18 years old[46]. For older participants (≥19 years) we used the Brief Symptom Inventory-18 (BSI-18 anxiety and depression score), an 18-item self-report questionnaire, assessing psychological distress in adults[45], previously also used in cohorts of young and older adult survivors of childhood cancer[48,49]. Internal consistency coefficients measured by Cronbach's alphas were all satisfactory, >0.80[50]. Age-adjusted scores one standard deviation above the population mean were considered as impaired.

4.3.4. Sequencing and quality control

Whole-exome sequencing (WES) was performed on germline DNA, extracted from peripheral blood samples from participants in the PETALE cohort, using standard protocols as described previously[33]. Whole exomes were captured in solution with Agilent's SureSelect Human All Exon 50Mb kits and sequenced on either Life Technologies SOLiD System 4.0 (mean coverage = 40X) or Illumina HiSeq 2500 platform (mean coverage = 113.1X) at SJUHC integrated clinical genomic center in pediatrics. Only missense, nonsense, and splicing common and rare variants with predicted functional impact (Sift (<0.1) and/or PolyPhen2 (≥0.85)) were considered[51,52]. Variants were defined as rare (minor allele frequency, MAF<5%) and common (MAF≥5%) according to the reported frequency for European populations in public datasets[53]. Variants exceeding a missing rate of 20%, not in Hardy-Weinberg Equilibrium (p=0.05/number of tests)[54], and common variants with pairwise linkage disequilibrium (LD, r2≥0.8) were excluded. Therefore, 5312 common genetic variants corresponding to 3793 genes, and 58924 rare genetic variants corresponding to 11441 genes that satisfied the above-mentioned filtering criteria entered the association analyses.

4.3.5. Association analyses

The P-value threshold of 5×10^{-8}, commonly used to identify an association between a common genetic variant and an outcome of interest in a typical GWAS[55-57] is not applicable to our study since variants for analysis were selected from WES data to focus only on those predicted by various instruments to affect coding protein function[51,52]. As a result, 5312 common variants were analyzed, which is far less than the typical number of GWAS variants; and the Benjamini-Hochberg procedure for false discovery rate (FDR)[58,59] was used to adjust for number of variants tested with an adjusted cut-off value of <5% considered to be statistically significant[57]. The analyses between common genetic variants and neurocognitive outcomes, and anxiety/depression were performed by the allelic chi-square or Fisher's exact test implemented in PLINK v.1.07[60,61]. Analyses were performed in 229 sequenced patients and

stratified by sex and risk groups with different treatment intensity because these factors have an established role in modulating neurocognitive outcomes[62,63]. For top-ranking associations, multivariable logistic regression analysis was also performed to assess the effect of genotype when controlling for non-genetic covariates. Multiple regression models included the following adjustment variables: age at the time of diagnosis (continuous variable); time since the end of treatment (continuous variable); sex: males/females (categorical variable); DFCI Protocol: 87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5 (categorical variable); risk: SR/HR (categorical variable); treatment combined variable with the following scores: patients who received only chemotherapy =1, patients who received chemotherapy and cranial radiation =2 (categorical variable). The associations that remained significant through multivariable regression models were retained for further analyses. The effect of genotype was quantified by odds ratio (OR) with 95% CI according to the most representative model, which was either dominant or recessive in all cases. In addition, the potential additive effect of combining risk loci by recoding genotypes as having none, one, or two or more risk alleles has also been explored.

The effects of rare variants were evaluated using the SKAT-O test (Optimal Sequence Kernel Association Test)[64-66] implemented in SKAT package v.2.0.1[67]. Only genes with at least two variants that satisfied filtering criteria were retained for the SKAT-O test. To evaluate individual variant contributions to association signals, a collapsing approach[68,69], with the iterative exclusion of every single variant, was additionally executed in SPSS v.25.0. Similar to the common variant analyses, genetic associations were assessed for each of the neurocognitive outcomes in the entire cohort, in different sex groups, as well as in SR and HR groups, and through multivariable models adjusted for the covariates described above. The multiple test adjustment (FDR) for the number of genes tested was included in all analyses.

4.3.6. Replication cohort

The replication cohort consisted of childhood ALL survivors of European ancestry enrolled in the St. Jude Lifetime Cohort (SJLIFE) study with whole-genome sequencing (**Table 6**). The maximum number of participants in the replication cohort with available outcome data was 675; the total number varied depending on the outcomes and subgroup studied. Participants were selected to resemble the discovery cohort based on demographics and treatment characteristics. They were younger than 19 years at diagnosis, older than 12 years at evaluation, with no history of relapse within 5 years of the primary ALL diagnosis date, Down syndrome, or hematopoietic stem cell transplantation. The median age at diagnosis was 4.9 years, the median time since 5-year survival from ALL diagnosis was 25.8 years for neurocognitive evaluation, and 25.6 years for anxiety and depression evaluation; 50.5% of participants were males. All outcome measures were the same as in the discovery cohort with the exception of the Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety and Depression Scales[70-72] that were used to assess anxiety and depression in participants under 18 years of age. Associations that remained significant in the discovery cohort using multivariable regression models were analyzed in replication cohort by Fisher exact test for allelic contingency tables and logistic regression adjusting for continuous age at diagnosis, sex, continuous time between date of becoming a 5-year ALL survivor and date of test measurement, whether the survivor was treated with chemotherapy only versus chemotherapy plus radiation, and the top 20 principal components adjusting for genetic ancestry. Variant rs750295511 (*MUC16*) did not pass quality control and was excluded from the analysis. Stratification by risk group designation was not available for the SJLIFE cohort. Rare variant replication analyses were not performed since rare variant associations in the discovery cohort were only detected in risk group stratified analyses.

4.4. RESULTS

4.4.1. Discovery cohort characteristics

Demographics and clinical characteristics of PETALE participants are presented in **Table 1.** The analyses were performed in either all patients or subgroups. These included patients assigned to standard (SR, 45.9%) and high risk (HR, 54.1%) groups, males (48.9%) and females (51.1%). The median age of ALL survivors at the time of evaluation was 21 years. The most prevalent deficit in neurocognitive test performance was noted for digit span (19.7%) followed by verbal fluency (19.2%) and trail making test (8.7%). Moderate-severe anxiety was noted in 9.2% survivors, whereas 10.5% of survivors were affected by moderate-severe depression.

4.4.2. Common variants

Among the common genetic variants, the top-ranking associations obtained using PLINK (**Supplemental Tables 1-3)** were assessed through multivariate regression models that also included non-genetic co-variables. Only the associations that remained significant in these models were retained for further analyses (**Supplemental Tables 4-7).**

Accordingly, significant associations were detected between Trail making test and rs17407084 variant in the *AK8* gene (OR=7.3, 95% CI, 2.7-19.7; p=4.52E-04), as well as between Moderate-severe anxiety and the following variants: rs740965 in *PTPRZ1* (OR=5.1, 95% CI, 1.98-12.9; p=1.00E-03), rs2748431 in *MUC16* (OR=8.3, 95% CI, 3.1-22.5; p=3.10E-05), and rs2748431in *TNRC6C-AS1* (OR=6.1, 95% CI, 2.0-18.1; p=1.00E-03). These association are presented in their best genetic models (**Tables 2** and **3)** as well as by Manhattan plots (**Supplemental Figure 1** and **2).** Additionally, the combined effect of the rs740965 (*PTPRZ1*), rs2748431 (*MUC16*), and rs2748431 (*TNRC6C-AS1)* variants was identified by recoding genotypes as having none, one or two and more alleles at risk (**Supplemental Figure 3).**

Following stratification according to sex we identified variant rs61732180 in the *ZNF382* gene that was associated with the increased risk of deficit scores in the Trail making test in male participants (OR=20.2, 95% CI, 4.3-95.4; p=2.62E-04, **Table 2**). Male carriers of the variant alleles in the rs7285167 (*APOL2*) and rs61731441 (*C6orf165*) genes were more prone to Moderate-severe anxiety (OR=9.6, 95% CI, 2.3-40.5; p=3.00E-03 and OR=10.9, 95% CI, 2.5-47.0; p=2.00E-03, respectively, **Table 3**), whereas male carriers of the variant allele in the rs35672330 (*EXO5*) gene were more prone to Moderate-severe depression (OR=20.3, 95% CI, 4.1-99.8; p=4.90E-04, **Table 4**). Additionally, the combined effect of the rs7285167 (*APOL2*) and rs61731441 (*C6orf165*) variants was seen when they were tested by recoding genotypes as having none, one or two and more alleles at risk (**Supplemental Figure 4).**

We did not find any significant common variant association that would satisfy multiple testing adjustments when performing an exome-wide association with the remaining neurocognitive phenotypes either in entire group or following stratification.

4.4.3. Replication results.

Common variants associations that remained significant in the discovery cohort using multivariable regression models were further analyzed for an association using the respective phenotype measures in the independent cohort of ALL survivors (SJLIFE cohort) (**Tables 5a** and **5b**). An association between deficit in the trail making test performance and the minor allele of *ZNF382* rs61732180 was not observed in males, however it was seen in all survivors (OR=1.4; 95% CI, 1.04-1.90; p=0.025) for both allelic Fisher and adjusted logistic regression analyses (**Tables 5a** and **5b, Supplemental Tables 8a** and **8b**). Additionally, variant rs61732180 in the *ZNF382* gene remained significant in the combined discovery and replication set that included male participants when assessed through meta-analysis (**Supplemental Figure 5**).

An allelic association between moderate-severe anxiety and *APOL2* rs7285167 was not observed in males but was seen with p-value<0.05 in the male-restricted adjusted logistic regression analysis and the

full SJLIFE cohort allelic and adjusted logistic regression analyses (**Tables 6a** and **6b, Supplemental Tables 8a** and **8b**); however, it had the opposite effect and therefore cannot be considered as confirmed through the replication study. Moreover, although an association between Moderate-severe depression and the minor allele of *EXO5* rs35672330 was not observed in males, the effect of this allele was detected in the pooled discovery and replication cohort through the meta-analysis (**Supplemental Figure 6**).

4.4.4. Rare variants

The analysis of functionally predicted rare variants in PETALE cohort led to the detection of the associations between the moderate-severe anxiety and rare variants in the *PCMTD1* and *CYP2W1* genes in the HR patients (p=9.4E-6 and p=1.3E-5, respectively, **Table 6**). Using the collapsing approach, we explored variant combinations that contributed to the observed association signal, consequently identifying two variants in *PCMTD1* (rs201786115 and rs200377849), and one in *CYP2W1* (rs3735684). No significant association was obtained for moderate-severe depression and neurocognitive outcomes through the rare variants' analysis.

4.5. DISCUSSION

4.5.1. Neurocognitive function

In our study, functionally predicted germline common variants in the *AK8* and *ZNF382* genes were found to be significantly associated with deficits in the performance of the trail making test in PETALE participants. The association with the *AK8* gene was detected when the entire cohort of survivors was analyzed and also in the standard-risk group, whereas that of *ZNF382* was only context-dependent, and was detected upon sex stratification. The *ZNF382* association merits special attention since its effect was observed in the replication cohort. The detailed description of the *AK8* gene function is provided in Supplementary Material (**Item S1**).

Zinc finger protein 382, encoded by the *ZNF382* gene, is a member of the largest family of transcriptional regulators - Krüppel-associated box domain (KRAB) zinc finger proteins[73]. It plays critical roles as a transcription inhibitor and has been suggested to be a tumor suppressor in various types of human cancer, including pediatric acute myeloid leukemia[74,75]. Interestingly, ZNF382 inhibits the activating protein 1 (AP-1) and nuclear factor kappa-B (NF-kB) signaling. NF-κB involvement was detected in the different categories of neurons including both excitatory (glutamatergic) and inhibitory (GABAergic), as well as in the neural sub-compartment of the synapse; thus suggesting that neuronal NF-κB signaling pathway functions under normal physiological conditions to promote synaptic growth and to improve synaptic activity and long-lasting forms of plasticity[76]. Moreover, its activation by excitatory neurotransmission and participation in multiple forms of structural and synaptic plasticity is probably at the basis of the function of this transcription factor in cognitive behaviors[76]. On the other hand, NF-κB, a key regulator of innate immunity, is over-activated in a number of neurodegenerative diseases, including Alzheimer's disease (AD)[77]. Although the male-specific association between the deficit score in neurocognitive test performance and the minor rs61732180 *ZNF382* allele identified in our study was not observed in males from the SJLIFE replication cohort, however, a p-value<0.05 was obtained for the full SJLIFE

population. Therefore, the involvement of the *ZNF382* gene in neurocognitive function may warrant further investigation.

Childhood ALL patients who have received CNS-directed chemotherapy demonstrate persistent and significant neurocognitive impairment that manifests after treatment[5,78]. Specifically, treatment of ALL may result in smaller gray and white matter volumes as well as alterations in white matter microstructure, often associated with decreased neurocognitive performance[79]. Our previous work identified, using a candidate gene approach, a panel of several genes that showed an effect on neurocognitive decline in the PETALE cohort. Two variants, rs1805087 in the *MTR* gene and rs58225473 in the *CACNB2* gene, deserve mention as these associations were confirmed in the independent SJLIFE replication cohort.

Interestingly, genes identified in the present study through a hypothesis-free approach regarding the impaired neurocognitive function are associated to varying degrees with oxidative stress (*AK8* gene) or with immune regulation (*ZNF382* gene); and all of them, in one way or another, are also related to the function of the CNS. There is a lot of evidence to suggest that the brain is very susceptible to oxidative damage due to its high metabolic demand[80,81]. The oxidative stress, which is related to elevated intracellular levels of reactive oxygen species (ROS) is a key mediator of neuroinflammation, metabolic changes, bioenergetic deficiency, and neuronal apoptosis [82]. ROS generated during chemotherapy may be associated with various harmful events, including neurotoxicity[83]. For example, methotrexate (MTX) promotes oxidative stress in several organs, including the brain[83]. In addition, MTX inhibits the activation of NF-κB, the already mentioned protein complex, which, among other functions, plays a central role in DNA transcription and the regulation of inflammation[84,85].

4.5.2. Anxiety and Depression

We found that functionally predicted germline common variants in the *PTPRZ1, MUC16, TNRC6C-AS1, APOL2,* and *C6orf165* genes were significantly associated with moderate-severe anxiety in ALL survivors; whereas the *EXO5* gene was associated with the increased risk of moderate-severe depression.

Noteworthy, most of the associations were sex-specific and were detected in males, which may be partly explained by the fact that self-reported anxiety/depression includes more social variance and less biogenetic variance in females than in males[86]. Additionally, an association was found between the moderate-severe anxiety in HR patients and rare variants enrichment in the *PCMTD1* and *CYP2W1* genes. The *EXO5* association deserves special mention since its effect was also seen in the pooled discovery and replication cohort.

EXO5 (Exonuclease 5) is a single-stranded DNA-specific bidirectional exonuclease that functions in the repair of nuclear DNA[87]. In a recent study, the *EXO5* gene was identified as a risk gene involved in prostate tumorigenesis[88]. Although there is no data available on the potential involvement of *EXO5* in mood disorders or CNS function, it is interesting to note that the effect of the *EXO5* gene detected in our cohort was also gender-specific and has been identified in male individuals.

The detailed description of the gene functions of the associations reported above is provided in Supplementary Material (**Item S1**).

Emotional challenges that arise from diagnosis and treatment of cancer in childhood can be seen as a highly traumatic event and can have long-term consequences. There is growing evidence that exposure to psychological distress at an early age can seriously affect brain maturation and development[89]; in addition, childhood stress can increase vulnerability to later development of mental disorders, such as depression and anxiety[90]. Particularly, during peri-adolescence, brain areas critical for emotional regulation, such as the prefrontal cortex, hippocampus, and amygdala are still developing and are highly sensitive to stress[89]. In a similar stressful environment, individuals will respond to stress differently as only part of them will demonstrate vulnerability, while others will remain resilient[30,91].

The loci reported in the current study have not previously been identified as potential risk predictors of anxiety and/or depression susceptibility, however, given their important biological role, may warrant

further study. In addition, the significant combined effect demonstrated in our study indicates that a single marker and/or single genotype may not be sufficient to explain the etiology of psychological distress phenotypes, given their complexity and environmental influences.

We acknowledge that our study has certain limitations. For example, limited sample size may affect the accuracy of the results, especially in the context of a stratified analysis. Some phenotypes, such as trail making test, appear to occur with the similar frequency as in the general population, but the cause of their development, including genetic predisposition, may differ between patients exposed to the treatment and untreated individuals[92]. At the same time, it is also important to note that survivors may tend to systematically report lower or normal distress rates as a result of a tendency to overnormalize their situation. If this is the case in the present group, then it is likely that those with moderate-severe levels experience feelings to an even more significant degree. Thus, despite the small number of affected individuals in the study group it is nevertheless legitimate to explore why some individuals are more vulnerable. The association results obtained for rare variants in the discovery cohort (not evaluated in the SJLIFE cohort) should be taken with caution given their low number. Most of the common variant associations found in the PETALE cohort were not replicated in the SJLIFE cohort. This can be explained by several reasons. Mainly, despite the use of similar inclusion/exclusion criteria and the use of similar outcomes between the two cohorts, it is possible that the small sample sizes in both cohorts, differences in treatment protocols, and/or timing between ALL diagnosis and evaluation contributed to the observed discrepancies. Finally, we can not disregard the possibility that some of the associations observed in the discovery cohort could have been obtained by chance.

4.6. CONCLUDING REMARKS

Using WES data and a hypothesis-free approach, we identified several genes as potential modulators of the risk of developing treatment-related neurocognitive complications, as well as anxiety and depression. The association between deficit in the trail making test and variant rs61732180 in the *ZNF382* and rs35672330 in *EXO5* genes are of particular interest since association were also found in the replication cohort or poled discovery and replication cohorts.

Multiple evidence has been collected nowadays for potential genetic and epigenetic risk markers of the long-term treatment-related neurocognitive and emotional complications in survivors of childhood cancer. In addition, accumulating data suggest that genetic factors contribute significantly to resilient responses to trauma and stress[93]. Large genome-wide association studies on the genetic architecture of mental disorders indicate its polygenic nature[94-96]. Therefore, future studies will be required not only to verify current results, but also for multilevel integration of several approaches including polygenic score models along with other non-genetic factors, in order to identify markers of neurocognitive and emotional disorders and implement them into clinical practice.

4.7. DECLARATIONS

Ethics approval and consent to participate

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of SJUHC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon request and revision of the projects for which the data might be used.

Competing interests

The authors declare no competing financial interests.

Funding

Institute of Cancer Research (ICR) of the Canadian Institutes of Health Research (CIHR) **grant number: 118694** awarded to Dr Daniel Sinnett, Dr Maja Krajinovic, Dr Caroline Laverdière, Dr Philippe Robaey, Dr Sarah Lippé, Dr Serge Sultan (http://www.cihr-irsc.gc.ca); C17 Council, Cancer Research Society (CRS), Canadian Cancer Society Research Institute (CCSRI), Ontario Institute for Cancer Research (OICR), Pediatric Oncology Group of Ontario (POGO), Garron Family Cancer Centre at SickKids Hospital (Ontario), The Terry Fox foundation, **grant number: TFF - 105266** awarded to Dr Daniel Sinnett and Dr Maja Krajinovic (https://www.terryfox.org/), FRQS Applied Medical Genetics Network and Sainte-Justine Hospital Foundation supported the PETALE study. Kateryna Petrykey is a scholar of the Cole Foundation and the Université de Montréal.

Acknowledgements

The authors would like to thank all childhood ALL survivors and their parents for the participation in the PETALE study, as well as all study collaborators for their valuable contribution.

We would like to thank Dr. Kevin R. Krull from the St. Jude Children's Research Hospital for a critical review of the manuscript. We also thank Aziz Rezgui for his contributions to patient sampling and data processing for the PETALE cohort.

4.8. TABLES

4.8.1. Table 1. Patient demographics and clinical characteristics, N=229.

	N	%
Sex		
Males	112	48.9
Females	117	51.1
Prognostic risk group		
Standard risk	105	45.9
High risk	124	54.1
Criteria for High-risk stratification*		
Age	41	33.1
WBC	33	26.6
T-cell markers	5	4.1
Combination Age and/or WBC and/or T-cell markers	21	16.9
CNS involvement	16	12.9
MRD+	2	1.6
Other**	6	4.8
DFCI protocol		
(87-01)	18	7.9
(91-01)	47	20.5
(95-01)	68	29.7
(00-01)	72	31.4
(05-01)	24	10.5
Cranial radiation therapy		
Median 18 (Gy)		
Yes	134	58.5
No	95	41.5
Age at diagnosis - median (range)		
		4 (0-18)
Time since the end of treatment		
More than 10 years	174	76
Less than 10 years	55	24
Median	Min	Max
13	3	24

TESTED OUTCOMES			
Binary outcomes	**Affected**	**Unaffected**	**Missing**
Neurocognitive outcomes			
Trail making test	20 (8.7%)	209 (91.3%)	0 (0.0%)
Verbal fluency	44 (19.2 %)	184 (80.4%)	1 (0.4%)
Digit span	45 (19.7%)	187 (80.3%)	0 (0.0%)

Emotional distress outcomes			
Moderate-severe anxiety	21 (9.2%)	180 (78.6%)	28 (12.2%)
Moderate-severe depression	24 (10.5%)	177 (77.3%)	28 (12.2%)

DFCI, Dana-Farber Cancer Institute. WBC: White Blood Cell; CNS: Central Nervous System; MRD: Minimal Residual Disease.

*Criteria for High-risk stratification were mainly attributed based on age, white blood cell count, immunophenotype (presence of T-cell markers) and combination of these factors; as well as central nervous system (CNS) status and Minimal residual disease at diagnosis.

**This category represents patients with chromosomal abnormalities and/or combination of factors.

4.8.2. Table 2. Top-ranking associations of the common variants analysis regarding the performance of the Trail making test.

Genotype	Case*, N (%)	Control*, N (%)	Model	Case*, N (%)	Control*, N (%)	P value**	OR [95%-CI]	P value adj***
				AK8 rs17407084				
				All cohort, N=229				
TT	11 (55.0)	188 (90.0)	**TT**	11 (55.0)	188 (90.0)	2.16E-04[d]	7.3 [2.7-19.7]	4.52E-04
TC	7 (35.0)	21 (10.0)	**TC+CC**	9 (45.0)	21 (10.0)			
CC	2 (10.0)	0 (0.0)						
				ZNF382 rs61732180				
				Males, N=112				
CC	3 (27.3)	65 (64.4)	**CC+CT**	6 (54.5)	97 (96.0)	3.58E-04[r]	20.2 [4.3-95.4]	2.62E-04
CT	3 (27.3)	32 (31.6)	**TT**	5 (45.5)	4 (4.0)			
TT	5 (45.4)	4 (4.0)						
				AK8 rs17407084				
				Standard risk, N=105				
TT	2 (28.6)	90 (91.8)	**TT**	2 (28.6)	90 (91.8)	2.44E-04[d]	28.1 [4.7-168.8]	4.40E-04
TC	4 (57.1)	8 (8.2)	**TC+CC**	5 (71.4)	8 (8.2)			
CC	1 (14.3)	0 (0.0)						

AK8:Adenylate Kinase 8 ATP-AMP Transphosphorylase 8; ZNF: Zinc Finger Protein 382; FDR: false discovery rate; OR: odds ratio.

*Participants with and without indicated complications are defined as cases and controls, respectively.

**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (a: Additive; d: Dominant, r: Recessive).

***P value adj: p value from logistic regression adjusted for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.8.3. Table 3. Top-ranking associations between common variants and Moderate-severe anxiety.

Genotype	Case*, N (%)	Control*, N (%)	Model	Case*, N (%)	Control*, N (%)	P value**	OR [95%-CI]	P value adj***
PTPRZ1 rs740965								
All cohort, N=200								
TT	10 (47.6)	147 (82.1)	**TT**	10 (47.6)	147 (82.1)	1.00E-03[d]	5.1 [1.98-12.9]	1.00E-03
TG	8 (38.1)	31 (17.3)	**TG+GG**	11 (52.4)	32 (17.9)			
GG	3 (14.3)	1 (0.6)						
MUC16 rs750295511								
All cohort, N=197								
AA	10 (50.0)	158 (89.3)	**AA**	10 (50.0)	158 (89.3)	7.20E-05[d]	8.3 [3.1-22.5]	3.10E-05
TA	10 (50.0)	19 (10.7)	**AT+TT**	10 (50.0)	19 (10.7)			
TT	0 (0.0)	0 (0.0)						
TNRC6C-AS1 rs2748431								
All cohort, N=167								
GG	12 (63.2)	135 (91.2)	**GG**	12 (63.2)	135 (91.2)	2.00E-03[d]	6.1 [2.0-18.1]	1.00E-03
GA	5 (26.3)	11 (7.4)	**GA+AA**	7 (36.8)	13 (8.8)			
AA	2 (10.5)	2 (1.4)						
APOL2 rs7285167								
Males only, N=95								
GG	5 (50.0)	77 (90.6)	**GG**	5 (50.0)	77 (90.6)	4.00E-03[d]	9.6 [2.3-40.5]	3.00E-03
GA	4 (40.0)	8 (9.4)	**GA+AA**	5 (50.0)	8 (9.4)			
AA	1 (10.0)	0 (0.0)						
C6orf165 rs61731441								
Males only, N=95								
GG	3 (30.0)	70 (82.4)	**GG**	3 (30.0)	70 (82.4)	1.00E-03[d]	10.9 [2.5-47.0]	2.00E-03
GA	6 (60.0)	15 (17.6)	**GA+AA**	7 (70.0)	15 (17.6)			
AA	1 (10.0)	0 (0.0)						

PTPRZ1: Protein Tyrosine Phosphatase Receptor Type Z1; MUC16: Mucin 16, Cell Surface Associated; TNRC6C-AS1: TNRC6C antisense RNA 1; APOL2: Apolipoprotein L2; C6orf165: Cilia And Flagella Associated Protein 206;
FDR: false discovery rate; OR: odds ratio.
*Participants with and without indicated complications are defined as cases and controls, respectively.

**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (d: Dominant).

***P value adj: p value from logistic regression adjusted for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.8.4. Table 4. Top-ranking associations between common variants and Moderate-severe depression.

Genotype	Case*, N (%)	Control*, N (%)	Model	Case*, N (%)	Control*, N (%)	P value**	OR [95%-CI]	P value adj***
EXO5 rs35672330								
Males only, N=94								
TT	4 (44.4)	81(94.2)	**TT**	4 (44.4)	81(94.2)	4.52E-04[d]	20.3 [4.1-99.8]	4.90E-04
TC	4 (44.4)	5 (5.8)	**TC+CC**	5 (55.6)	5 (5.8)			
CC	1 (11.2)	0 (0.0)						

EXO5: Exonuclease 5; FDR: false discovery rate; OR: odds ratio.

*Participants with and without indicated complications are defined as cases and controls, respectively.

**P values are calculated by chi-square or Fisher exact test, as appropriate. The most representative genetic model used is indicated (d: Dominant, r: Recessive).

***P value adj: p value from logistic regression adjusted for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.8.5. Table 5. Replication results, SJLIFE cohort, N=675.

a. Fisher exact test results (allelic test), replication cohort of SJLIFE.

Outcome	Category	Gene	Variant	EAF	cases			controls			Fisher exact test			
					n.2	n.1	n.0	n.2	n.1	n.0	OR	CI95l_OR	CI95u_OR	P value
Trail making test	All	*AK8*	rs17407084	0.05	0	12	140	1	52	470	0.76	0.36	1.45	0.452
	Males	*ZNF382*	rs61732180	0.26	7	36	40	14	94	146	1.36	0.90	2.04	0.125
Moderate-severe anxiety	All	*PTPRZ1*	rs740965	0.15	2	20	45	13	150	445	1.29	0.77	2.09	0.305
		TNRC6C-AS1	rs2748431	0.07	0	9	58	5	77	526	0.93	0.40	1.92	1.000
	Males	*APOL2*	rs7285167	0.08	0	2	32	0	52	254	0.33	0.04	1.29	0.153
		C6orf165	rs61731441	0.10	1	3	30	1	61	244	0.69	0.21	1.80	0.529
Moderate-severe depression	Males	*EXO5*	rs35672330	0.06	0	10	45	0	33	251	1.62	0.69	3.50	0.201

b. Logistic Regression results, replication cohort of SJLIFE.

Outcome	Category	Gene	Variant	OR	CI95l_OR	CI95u_OR	beta	CI95l	CI95u	LRT_P
Trail making test	All	*AK8*	rs17407084	0.75	0.36	1.44	-0.29	-1.02	0.36	0.394
	Males	*ZNF382*	rs61732180	1.40	0.92	2.15	0.34	-0.09	0.76	0.119
Moderate-severe anxiety	All	*PTPRZ1*	rs740965	1.52	0.90	2.50	0.42	-0.10	0.91	0.111
		TNRC6C-AS1	rs2748431	0.86	0.38	1.70	-0.16	-0.96	0.53	0.674
	Males	*APOL2*	rs7285167	0.24	0.04	0.94	-1.41	-3.33	-0.06	0.040
		C6orf165	rs61731441	0.71	0.23	1.83	-0.34	-1.49	0.60	0.500
Moderate-severe depression	Males	*EXO5*	rs35672330	2.18	0.91	4.97	0.78	-0.10	1.60	0.081

AK8:Adenylate Kinase 8 ATP-AMP Transphosphorylase 8; ZNF: Zinc Finger Protein 382; PTPRZ1: Protein Tyrosine Phosphatase Receptor Type Z1; TNRC6C-AS1: TNRC6C antisense RNA 1; APOL2: Apolipoprotein L2; C6orf165: Cilia And Flagella Associated Protein 206; EXO5: Exonuclease 5.

EAF: Effect allele frequency; n.2: 2 copies of the effect allele; n.1: 1 copy of the effect allele; n.0: 0 copies of the effect allele; OR: odds ratio for each additional copy of the effect allele; CI95l_OR: lower boundary of OR's 95% confidence interval; CI95u_OR: upper boundary of OR's 95% confidence interval; beta: beta coefficient for each additional copy of the effect allele; CI95l: lower boundary of beta's 95% confidence interval; CI95u: upper boundary of beta's 95% confidence interval; LRT_P: Likelihood ratio test p-value.

* Logistic regression analyses adjusted for continuous age at diagnosis, sex, continuous time between date of becoming a 5-year ALL survivor and date of test measurement, whether the survivor was treated with chemotherapy only versus chemotherapy plus radiation, and the top 20 principal components adjusting for genetic ancestry.

4.8.6. Table 6. Top-ranking associations of the rare variants identified through the SKAT-O test regarding the Moderate-severe anxiety in HR patients, N=124.

Gene	SNVs tested		MAF	P value SKAT-O	FDR-BH	11/ 12+22		11	12+22	P value Fisher test	OR [95% CI]
	position (hg19)	rs number									
PCMTD1	**chr8:52732981***	rs201786115	0.013	9.4E-06	0.03						**21.5 (4.4-105.9)***
	chr8:52733110		0.008								
	chr8:52733164	rs149898988	0.021			**Significant combination of collapsed variants****					
	chr8:52733209	rs202074278	0.022			91/9	Affected	5 (50.0%)	5 (50.0%)	4E-04	
	chr8:52733214*	rs200377849	0.025				Unaffected	86 (95.6%)	4 (4.4%)		
	chr8:52733227*		0.016								
CYP2W1	**chr7:1024855***	rs3735684	0.038	1.3E-05	0.03	**Individual contribution***					**21.9 (4.2-114.8)***
	chr7:1024874		0.005			85/8	Affected	6 (54.5%)	5 (45.5%)	4E-04	
	chr7:1024921		0.005				Unaffected	79 (96.3%)	3 (3.7%)		

PCMTD1: Protein-L-Isoaspartate (D-Aspartate) O-Methyltransferase Domain Containing 1; CYP2W1: Cytochrome P450 Family 2 Subfamily W Member 1; SNV: single nucleotide variation; MAF: minor allele frequency; *FDR-BH: Benjamini–Hochberg false discovery rate; OR: odds ratio; CI: confidence interval.*

*SNVs that are identified as the most important contributors to the association signal are highlighted.

**Collapsed variants (carriers of at least one of rare variants were included into the model, variants with missing values were excluded.

Genotypes were recoded as follows: 11-homozygote wild type; 12-heterozygote variant; 22-homozygote variant.

4.9. REFERENCES

1. Gaynon PS, Angiolillo AL, Carroll WL, et al. Long-term results of the children's cancer group studies for childhood acute lymphoblastic leukemia 1983-2002: a Children's Oncology Group Report. *Leukemia*. 2010;24(2):285-297.
2. Salzer WL, Devidas M, Carroll WL, et al. Long-term results of the pediatric oncology group studies for childhood acute lymphoblastic leukemia 1984-2001: a report from the children's oncology group. *Leukemia*. 2010;24(2):355-370.
3. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatr Clin North Am*. 2015;62(1):47-60.
4. Cheung YT, Krull KR. Neurocognitive outcomes in long-term survivors of childhood acute lymphoblastic leukemia treated on contemporary treatment protocols: A systematic review. *Neurosci Biobehav Rev*. 2015;53:108-120.
5. Krull KR, Brinkman TM, Li C, et al. Neurocognitive outcomes decades after treatment for childhood acute lymphoblastic leukemia: a report from the St Jude lifetime cohort study. *J Clin Oncol*. 2013;31(35):4407-4415.
6. Jacola LM, Edelstein K, Liu W, et al. Cognitive, behaviour, and academic functioning in adolescent and young adult survivors of childhood acute lymphoblastic leukaemia: a report from the Childhood Cancer Survivor Study. *Lancet Psychiatry*. 2016;3(10):965-972.
7. Zeltzer LK, Chen E, Weiss R, et al. Comparison of psychologic outcome in adult survivors of childhood acute lymphoblastic leukemia versus sibling controls: a cooperative Children's Cancer Group and National Institutes of Health study. *J Clin Oncol*. 1997;15(2):547-556.
8. Giedd JN, Blumenthal J, Jeffries NO, et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*. 1999;2(10):861-863.
9. Lenroot RK, Giedd JN. Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neurosci Biobehav Rev*. 2006;30(6):718-729.
10. Konrad K, Firk C, Uhlhaas PJ. Brain development during adolescence: neuroscientific insights into this developmental period. *Dtsch Arztebl Int*. 2013;110(25):425-431.
11. Pääkkö E, Vainionpää L, Lanning M, Laitinen J, Pyhtinen J. White matter changes in children treated for acute lymphoblastic leukemia. *Cancer*. 1992;70(11):2728-2733.
12. Vázquez E, Lucaya J, Castellote A, et al. Neuroimaging in Pediatric Leukemia and Lymphoma: Differential Diagnosis. *RadioGraphics*. 2002;22(6):1411-1428.
13. Campbell LK, Scaduto M, Sharp W, et al. A meta-analysis of the neurocognitive sequelae of treatment for childhood acute lymphocytic leukemia. *Pediatric Blood & Cancer*. 2007;49(1):65-73.
14. Reddick WE, Shan ZY, Glass JO, et al. Smaller white-matter volumes are associated with larger deficits in attention and learning among long-term survivors of acute lymphoblastic leukemia. *Cancer*. 2006;106(4):941-949.
15. Campbell LK, Scaduto M, Sharp W, et al. A meta-analysis of the neurocognitive sequelae of treatment for childhood acute lymphocytic leukemia. *Pediatr Blood Cancer*. 2007;49(1):65-73.
16. Iyer NS, Balsamo LM, Bracken MB, Kadan-Lottick NS. Chemotherapy-only treatment effects on long-term neurocognitive functioning in childhood ALL survivors: a review and meta-analysis. *Blood*. 2015;126(3):346-353.
17. Ashford J, Schoffstall C, Reddick WE, et al. Attention and working memory abilities in children treated for acute lymphoblastic leukemia. *Cancer*. 2010;116(19):4638-4645.

18. Kingma A, van Dommelen RI, Mooyaart EL, Wilmink JT, Deelman BG, Kamps WA. Slight cognitive impairment and magnetic resonance imaging abnormalities but normal school levels in children treated for acute lymphoblastic leukemia with chemotherapy only. *J Pediatr.* 2001;139(3):413-420.
19. Lofstad GE, Reinfjell T, Hestad K, Diseth TH. Cognitive outcome in children and adolescents treated for acute lymphoblastic leukaemia with chemotherapy only. *Acta Paediatr.* 2009;98(1):180-186.
20. Rodgers J, Marckus R, Kearns P, Windebank K. Attentional ability among survivors of leukaemia treated without cranial irradiation. *Arch Dis Child.* 2003;88(2):147-150.
21. Daams M, Schuitema I, van Dijk BW, et al. Long-term effects of cranial irradiation and intrathecal chemotherapy in treatment of childhood leukemia: a MEG study of power spectrum and correlated cognitive dysfunction. *BMC Neurol.* 2012;12:84.
22. Edelmann MN, Krull KR, Liu W, et al. Diffusion tensor imaging and neurocognition in survivors of childhood acute lymphoblastic leukaemia. *Brain.* 2014;137(Pt 11):2973-2983.
23. Mennes M, Stiers P, Vandenbussche E, et al. Attention and information processing in survivors of childhood acute lymphoblastic leukemia treated with chemotherapy only. *Pediatr Blood Cancer.* 2005;44(5):478-486.
24. Carey ME, Haut MW, Reminger SL, Hutter JJ, Theilmann R, Kaemingk KL. Reduced frontal white matter volume in long-term childhood leukemia survivors: a voxel-based morphometry study. *AJNR Am J Neuroradiol.* 2008;29(4):792-797.
25. Schultz KA, Ness KK, Whitton J, et al. Behavioral and social outcomes in adolescent survivors of childhood cancer: a report from the childhood cancer survivor study. *J Clin Oncol.* 2007;25(24):3649-3656.
26. Zeltzer LK, Recklitis C, Buchbinder D, et al. Psychological status in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. *J Clin Oncol.* 2009;27(14):2396-2404.
27. Michel G, Rebholz CE, von der Weid NX, Bergstraesser E, Kuehni CE. Psychological distress in adult survivors of childhood cancer: the Swiss Childhood Cancer Survivor study. *J Clin Oncol.* 2010;28(10):1740-1748.
28. Stuber ML, Meeske KA, Krull KR, et al. Prevalence and predictors of posttraumatic stress disorder in adult survivors of childhood cancer. *Pediatrics.* 2010;125(5):e1124-1134.
29. Baytan B, Asut C, Cirpan Kantarcioglu A, Sezgin Evim M, Gunes AM. Health-Related Quality of Life, Depression, Anxiety, and Self-Image in Acute Lymphocytic Leukemia Survivors. *Turk J Haematol.* 2016;33(4):326-330.
30. Pepin AJ, Lippe S, Krajinovic M, et al. How to interpret high levels of distress when using the Distress Thermometer in the long-term follow-up clinic? A study with Acute Lymphoblastic Leukemia survivors. *Pediatr Hematol Oncol.* 2017;34(3):133-137.
31. Oancea SC, Brinkman TM, Ness KK, et al. Emotional distress among adult survivors of childhood cancer. *J Cancer Surviv.* 2014;8(2):293-303.
32. Jain N, Brouwers P, Okcu MF, Cirino PT, Krull KR. Sex-specific attention problems in long-term survivors of pediatric acute lymphoblastic leukemia. *Cancer.* 2009;115(18):4238-4245.
33. Petrykey K, Lippé S, Robaey P, et al. Influence of genetic factors on long-term treatment related neurocognitive complications, and on anxiety and depression in survivors of childhood acute lymphoblastic leukemia: The Petale study. *PLoS One.* 2019;14(6):e0217314.

34. Marcoux S, Drouin S, Laverdiere C, et al. The PETALE study: Late adverse effects and biomarkers in childhood acute lymphoblastic leukemia survivors. *Pediatr Blood Cancer.* 2017;64(6).
35. Moghrabi A, Levy DE, Asselin B, et al. Results of the Dana-Farber Cancer Institute ALL Consortium Protocol 95-01 for children with acute lymphoblastic leukemia. *Blood.* 2007;109(3):896-904.
36. Silverman LB, Stevenson KE, O'Brien JE, et al. Long-term results of Dana-Farber Cancer Institute ALL Consortium protocols for children with newly diagnosed acute lymphoblastic leukemia (1985-2000). *Leukemia.* 2010;24(2):320-334.
37. Krull KR, Okcu MF, Potter B, et al. Screening for neurocognitive impairment in pediatric cancer long-term survivors. *J Clin Oncol.* 2008;26(25):4138-4143.
38. Delis DC KE, Kramer JH. . Delis-Kaplan executive function system (D-KEFS). *Psychological Corporation.* 2001.
39. D W. *Wechsler adult intelligence scale - Fourth Edition (WAIS-IV).* San Antonio, TX2008.
40. Yochim B, Baldo J, Nelson A, Delis DC. D-KEFS Trail Making Test performance in patients with lateral prefrontal cortex lesions. *J Int Neuropsychol Soc.* 2007;13(4):704-709.
41. Delis DC, Kramer JH, Kaplan E, Holdnack J. Reliability and validity of the Delis-Kaplan Executive Function System: an update. *J Int Neuropsychol Soc.* 2004;10(2):301-303.
42. Erdodi LA, Abeare CA, Lichtenstein JD, et al. Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychol Assess.* 2017;29(2):148-157.
43. Harrison AG, Armstrong IT, Harrison LE, Lange RT, Iverson GL. Comparing Canadian and American normative scores on the Wechsler Adult Intelligence Scale-Fourth Edition. *Arch Clin Neuropsychol.* 2014;29(8):737-746.
44. Boulet-Craig A, Robaey P, Laniel J, et al. DIVERGT screening procedure predicts general cognitive functioning in adult long-term survivors of pediatric acute lymphoblastic leukemia: A PETALE study. *Pediatr Blood Cancer.* 2018;65(9):e27259.
45. Derogatis LR. *Brief Symptom Inventory (BSI) 18: Administration, scoring, and procedures manual.* Minneapolis, MN: NCS Pearson; 2000.
46. Beck JS BA, Jolly JB, Steer RA. . *Beck Youth Inventories-Second Edition for children and adolescent's manual.* San Antonio, TX: PsychCorp; 2005.
47. Brinkman TM, Zhu L, Zeltzer LK, et al. Longitudinal patterns of psychological distress in adult survivors of childhood cancer. *Br J Cancer.* 2013;109(5):1373-1381.
48. Recklitis CJ, Parsons SK, Shih MC, Mertens A, Robison LL, Zeltzer L. Factor structure of the brief symptom inventory--18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. *Psychol Assess.* 2006;18(1):22-32.
49. Gianinazzi ME, Rueegg CS, Wengenroth L, et al. Adolescent survivors of childhood cancer: are they vulnerable for psychological distress? *Psychooncology.* 2013;22(9):2051-2058.
50. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80(1):99-103.
51. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.

52. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 2006;7:61-80.
53. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
54. Rohlfs RV, Weir BS. Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics.* 2008;180(3):1609-1616.
55. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics.* 2016;24(8):1202-1205.
56. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics.* 2014;15(5):335-346.
57. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440-9445.
58. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1-2):279-284.
59. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. *Genetics.* 2005;171(2):783-790.
60. Whole genome association analysis toolset. http://zzz.bwh.harvard.edu/plink/index.shtml. Accessed.
61. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
62. Buizer AI, de Sonneville LM, van den Heuvel-Eibrink MM, Veerman AJ. Chemotherapy and attentional dysfunction in survivors of childhood acute lymphoblastic leukemia: effect of treatment intensity. *Pediatr Blood Cancer.* 2005;45(3):281-290.
63. Waber DP, Tarbell NJ, Kahn CM, Gelber RD, Sallan SE. The relationship of sex and treatment modality to neuropsychologic outcome in childhood acute lymphoblastic leukemia. *J Clin Oncol.* 1992;10(5):810-817.
64. Lee S, Abecasis Gonçalo R, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. *The American Journal of Human Genetics.* 2014;95(1):5-23.
65. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012;91(2):224-237.
66. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics.* 2012;13(4):762-775.
67. CRAN. Package SKAT: SNP-Set (Sequence) Kernel Association Test. https://cran.r-project.org/web/packages/SKAT/index.html. Accessed.
68. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet Epidemiol.* 2011;35 Suppl 1(Suppl 1):S12-17.
69. Sung YJ, Korthauer KD, Swartz MD, Engelman CD. Methods for collapsing multiple rare variants in whole-genome sequence data. *Genet Epidemiol.* 2014;38 Suppl 1(0 1):S13-20.
70. The Genome Aggregation Database (gnomAD). In.
71. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.

72. Schalet BD, Pilkonis PA, Yu L, et al. Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. *J Clin Epidemiol.* 2016;73:119-127.
73. Ecco G, Imbeault M, Trono D. KRAB zinc finger proteins. *Development.* 2017;144(15):2719-2729.
74. Tao YF, Hu SY, Lu J, et al. Zinc finger protein 382 is downregulated by promoter hypermethylation in pediatric acute myeloid leukemia patients. *Int J Mol Med.* 2014;34(6):1505-1515.
75. Chen S, Xiao Z, Zhou J, et al. ZNF382: A transcription inhibitor down-regulated in multiple tumors due to promoter methylation. *Clin Chim Acta.* 2020;500:220-225.
76. Dresselhaus EC, Meffert MK. Cellular Specificity of NF-κB Function in the Nervous System. *Frontiers in Immunology.* 2019;10(1043).
77. Jones SV, Kounatidis I. Nuclear Factor-Kappa B and Alzheimer Disease, Unifying Genetic and Environmental Risk Factors from Cell to Humans. *Frontiers in Immunology.* 2017;8(1805).
78. Jacola LM, Krull KR, Pui CH, et al. Longitudinal Assessment of Neurocognitive Outcomes in Survivors of Childhood Acute Lymphoblastic Leukemia Treated on a Contemporary Chemotherapy Protocol. *J Clin Oncol.* 2016;34(11):1239-1247.
79. Boulet-Craig A, Robaey P, Barlaam F, et al. Visual short-term memory activation patterns in adult survivors of childhood acute lymphoblastic leukemia. *Cancer.* 2019;125(20):3639-3648.
80. Ma MW, Wang J, Zhang Q, et al. NADPH oxidase in brain injury and neurodegenerative disorders. *Molecular Neurodegeneration.* 2017;12(1):7.
81. Salim S. Oxidative Stress and the Central Nervous System. *J Pharmacol Exp Ther.* 2017;360(1):201-205.
82. Areti A, Yerra VG, Naidu V, Kumar A. Oxidative stress and nerve damage: role in chemotherapy induced peripheral neuropathy. *Redox Biol.* 2014;2:289-295.
83. Taylor OA, Hockenberry MJ, McCarthy K, et al. Evaluation of Biomarkers of Oxidative Stress and Apoptosis in Patients With Severe Methotrexate Neurotoxicity: A Case Series. *J Pediatr Oncol Nurs.* 2015;32(5):320-325.
84. Mattson MP, Meffert MK. Roles for NF-kappaB in nerve cell survival, plasticity, and disease. *Cell Death Differ.* 2006;13(5):852-860.
85. Liu T, Zhang L, Joo D, Sun S-C. NF-κB signaling in inflammation. *Signal Transduction and Targeted Therapy.* 2017;2(1):17023.
86. Shi P, Yang A, Zhao Q, Chen Z, Ren X, Dai Q. A Hypothesis of Gender Differences in Self-Reporting Symptom of Depression: Implications to Solve Under-Diagnosis and Under-Treatment of Depression in Males. *Frontiers in Psychiatry.* 2021;12.
87. Sparks JL, Kumar R, Singh M, Wold MS, Pandita TK, Burgers PM. Human exonuclease 5 is a novel sliding exonuclease required for genome stability. *J Biol Chem.* 2012;287(51):42773-42783.
88. Ali S, Zhang Y, Zhou M, et al. Functional deficiency of DNA repair gene EXO5 results in androgen-induced genomic instability and prostate tumorigenesis. *Oncogene.* 2020;39(6):1246-1259.
89. Breton JM, Barraza M, Hu KY, Frias SJ, Long KLP, Kaufer D. Juvenile exposure to acute traumatic stress leads to long-lasting alterations in grey matter myelination in adult female but not male rats. *Neurobiology of Stress.* 2021;14:100319.

90. Heim C, Nemeroff CB. The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol Psychiatry.* 2001;49(12):1023-1039.
91. Anestin AS, Lippé S, Robaey P, et al. Psychological risk in long-term survivors of childhood acute lymphoblastic leukemia and its association with functional health status: A PETALE cohort study. *Pediatr Blood Cancer.* 2018;65(11):e27356.
92. Chantziara S, Musoro J, Rowsell AC, et al. Quality of life of long-term childhood acute lymphoblastic leukemia survivors: Comparison with healthy controls. *Psychooncology.* 2022;31(12):2159-2168.
93. Wu G, Feder A, Cohen H, et al. Understanding resilience. *Frontiers in Behavioral Neuroscience.* 2013;7(10).
94. Musliner KL, Mortensen PB, McGrath JJ, et al. Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA Psychiatry.* 2019;76(5):516-525.
95. Power RA, Tansey KE, Buttenschøn HN, et al. Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biol Psychiatry.* 2017;81(4):325-335.
96. Schulze TG, Akula N, Breuer R, et al. Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World J Biol Psychiatry.* 2014;15(3):200-208.
97. Panayiotou C, Solaroli N, Xu Y, Johansson M, Karlsson A. The characterization of human adenylate kinases 7 and 8 demonstrates differences in kinetic parameters and structural organization among the family of adenylate kinase isoenzymes. *Biochemical Journal.* 2011;433(3):527-534.
98. Robinson G, Parker M, Kranenburg TA, et al. Novel mutations target distinct subgroups of medulloblastoma. *Nature.* 2012;488(7409):43-48.
99. Lesch K-P, Timmesfeld N, Renner TJ, et al. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *Journal of Neural Transmission.* 2008;115(11):1573-1585.
100. Hawi Z, Cummins TD, Tong J, et al. The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry.* 2015;20(3):289-297.
101. Ionescu MI. Adenylate Kinase: A Ubiquitous Enzyme Correlated with Medical Conditions. *The Protein Journal.* 2019;38(2):120-133.
102. Harroch S, Furtado GC, Brueck W, et al. A critical role for the protein tyrosine phosphatase receptor type Z in functional recovery from demyelinating lesions. *Nat Genet.* 2002;32(3):411-414.
103. Lamprianou S, Chatzopoulou E, Thomas JL, Bouyain S, Harroch S. A complex between contactin-1 and the protein tyrosine phosphatase PTPRZ controls the development of oligodendrocyte precursor cells. *Proc Natl Acad Sci U S A.* 2011;108(42):17498-17503.
104. Tanga N, Kuboyama K, Kishimoto A, et al. Behavioral and neurological analyses of adult mice carrying null and distinct loss-of-receptor function mutations in protein tyrosine phosphatase receptor type Z (PTPRZ). *PLoS One.* 2019;14(6):e0217880.
105. Sacchet MD, Gotlib IH. Myelination of the brain in Major Depressive Disorder: An in vivo quantitative magnetic resonance imaging study. *Scientific Reports.* 2017;7(1):2200.

106. Haridas D, Ponnusamy MP, Chugh S, Lakshmanan I, Seshacharyulu P, Batra SK. MUC16: molecular analysis and its functional implications in benign and malignant conditions. *Faseb j.* 2014;28(10):4183-4199.
107. Lakshmanan I, Ponnusamy MP, Das S, et al. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. *Oncogene.* 2012;31(7):805-817.
108. Haridas D, Chakraborty S, Ponnusamy MP, et al. Pathobiological implications of MUC16 expression in pancreatic cancer. *PLoS One.* 2011;6(10):e26839.
109. Kanwal M, Ding XJ, Song X, Zhou GB, Cao Y. MUC16 overexpression induced by gene mutations promotes lung cancer cell growth and invasion. *Oncotarget.* 2018;9(15):12226-12239.
110. Streppel MM, Vincent A, Mukherjee R, et al. Mucin 16 (cancer antigen 125) expression in human tissues and cell lines and correlation with clinical outcome in adenocarcinomas of the pancreas, esophagus, stomach, and colon. *Hum Pathol.* 2012;43(10):1755-1763.
111. Darby MM, Yolken RH, Sabunciyan S. Consistently altered expression of gene sets in postmortem brains of individuals with major psychiatric disorders. *Transl Psychiatry.* 2016;6(9):e890.
112. Villegas VE, Zaphiropoulos PG. Neighboring Gene Regulation by Antisense Long Non-Coding RNAs. *International Journal of Molecular Sciences.* 2015;16(2):3251-3266.
113. Jovčevska I, Videtič Paska A. Neuroepigenetics of psychiatric disorders: Focus on lncRNA. *Neurochemistry International.* 2021;149:105140.
114. Muhanhali D, Zhai T, Jiang J, Ai Z, Zhu W, Ling Y. Long Non-coding Antisense RNA TNRC6C-AS1 Is Activated in Papillary Thyroid Cancer and Promotes Cancer Progression by Suppressing TNRC6C Expression. *Frontiers in Endocrinology.* 2018;9(360).
115. An T, Zhang J, Ma Y, et al. Relationships of Non-coding RNA with diabetes and depression. *Sci Rep.* 2019;9(1):10707.
116. Wei CW, Luo T, Zou SS, Wu AS. The Role of Long Noncoding RNAs in Central Nervous System and Neurodegenerative Diseases. *Front Behav Neurosci.* 2018;12:175.
117. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics.* 2018;50(8):1112-1121. doi:10.1038/s41588-018-0147-3. Accessed 2018/07//.
118. Vanhollebeke B, Pays E. The function of apolipoproteins L. *Cellular and Molecular Life Sciences CMLS.* 2006;63(17):1937-1944.
119. Elliott DA, Weickert CS, Garner B. Apolipoproteins in the brain: implications for neurological and psychiatric disorders. *Clin Lipidol.* 2010;51(4):555-573.
120. Mimmack ML, Ryan M, Baba H, et al. Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proc Natl Acad Sci U S A.* 2002;99(7):4680-4685.
121. Takahashi S, Cui YH, Han YH, et al. Association of SNPs and haplotypes in APOL1, 2 and 4 with schizophrenia. *Schizophr Res.* 2008;104(1-3):153-164.
122. Miller JRC, Lo KK, Andre R, et al. RNA-Seq of Huntington's disease patient myeloid cells reveals innate transcriptional dysregulation associated with proinflammatory pathway activation. *Human Molecular Genetics.* 2016;25(14):2893-2904.
123. van Dam TJP, Kennedy J, van der Lee R, et al. CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS One.* 2019;14(5):e0216705.

124. Lu B, Jiang R, Xie B, Wu W, Zhao Y. Fusion genes in gynecologic tumors: the occurrence, molecular mechanism and prospect for therapy. *Cell Death & Disease.* 2021;12(8):783.
125. Kondkar AA. Updates on Genes and Genetic Mechanisms Implicated in Primary Angle-Closure Glaucoma. *Appl Clin Genet.* 2021;14:89-112.
126. Gervasini G, Murillo SGd, Ladero JM, Agúndez JA. CYP2W1 variant alleles in Caucasians and association of the CYP2W1 G541A (Ala181Thr) polymorphism with increased colorectal cancer risk. *Pharmacogenomics.* 2010;11(7):919-925.
127. Chung FF, Mai CW, Ng PY, Leong CO. Cytochrome P450 2W1 (CYP2W1) in Colorectal Cancers. *Curr Cancer Drug Targets.* 2016;16(1):71-78.
128. Zemanová N, Anzenbacher P, Zapletalová I, et al. The role of the microbiome and psychosocial stress in the expression and activity of drug metabolizing enzymes in mice. *Scientific Reports.* 2020;10(1):8529.
129. Konstandi M, Johnson EO, Lang MA. Consequences of psychophysiological stress on cytochrome P450-catalyzed drug metabolism. *Neuroscience & Biobehavioral Reviews.* 2014;45:149-167.

4.10. SUPPLEMENTAL MATERIAL

4.10.1. Supplemental table 1. Trail making test, PLINK results' summary.

All cohort, N=229								
Gene	**rs number**	**position (hg19)**	**P value allelic test*(Chi2 or Fisher)**	**FDR**	**OR allelic [95%CI]**	**P value adj****	**Allelic frequencies**	
							Affected	**Unaffected**
AK8	rs17407084	chr9:135730257	9.98E-08	0.0005	7.17 [3.2-16.3]	1E-04	11/29	21/397
Females only, N=117								
KIR3DL1	rs45542639	chr19:55340906	2.12E-07	0.0011	11.5 [3.86-34.5]	0.0010	10/6	25/173
PPARG	rs1801282	chr3:12393125	6.49E-07	0.0017	10 [3.5-28.86]	0.0004	8/10	16/200
PLAUR	rs4760	chr19:44153100	1.06E-06	0.0019	9.19 [3.29-25.69]	0.0017	9/9	21/193
AK8	rs17407084	chr9:135730257	3.56E-06	0.0048	10.3 [3.2-33.11]	0.0027	6/12	10/206
Males only, N=112								
DUOXA1	rs61751061	chr15:45409732	6.14E-06	0.0111	9.3 [3.03-28.4]	0.0009	7/15	9/179
DUOX2	rs269868	chr15:45392075	6.22E-06	0.0111	8.96 [2.98-26.9]	0.0008	7/15	10/179

ZNF382	rs61732180	chr19:37118439	3.83E-05	0.05	5.85 [2.3-14.6]	0.0013	13/9	40/162
Standard risk, N=105								
AK8	rs17407084	chr9:135730257	1.92E-08	0.0001	17.62[4.9-62.96]	0.0005	6/8	8/188

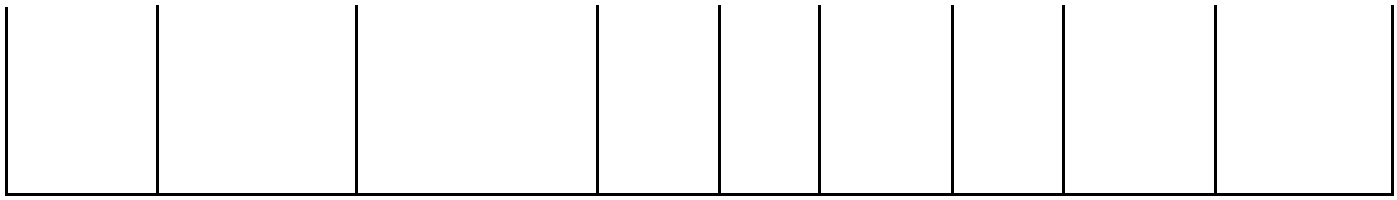
AK8: Adenylate Kinase 8 ATP-AMP Transphosphorylase 8; KIR3DL1: Killer Cell Immunoglobulin Like Receptor; PPARG: Peroxisome Proliferator Activated Receptor Gamma; PLAUR: Urokinase Plasminogen Activator Surface Receptor; DUOXA1: Dual Oxidase Maturation Factor 1; DUOX2: Dual Oxidase 2; ZNF: Zinc Finger Protein 382.

*Case/control association test which is based on comparing allele frequencies between cases and controls, chi-square p values.

**P value adj: p value from logistic regression in which each genetic variable is adjusted separately for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.10.2. Supplemental table 2. Moderate-severe anxiety, PLINK results' summary.

All cohort, N=229								
Gene	**rs number**	**position (hg19)**	**P value allelic test (Chi2 or Fisher)**	**FDR**	**OR allelic [95%CI]**	**P value adj***	**Allelic frequencies**	
							Affected	**Unaffected**
PTPRZ1	rs740965	chr7:121513561	4.40E-06	0.02	4.9 [2.4-10.3]	9.22E-05	14/28	33/325
MUC16	rs750295511	chr19:9009325	6.57E-06	0.02	5.9 [2.5-13.8]	3.78E-05	10/30	19/335
TNRC6C-AS1	rs2748431	chr17:76105754	2.87E-05	0.05	5.8 [2.3-14.5]	1.17E-03	9/29	15/281
Males only, N=112								
MICB	rs3134900	chr6:31473957	6.87E-07	0.004	12.4 [3.8-40.7]	3.77E-03	7/13	7/161
ARL16	rs8066889	chr17:79650828	2.57E-05	0.046	8.6 [2.7-26.7]	6.80E-03	7/13	9/143
APOL2	rs7285167	chr22:36623920	4.21E-05	0.047	8.7 [2.6-28.6]	1.38E-03	6/14	8/162
C6orf165	rs61731441	chr6:88125542	5.27E-05	0.047	6.9 [2.4-19.5]	1.45E-03	8/12	15/155



PTPRZ1: Protein Tyrosine Phosphatase Receptor Type Z1; MUC16: Mucin 16, Cell Surface Associated; TNRC6C-AS1: TNRC6C antisense RNA 1; MICB: MHC Class I Polypeptide-Related Sequence B; ARL16: ADP Ribosylation Factor Like GTPase 16; APOL2: Apolipoprotein L2; C6orf165: Cilia And Flagella Associated Protein 206.

*Case/control association test which is based on comparing allele frequencies between cases and controls, chi-square p values.

**P value adj: p value from logistic regression in which each genetic variant is adjusted for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.10.3. Supplemental table 3. Moderate-severe depression, PLINK results' summary.

Males only, N=112								
Gene	rs number	position (hg19)	P value allelic test (Chi2 or Fisher)	FDR	OR allelic [95%CI]	P value adj*	Allelic frequencies	
							Affected	Unaffected
EXO5	rs35672330	chr1:40980668	1.45E-07	7.78E-04	16.7 [4.4-62.7]	7.00E-04	6/12	5/167

EXO5: Exonuclease 5.

*Case/control association test which is based on comparing allele frequencies between cases and controls, chi-square p values.

**P value adj: p value from logistic regression adjusted for age at diagnosis, sex, time since the end of treatment, protocol, and treatment variable with the following scores: patients who received chemotherapy only =1, patients who received chemotherapy and radiotherapy =2 (categorical variable).

4.10.4. Supplemental table 4. Multivariate regression analysis of the top-ranking associations identified in female participants tested for the Trail making test.

	B	S.E.	Wald	df	P value
KIR3DL1_chr19_55340906	14.166	9.418	2.262	1	0.133
PPARG_chr3_12393125	12.836	8.815	2.12	1	0.145
PLAUR_chr19_44153100	8.485	5.493	2.386	1	0.122
Age at diagnosis	-1.333	1.076	1.535	1	0.215
Time between end of treatment and evaluation	-0.113	0.383	0.088	1	0.767
DFCI: 87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5	8.841	6.549	1.822	1	0.177
Treatment code	3.769	2.707	1.939	1	0.164
Constant	-50.899	34.939	2.122	1	0.145

KIR3DL1: Killer Cell Immunoglobulin Like Receptor; PPARG: Peroxisome Proliferator Activated Receptor Gamma; PLAUR: Urokinase Plasminogen Activator Surface Receptor; S.E.: standard error; df: degrees of freedom; DFCI, Dana-Farber Cancer Institute.

Multivariate analysis in which all genetic variables associated with a given outcome are entered in the model in one step with non-genetic covariates. Each model thus includes genetic variables in their dominant model and non-genetic co-variables: age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment (patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy).

4.10.5. Supplemental table 5. Multivariate regression analysis of the top-ranking associations identified in male participants tested for the Trail making test.

	B	S.E.	Wald	df	P value
ZNF382_chr19_37118439	4.082	1.148	12.643	1	**0.0004**
DUOXA1_chr15_45409732	1.745	2.373	0.541	1	0.462
DUOX2_chr15_45392075	1.426	2.421	0.347	1	0.556
Age at diagnosis	-0.061	0.106	0.327	1	0.567
Time between end of treatment and evaluation	0.21	0.234	0.804	1	0.37
DFCI: 87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5	0.368	1.008	0.134	1	0.715
Treatment code	-0.649	1.094	0.352	1	0.553
Constant	-7.158	6.255	1.31	1	0.252

DUOXA1: Dual Oxidase Maturation Factor 1; DUOX2: Dual Oxidase 2; ZNF: Zinc Finger Protein 382; S.E., standard error; df: degrees of freedom; DFCI, Dana-Farber Cancer Institute.

Multivariate analysis in which all genetic variables associated with a given outcome are entered in the model in one step with non-genetic covariates. Each model thus includes genetic variables in their dominant model and non-genetic co-variables: age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment (patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy).

4.10.6. Supplemental table 6. Multivariate regression analysis of the top-ranking associations identified in the whole cohort of survivors evaluated for the Moderate-severe anxiety.

	B	S.E.	Wald	df	P value
PTPRZ1_chr7_121513561	1.359	0.579	5.511	1	**0.019**
MUC16_chr19_9009325	2.359	0.655	12.96	1	**0.0003**
TNRC6C_AS1_chr17_76105754	1.929	0.668	8.345	1	**0.004**
Age at diagnosis	0.033	0.067	0.236	1	0.627
Time between end of treatment and evaluation	-0.125	0.164	0.58	1	0.447
DFCI: 87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5	-0.192	0.76	0.064	1	0.8
Treatment code	-0.429	0.623	0.474	1	0.491
Constant	-1.374	4.515	0.093	1	0.761

PTPRZ1: Protein Tyrosine Phosphatase Receptor Type Z1; MUC16: Mucin 16, Cell Surface Associated; TNRC6C-AS1: TNRC6C antisense RNA 1; S.E., standard error; df: degrees of freedom; DFCI, Dana-Farber Cancer Institute.

Multivariate analysis in which all genetic variables associated with a given outcome are entered in the model in one step with non-genetic covariates. Each model thus includes genetic variables in their dominant model and non-genetic co-variables: age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment (patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy).

4.10.7. Supplemental table 7. Multivariate regression analysis of the top-ranking associations identified in male participants evaluated for the Moderate-severe anxiety.

	B	S.E.	Wald	df	P value
MICB_chr6_31473957	1.924	1.098	3.072	1	0.08
ARL16_chr17_79650828	3.828	2.127	3.239	1	0.072
APOL2_chr22_36623920	4.298	1.805	5.67	1	**0.017**
C6orf165_chr6_88125542	3.237	1.37	5.584	1	**0.018**
Age at diagnosis	0.05	0.135	0.138	1	0.71
Time between end of treatment and evaluation	0.062	0.27	0.053	1	0.818
87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5	0.594	1.36	0.191	1	0.662
Treatment code	1.797	1.515	1.407	1	0.236
Constant	-9.65	8.691	1.233	1	0.267

MICB: MHC Class I Polypeptide-Related Sequence B; ARL16: ADP Ribosylation Factor Like GTPase 16; APOL2: Apolipoprotein L2; C6orf165: Cilia And Flagella Associated Protein 206; S.E., standard error; df: degrees of freedom; DFCI, Dana-Farber Cancer Institute.

Multivariate analysis in which all genetic variables associated with a given outcome are entered in the model in one step with non-genetic covariates. Each model thus includes genetic variables in their dominant model and non-genetic co-variables: age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment (patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy).

4.10.8. Supplemental table 8.

a. Fisher exact test additional results (allelic test), replication cohort of SJLIFE.

Outcome	Category	Gene	Variant	EAF	cases			controls			Fisher exact test			
					n.2	n.1	n.0	n.2	n.1	n.0	OR	CI95l	CI95u	P value
Trail making test	All	*ZNF382*	rs61732180	0.23	11	64	77	21	187	315	1.41	1.04	1.90	**0.025**
Moderate-severe anxiety	All	*APOL2*	rs7285167	0.08	0	4	63	1	100	507	0.34	0.09	0.91	**0.026**

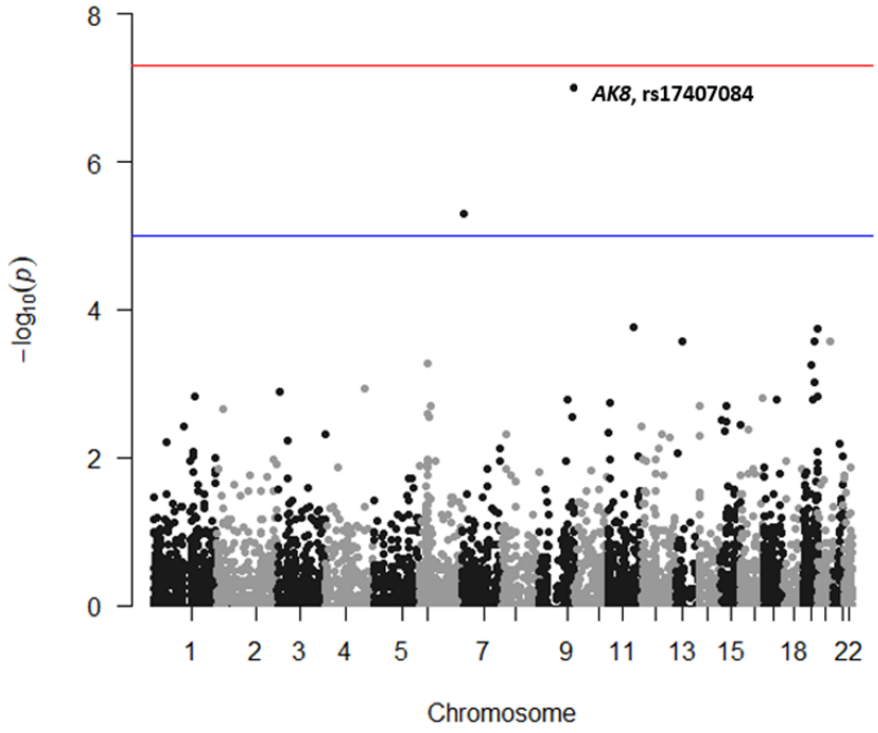
b. Logistic Regression* additional results, replication cohort of SJLIFE.

Outcome	Category	Gene	Variant	OR	CI95l_OR	CI95u_OR	beta	CI95l	CI95u	LRT_P
Trail making test	All	*ZNF382*	rs61732180	1.44	1.05	1.98	0.37	0.05	0.69	**0.025**
Moderate-severe anxiety	All	*APOL2*	rs7285167	0.37	0.11	0.96	-0.99	-2.22	-0.05	**0.039**

ZNF: Zinc Finger Protein 382; APOL2: Apolipoprotein L2.

EAF: Effect allele frequency; n.2: 2 copies of the effect allele; n.1: 1 copy of the effect allele; n.0: 0 copies of the effect allele; OR: odds ratio for each additional copy of the effect allele; CI95l_OR: lower boundary of OR's 95% confidence interval; CI95u_OR: upper boundary of OR's 95% confidence interval; beta: beta coefficient for each additional copy of the effect allele; CI95l: lower boundary of beta's 95% confidence interval; CI95u: upper boundary of beta's 95% confidence interval; LRT_P: Likelihood ratio test p-value.

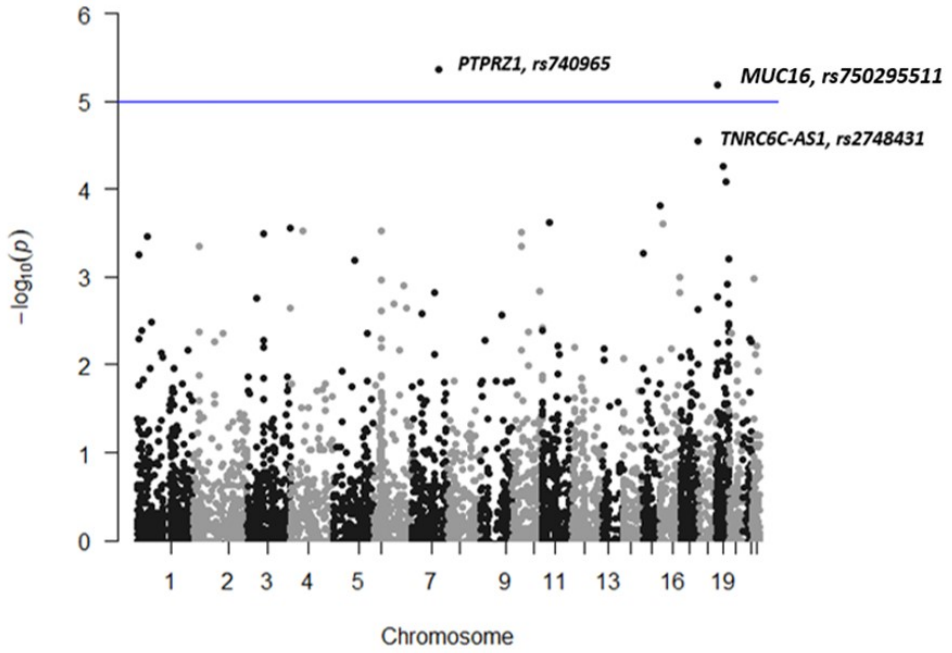
* Logistic regression analyses adjusted for continuous age at diagnosis, sex, continuous time between date of becoming a 5-year ALL survivor and date of test measurement, whether the survivor was treated with chemotherapy only versus chemotherapy plus radiation, and the top 20 principal components adjusting for genetic ancestry.



4.10.9. Supplemental Figure 1. P value distribution of the common variants tested for Trail making test represented as a Manhattan plot, all cohort, N=229.

Manhattan plot was created to display the statistical significance between common genetic variants used in the exome-wide association study and neurocognitive complications as measured by Trail making test and ranked according to their associated P value.

AK8: Adenylate Kinase 8.

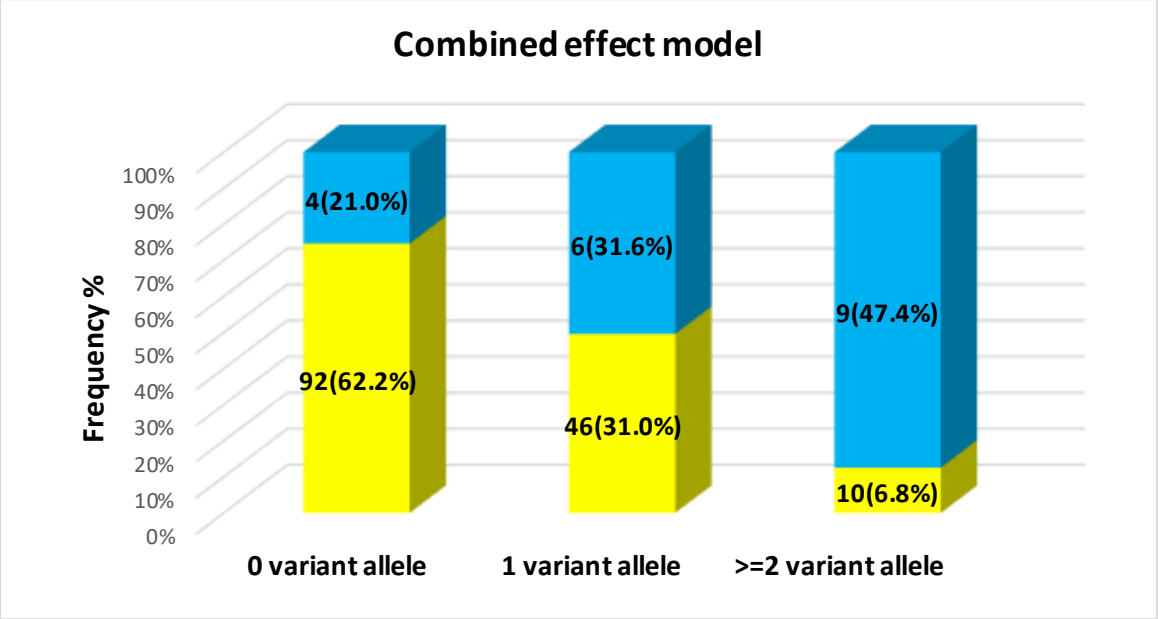


4.10.10. Supplemental Figure 2. P value distribution of the common variants tested for Moderate-Severe anxiety represented as a Manhattan plot, all cohort, N=229.

Manhattan plot was created to display the statistical significance between common genetic variants used in the exome-wide association study and mood disorders as measured by Moderate-Severe anxiety and ranked according to their associated P value.

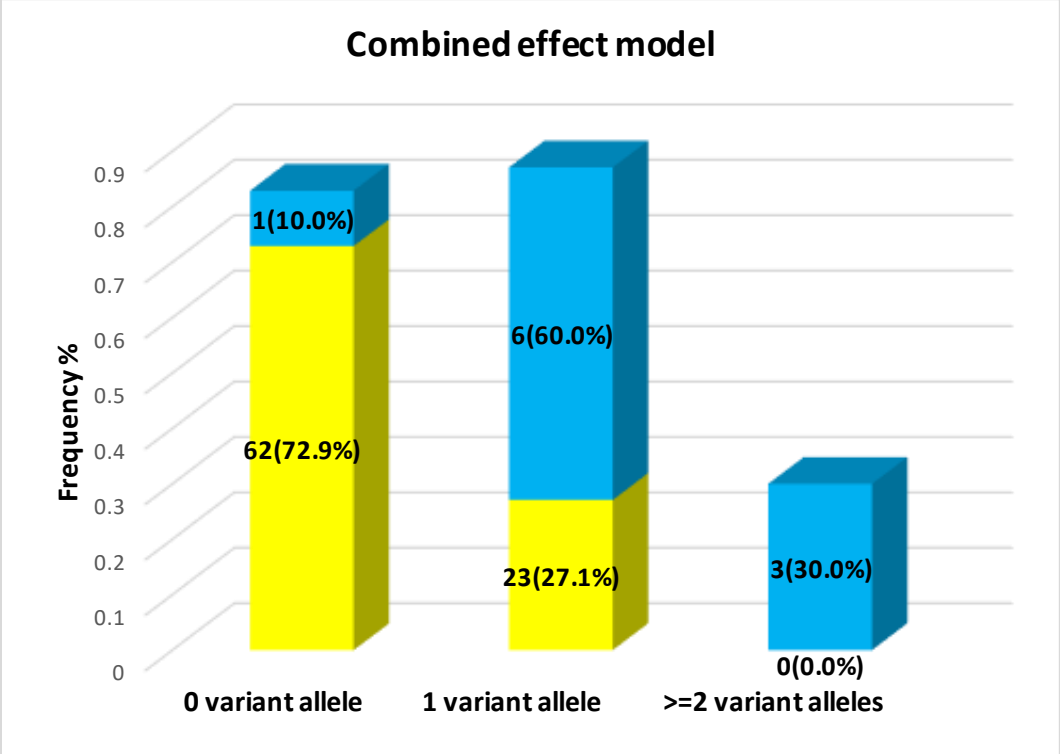
PTPRZ1: Protein Tyrosine Phosphatase Receptor Type Z1; MUC16: Mucin 16, Cell Surface Associated; TNRC6C-AS1: TNRC6C antisense RNA 1.

4.10.11. Supplemental Figure 3. Combined effect model of the PTPRZ1 rs740965, MUC16 rs750295511 and TNRC6C-AS1 rs2748431 variants.



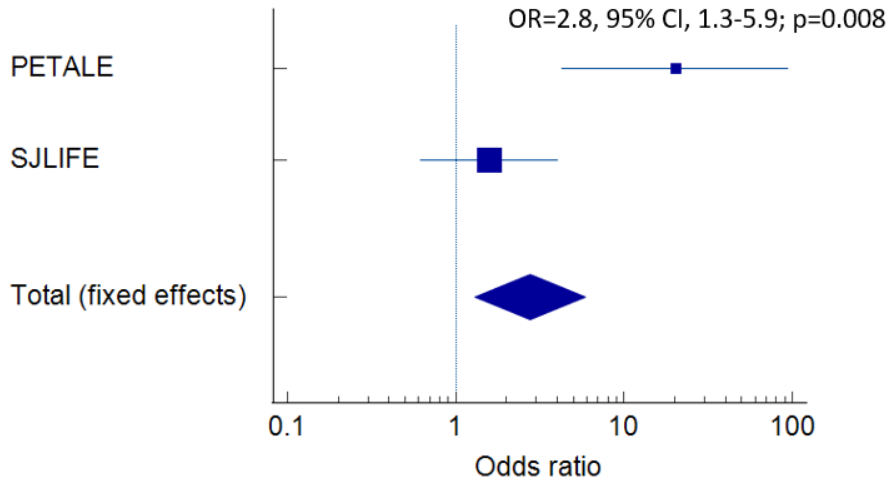
Association of moderate-severe anxiety with genetic variations in the *PTPRZ1, MUC16* and *TNRC6C-AS1* genes detected in the whole cohort of participants is presented as combined effect model. The combined effect was assessed through the multivariate regression model and was adjusted for non-genetic co-variables (age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment – patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy), p=2.00E-05, OR=5.1, 95% CI, 2.4-10.9. The frequency of patients with and without moderate-severe anxiety is represented by blue and yellow bars, respectively. The total number and percentage of samples per category is displayed inside the bars.

4.10.12. Supplemental Figure 4. Combined effect model of the APOL2 rs7285167 and C6orf165 rs61731441variants.



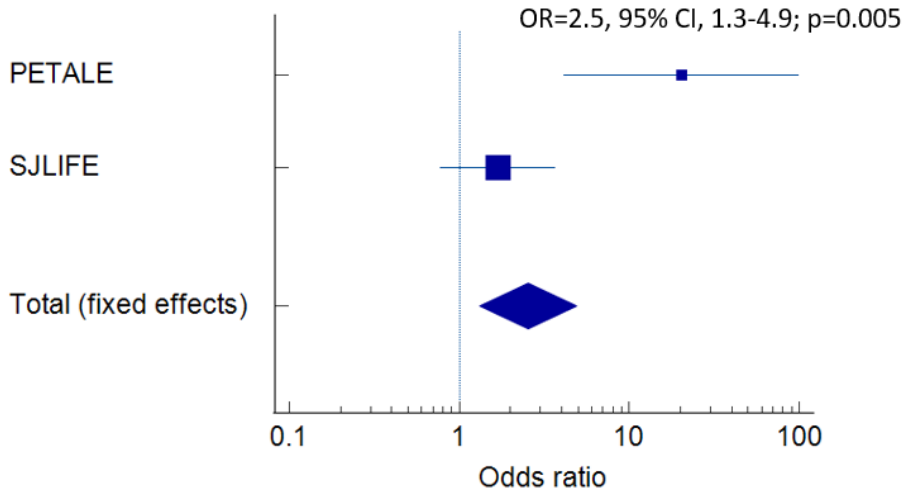
Association of moderate-severe anxiety with genetic variations in the *APOL2* and *C6orf165* genes detected in male participants is presented as combined effect model. The combined effect was assessed through the multivariate regression model and was adjusted for non-genetic co-variables (age at diagnosis, time between end of treatment and evaluation, DFCI protocol, treatment – patients who received chemotherapy only vs. patients who received chemotherapy and radiotherapy), p=0.001, OR=32.5, 95% CI, 3.8-275.8. The frequency of patients with and without moderate-severe anxiety is represented by blue and yellow bars, respectively. The total number and percentage of samples per category is displayed inside the bars.

4.10.13. Supplemental Figure 5. Meta-analysis of the effect of the variant rs61732180 ZNF382 on neurocognitive deficit in combined discovery and replication set that included male participants.



Plot represents the association of the variant rs61732180 *ZNF382* with the deficit in the trail making test performance as tested in the discovery cohort (PETALE), the replication cohort (SJLIFE) and the cohort combining them both (Total). Odd-Ratios (OR) comparing carriers to non-carriers, along with the 95% confidence intervals (95% CI) and the p-values of the associations are provided at the top of the graph. The Meta-analysis was performed using the Mantel-Haenszel method implemented in *MedCalc* software and assuming a fixed-effects model.

4.10.14. Supplemental Figure 6. Meta-analysis of the effect of the variant rs35672330 EXO5 on moderate-severe depression in combined discovery and replication set that included male participants.



Plot represents the association of the variant rs35672330 *EXO5* with the moderate-severe depression as tested in the discovery cohort (PETALE), the replication cohort (SJLIFE) and the cohort combining them both (Total). Odd-Ratios (OR) comparing carriers to non-carriers, along with the 95% confidence intervals (95% CI) and the p-values of the associations are provided at the top of the graph. The Meta-analysis was performed using the Mantel-Haenszel method implemented in *MedCalc* software and assuming a fixed-effects model.

4.10.15. Item S1. The description of the gene identified through an exome-wide association study between genetic factors and long-term treatment related neurocognitive deficits, and anxiety and depression in survivors of childhood acute lymphoblastic leukemia.

The *AK8* (*Adenylate Kinase 8*) gene encodes for a nucleoside monophosphate kinase, one of the essential regulators for maintaining the proper adenine nucleotide composition for many different cell functions, including nervous system development[97]. It is strongly expressed in the brain, and deletions of the *AK8* gene have been found in medulloblastoma[98]. Furthermore, it was identified as a gene associated with attention deficit / hyperactivity disorder (ADHD) in adults[99], and its potential implication in ADHD was confirmed in the more recent report that integrated candidate gene and genome-wide association studies using bioinformatics and complex network analysis[100]. In addition, adenylate kinases are modifiers of conditions, in which oxidative stress plays a critical role, such as metabolic syndrome or neurodegenerative diseases[101].

PTPRZ1 (Protein Tyrosine Phosphatase Receptor Type Z1) encodes protein tyrosine phosphatase receptor, which is expressed primarily in the nervous system including oligodendrocytes, astrocytes, and neurons[102]. Diverse evidence suggests that tyrosine phosphorylation is an essential component in myelin formation, differentiation of oligodendrocytes and Schwann cells, and recuperation from demyelinating lesions[102]. Demyelinating lesions are associated with multiple sclerosis, leukodystrophies and demyelinating diseases of the peripheral nerves[103]. Studies in a mouse model have shown that *Ptprz* is involved in the regulation of the dopaminergic system and plays an important role in behavioral responses[104]. Furthermore, emerging evidence suggests that changes in myelin levels have been associated with various mental illnesses[89]. Indeed, studies in a rodent model have demonstrated that juvenile exposure to acute traumatic stress leads to long-term changes in grey matter myelination in adult animals and was gender-specific[89]. Additionally, in a recent study, abnormalities in myelin levels in the

brain were detected by quantitative magnetic resonance imaging in individuals diagnosed with major depressive disorder[105].

The *MUC16* (*Mucin 16, Cell Surface Associated*) gene encodes one of the largest glycoproteins known, type I transmembrane mucin - member of the mucin family of proteins[106]. Besides their normal physiological role in protecting epithelial cells, mucins have been shown to be involved in various diseases, such as psoriasis, pulmonary tuberculosis, ulcerative colitis and cancer[106]. *MUC16* is overexpressed in several cancers[106], including breast[107], pancreatic[108] and lung cancer[109], as well as in digestive tract adenocarcinomas[110]. Furthermore, *MUC16* was also identified among the genes with consistently altered expression in post-mortem brains of individuals with bipolar disorders[111]. Nonetheless, more studies are needed to clarify its potential involvement in mood disorders.

TNRC6C-AS1 is an antisense RNA gene that belongs to the long non-coding RNA (lncRNA) class of cellular transcripts which have demonstrated multifaceted involvement in various fundamental biological processes[112]. While only a few lncRNAs have currently been characterized molecularly or functionally, their deregulation is often linked to cancer and to many developmental, cardiovascular, and neurological disorders[113]. *TNRC6C-AS1* is located on chromosome 17q25.3, and shares an overlapping region with the 3′UTR region of the *TNRC6C* gene (its cognate sense gene)[114]. *TNRC6C-AS1* regulates *TNRC6C* expression at both the mRNA and protein levels [114]. Interestingly, *TNRC6C-AS1* was found to be up-regulated in type 2 diabetes patients with depressive symptoms[115]. The contribution of lncRNAs has been confirmed in brain development, neuronal function, maintenance and differentiation, as well as in synaptic plasticity[116]. Also, growing evidence suggests the potential involvement of lncRNAs in the pathogenesis of neurodegenerative diseases[116], and psychiatric disorders[113]. Intriguingly, *TNRC6C* was reported to be significant in a recent large-scale genetic association analysis of educational

attainment, and was linked to higher mathematical ability[117]. However, understanding the role of *TNRC6C-AS1* in emotional distress will require further investigation.

Furthermore, our analysis also suggests that synergistic interactions might exist between the variants identified in the entire discovery cohort in relation to moderate-severe anxiety (rs740965 *PTPRZ1*, rs2748431 *MUC16*, and rs2748431 *TNRC6C-AS1*), which could explain the markedly significant associations in the combined variants model.

APOL2 (Apolipoprotein L-II) gene belongs to the APOL gene cluster that encodes high density lipoproteins, which has a key role in cholesterol transport[118]. Almost all cholesterol present in the CNS is synthesized *de novo* in brain cells, and is efficiently recirculated within the CNS. Accordingly, it exists in a compartment distinct from the rest of the body, and plays an important role in cellular processes such as modulation of gene transcription and signaling, both during the development of the nervous system and in the adult brain[119]. Although the biological function of *APOL2* in the brain remains unclear[118,119], its expression is detected in the brain[119], and the level of expression is significantly upregulated in schizophrenia[120]. Polymorphisms in this gene have been associated with schizophrenia risk[121].

C6orf165, also known as *CFAP206 (Cilia And Flagella Associated Protein 206),* is a protein-coding gene; however, the functions of the encoded protein (DUF3508) are currently poorly understood. In a recent study, which suggested that resting myeloid cells are a key source of immune dysfunction in Huntington's disease, *C6orf165* was identified as a gene with significant dysregulation of expression[122]. In addition, the *C6orf165* gene was characterized as a gene with ciliary functions[123]. However, *C6orf165* has no currently known associations with the CNS function and /or emotional distress in humans. Interestingly, through the combined effect model, we also noted a possible interaction between the *APOL2* and *C6orf165* genes.

PCMTD1(Protein-L-Isoaspartate (D-Aspartate) O-Methyltransferase Domain Containing 1) encodes a member of the methyltransferase superfamily[124]. *PCMTD1* locus may have a significant role in the development or progression of Primary Angle-Closure Glaucoma[125] and endometrioid ovarian cancer[124]. Other functions of *PCMTD1*, including those associated with emotional or behavioral problems, are not currently known in the literature.

CYP2W1 (cytochrome P450 family 2 subfamily W member 1) encodes a cytochrome P450 monooxygenase[126]. CYP2W1 was linked to metabolism of a variety of endogenous substrates, including lysophospholipids and certain pro-carcinogens such as polycyclic aromatic hydrocarbons[127]. The *CYP2W1* rs3735684 (Ala181Thr) identified in our study as the most significant contributor to the association signal through the rare variants' analysis, is also associated with increased risk of colorectal cancer[126]. It is interesting to note that in a mice model, psychosocial stress, along with the gut microbiota, were shown to have an important impact on the expression and activity of cytochrome P450 enzymes (CYP)[128]. Furthermore, the stress effectors, the adrenergic receptor-linked pathways, as well as glucocorticoids, play essential and distinct roles in stress-mediated regulation of CYPs in a species- and tissue-specific manner[129]. Nonetheless, the data linking the *CYP2W1* gene and emotional distress is not currently available.

Section B

Chapter 5

Genetic factors in treatment-related cardiovascular complications in survivors of childhood acute lymphoblastic leukemia.

Pharmacogenomics. 2021 Sep;22(14):885-901. doi: 10.2217/pgs-2021-0067. Epub 2021 Sep 10.

PMID: 34505544 DOI: 10.2217/pgs-2021-0067

Genetic factors in treatment-related cardiovascular complications in survivors of childhood acute lymphoblastic leukemia.

Running title: Genetics of cardiovascular complications in ALL.

Kateryna Petrykey[1,2], Aziz Rezgui[1], Mathilde Le Guern[1], Patrick Beaulieu[1], Pascal St-Onge[1], Simon Drouin[1], Laurence Bertout[1], Fan Wang[4], Jessica L. Baedke[4], Yutaka Yasui[4], Melissa M. Hudson[4,5], Marie-Josée Raboisson[1,3], Caroline Laverdière[1,3], Daniel Sinnett[1,3], Gregor U. Andelfinger[1,3] and Maja Krajinovic[1,2,3*].

1. Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada.
2. Department of Pharmacology and Physiology, Université de Montréal (Quebec), Canada.
3. Department of Pediatrics, Université de Montréal (Quebec), Canada.
4. Epidemiology and Cancer Control Department, St. Jude Children's Research Hospital, Memphis (Tennessee), USA.
5. Oncology Department, St. Jude Children's Research Hospital, Memphis (Tennessee), USA.

Corresponding author*:
Maja KRAJINOVIC
Sainte-Justine University Health Center (SJUHC)
3175 chemin de la Côte-Sainte-Catherine
Montreal, Quebec, Canada, H3T 1C5
Fax: 514-345-4731
Phone: 514-345-4931 #6259
E-mail: maja.krajinovic@umontreal.ca

Conflict of interest statement

The authors declare no potential conflicts of interest.

5.1. ABSTRACT

Aim: Cardiovascular disease represents one of the main causes of secondary morbidity and mortality in patients with childhood cancer.

Patients & methods: To further address this issue, we analyzed cardiovascular complications in relation to common and rare genetic variants derived through whole-exome sequencing from childhood acute lymphoblastic leukemia survivors (PETALE cohort).

Results: Significant associations were detected among common variants in the *TTN* gene, left ventricular ejection fraction ($p \leq 0.0005$), and fractional shortening ($p \leq 0.001$). Rare variants enrichment in the *NOS1, ABCG2* and *NOD2* was observed in relation to left ventricular ejection fraction, and in *NOD2* and *ZNF267* genes in relation to fractional shortening. Following stratification according to risk groups, the modulatory effect of rare variants was additionally found in the *CBR1, ABCC5* and *AKR1C3* genes. None of the associations was replicated in St-Jude Lifetime Cohort Study.

Conclusion: Further studies are needed to confirm whether the described genetic markers may be useful in identifying patients at increased risk of these complications.

Keywords

anthracycline-induced cardiotoxicity, doxorubicin, whole-exome sequencing, genetic association studies, pharmacogenomic markers, childhood acute lymphoblastic leukemia, late adverse effects, cancer survivors

5.2. INTRODUCTION

The introduction of multi-agent treatment protocols and the optimization of dosing regimens have led to a remarkable improvement in cure rates of childhood Acute Lymphoblastic Leukemia (ALL), the most frequent pediatric cancer. However, exposure to cytotoxic therapy at an early age has been associated with treatment-related late-adverse effects (LAEs) – often not clinically apparent until decades after treatment. Among LAEs, cardiovascular disease represents one of the main causes of secondary morbidity and mortality in patients with childhood cancer[1-3].

While the events leading to cardiovascular disease are diverse, one of the fundamental problems is the irreversible loss of cardiomyocytes, which in turn contributes to alterations in ventricular contractility and relaxation[4] as well as alteration of the function of endothelial cells affected by inflammation and by the production of reactive oxygen species[5].

Different molecular mechanisms were suggested behind the genesis and progression of cardiovascular disease[6-8]. For example, exosomes have also been described as playing an important role in various processes involved in the pathogenesis of cardiovascular disease such as cardiac fibrosis, hypertrophy, and myocardial apoptosis[9,10].

The interaction of at least three main factors determines the cardiovascular status of cancer patients: (a) the patient's underlying cardiovascular health status (including pre-existing heart disease and cardiovascular risk factors), (b) cancer itself that could directly or indirectly affect the heart and vasculature, and (c) cancer treatment (including conventional chemotherapy, targeted agents, and radiotherapy) which can damage several pathways that control functions of the cardiovascular system[11,12].

Anthracyclines, especially doxorubicin, are commonly used for a variety of hematological and solid malignancies including leukemia, lymphoma, Wilms tumor, neuroblastoma, hepatoblastoma, and high-risk embryonal malignancies[13-15]. However, anthracycline-related

cardiotoxicity including subclinical heart failure, reduction in left ventricular mass and wall thickness, or decline in systolic and diastolic function, can develop during, shortly after, or many years after exposure adversely affecting the long-term prognosis in pediatric cancer survivors[1,2,16-18].

Furthermore, despite an established dose-dependent association, there is a substantial variability regarding the occurrence and severity of anthracycline-induced cardiotoxicity in survivors of childhood cancer[19], which may be related to genetic factors.

In this study, we examined whether common and rare genetic polymorphisms contribute to this variability by altering the risk of treatment-related cardiotoxicity in combination with other factors. We performed candidate gene analyses in a well-described cohort of ALL survivors (PETALE)[20,21] in relation to treatment-related cardiac complications which identified several novel markers associated with these complications in the *TTN*, *NOS1, ABCG2, CBR1, ABCC5,* and *AKR1C3* genes. In addition, the association analyses were supplemented by a hypothesis-free approach - an exome-wide association study, which detected the contribution of the rare variants in the *NOD2* and *ZNF267* genes as potential modulators of the risk of developing treatment-related cardiotoxicity in survivors of childhood ALL.

5.3. STUDY POPULATION AND METHODS

5.3.1. Discovery cohort

A total of 233 patients treated for childhood ALL according to Dana Farber Cancer Institute (DFCI) ALL 87-01 to 05-01 protocols at Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada, were included in the analysis. The discovery set was composed of participants of previously described [20,21] PETALE study, a multidisciplinary research project aiming to identify and comprehensively characterize associated predictive biomarkers of long-term treatment-related complications in childhood ALL survivors. The participants were recruited during 2013-2015, and their characteristics are shown in **Table 1.** Briefly, eligible participants were younger than 19 years old at diagnosis, survived at least 5 years after diagnosis of ALL, without a history of relapse or refractory ALL, Down syndrome, or hematopoietic stem cell transplantation. Participants were all of European descent[21]. The median age of patients at the time of diagnosis was 4 years, the time from the end of treatment to evaluation ranged from 3-24 year with a median of 13 years (for 74.7% of participants, it was ≥10 years), both sexes were equally represented (51.1% of females).

The patients were classified into standard (SR) and high risk (HR) groups based on prognostic factors, including age at diagnosis, white blood cell count, immunophenotype, and central nervous system (CNS) involvement at diagnosis of ALL (2, 3). The frequency of patients assigned to SR and HR groups during the treatment was 45.5% and 54.5%, respectively. Anthracycline-exposed participants had a normal cardiac function before anthracycline exposure. During the induction phase of childhood ALL treatment, all patients were given two 30 mg/m^2 doses of doxorubicin. HR patients continued to receive one weekly doxorubicin dose during the consolidation phase, for a total cumulative dose of 294.52 mg/m^2 (median value, range 137.78 - 472.85 mg/m^2). HR patients treated with 95-01 protocol were randomized to receive (or not) dexrazoxane, a

cardioprotective agent, during induction and consolidation therapy, while all HR patients treated with 00-01 and 00-05 protocols, received dexrazoxane before doxorubicin administration (cumulative doses ranged from 1433.84 mg/m2 to 3706.17 mg/m2, with the median value of 2961.85 mg/m2).

5.3.2. Echography measurement of cardiac function and structures

Measurements and calculations were made according to the previously published recommendations[22] and were obtained at the last follow-up examination[20,23]. Participants were questioned and examined for symptomatic heart failure according to New York Heart Association functional criteria[24]. Height and body weight were measured at evaluation. Current use of medication modulating cardiovascular state was noted[20]. Echocardiographic parameters for the quantitation of left ventricular (LV) morphology and function included the LV end-diastolic diameter (EDD), fractional shortening (FS), and ejection fraction (EF). Two measurements of the LVEF were studied – LVEF M-Mode and LVEF 2D (Simpson). LVEDD values were transformed into z scores adjusted for height, weight, gender, and age. LVEDD z score outcome was calculated by splitting the patients into those ≤ 18 years old at the time of evaluation and those >18 years old. The z scores were calculated using the online calculator corresponding to the above-mentioned age groups[25,26] (http://www.parameterz.com/sites/m-mode).

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of SJUHC.

5.3.3. Sequencing and quality control

Whole-exome sequencing (WES) was performed on germline DNA, extracted from peripheral blood samples from participants of the PETALE cohort, using standard protocols as described previously[20]. Whole exomes were captured in solution with Agilent's SureSelect Human All Exon 50Mb kits and sequenced on either Life Technologies SOLiD System 4.0 (mean coverage = 40X) or Illumina HiSeq 2500 platform (mean coverage = 113.1X) at SJUHC integrated clinical genomic center in pediatrics. Reads were aligned to the hg19 reference genome using SOLiD LifeScope software[27] for the SOLiD samples and BWA-MEM[28] for the samples sequenced on the Illumina system. PICARD[29,30] was used to mark PCR duplicates and collect sequencing quality control metrics. Variant calling was performed using the Haplotype Caller and quality score recalibration was performed using Variant Recalibrator, both implemented in the Genome Analysis Tool Kit (GATK)[30]. Variants were selected based on the variant quality score (VQSR = PASS) and the minimum depth of coverage (DP≥10). The final germline variants were annotated by ANNOVAR[31]. Only missense, nonsense, and splicing common and rare variants with predicted functional impact (Sift (<0.1) and PolyPhen2 (≥0.85)) were considered[32,33]. Variants were defined as rare (minor allele frequency, MAF<5%) and common (MAF≥5%) according to the reported frequency for European populations in the 1000 Genomes[34] and ESP6500 datasets[35]. Variants exceeding a missing rate of 20%, not in Hardy-Weinberg Equilibrium (P<0.001), and common variants with pairwise linkage disequilibrium (LD, r2≥0.8) were excluded.

5.3.4. Association analyses

To conduct genetic association studies, we used a candidate gene approach with three gene sets (listed in **Supplemental Table 1)**. These sets included genetic variants in 51 genes selected for doxorubicin action pathways[36-40], as well as the 107 genes involved in the functioning of the cardiac system, and 107 genes relevant for the mitochondrial function selected due to the role of

oxidative stress in cardiovascular disease[41,42]. All candidate genes were selected using the KEGG PATHWAY Database[43]. A total of 80 common variants present in 49 genes (51 variants in cardiovascular system function, 16 variants in doxorubicin pathway, and 13 variants in mitochondria function) that satisfied all above-mentioned filtering criteria were used in association analyses. For an exome-wide association study, there were 5373 common genetic variants corresponding to 3793 genes that were retained for this analysis according to the described criteria.

The analyses between common genetic variants and each of the four echocardiographic parameters of interest were performed using Quantitative trait association based on Wald statistic and linear regression implemented in PLINK v.1.07 software[44,45], and SPSS v.25.0.0.0. Analyses were performed in 236 sequenced ALL survivors and stratified by risk groups with different treatment intensity, and treatment with or without dexrazoxane cardioprotection in HR patients because these factors have an established role in modulating studied outcomes[19,46,47]. The Benjamini-Hochberg procedure for false discovery rate (FDR)[48,49] was used to adjust for multiple testing with a cut-off value of < 5% considered statistically significant (top-ranking variants); additionally, $p<0.0006$ is the Bonferroni p-value corrected threshold for the number of variants tested. Multiple regression models included the following adjustment variables: age at the time of diagnosis (continuous variable); time since the end of treatment (continuous variable); sex: males/females (categorical variable); DFCI Protocol: 87-01=1, 91-01=2, 95-01=3, 00-01=4, 05-01=5 (categorical variable); and a risk-treatment combined variable with the following scores: SR patients =1, HR patients that received a cardioprotective drug (dexrazoxane)=2, HR patients that did not receive dexrazoxane=3 (categorical variable).

A total of 1124 rare variants present in 102 genes (845 variants in 55 genes related to cardiovascular system function, 216 variants in 33 genes in doxorubicin pathway, and 63 variants in 14 genes involved in mitochondria function) that satisfied all above-mentioned filtering criteria for variants selection were used in the first part of the association analyses using a candidate gene approach. Next, 59044 rare genetic variants corresponding to 11441 genes that satisfied the filtering criteria entered the exome-wide association analysis.

The effects of rare variants in both the candidate genes and exome-wide approaches were evaluated using the SKAT-O test (Optimal Sequence Kernel Association Test)[50,51] implemented in SKAT package v.1.3.2.1[52]. Only genes with at least two variants that satisfied filtering criteria were retained for the SKAT-O test. Multiple test adjustments (with a cut-off value of FDR < 5% considered statistically significant) were performed, including adjustment for every candidate genes' data set separately. To evaluate individual variant contributions to association signals, a collapsing approach[53,54], with the iterative exclusion of every single variant, was additionally executed. Similar to the common variant analyses, genetic associations were assessed for each of the four outcomes in the entire cohort, as well as in SR and HR groups, and through multivariable models adjusted for the same covariates.

5.3.5. Replication cohort

The replication cohort consisted of childhood ALL survivors of European ancestry enrolled in the St. Jude Lifetime Cohort (SJLIFE) study with whole-genome sequencing. SLIFE participants included in the replication analysis were limited to those with at least one of the echocardiographic measures identified to have a statistically significant genetic association in the PETALE cohort. The maximum number of participants in the replication cohort with available outcome data was 149; the total number analyzed varied depending on the outcomes and subgroup studied.

Therefore, the effect between the significant variants identified in the discovery cohort and LVEF M-mode and FS outcomes was tested in 90 patients in total (regardless of their risk assignment); while 56 and 102 SR patients were tested in relation to the LVEF M-mode and LVEF 2D changes, respectively. Participants were selected to resemble the discovery cohort based on demographics and treatment characteristics (risk groups). They were younger than 19 years at diagnosis, with no history of relapse within 5 years of the primary ALL diagnosis date, Down syndrome, or hematopoietic stem cell transplantation. The number of participants tested for each outcome including risk groups, along with the age at diagnosis, years since the end of treatment, and sex distribution are provided in **Supplemental Table 5**); on average the age at diagnosis was 4.7 years, time since the end of treatment was 27.4 years and 48.5% of participants were females. Participants met the definitions for the 1980s SR and HR ALL populations as defined previously[55], (risk group definitions are provided in **Supplemental Table 6)**. The latter criteria for the two risk groups were implemented since not all variables used for risk classification in the discovery cohort were available in the replication group. Unlike the discovery cohort, none of the HR survivors were exposed to dexrazoxane; this variable was not used in replication analysis. Analyses of top-ranking common variants were performed by linear regression adjusting for continuous age at diagnosis, sex, continuous time since the end of treatment, and the SR and HR groups. Analyses of top-ranking rare variants were performed by SKAT-O (when applicable) and by linear regression, adjusting for the same covariates. Upon quality control, if variants present within a gene of interest became monomorphic after excluding samples without phenotype data, they were not analyzed.

5.4. RESULTS

Demographics and clinical characteristics of PETALE participants are presented in **Table 1.** The analyses were performed in either all patients or subgroups that differ relative to doxorubicin treatment. These included patients assigned to standard (SR, 45.5%) and high risk (HR, 54.5%) groups who received different cumulative doxorubicin doses[20] and HR patients who received or not cardioprotectant (53.5% and 46.5%, among HR patients, respectively).

Among common variants implicated in cardiac system function, significant associations were detected between LVEF and FS M-mode with single nucleotide variations (SNVs) in the titin *(TTN)* gene when the entire cohort of survivors was analyzed. The higher mean values for the minor allele carriers, and thus better LVEF, were observed for all three *TTN* rs72648998, rs3829747, and rs2303838 polymorphisms (**Table 2a,** p=0.0003, 0.0003 and 0.0005, respectively). The association is also depicted by the Manhattan plot (**Figure 1**) and p-value distributions are represented in the histograms in **Supplemental Figure 1**. The same three *TTN* polymorphisms were also associated with the higher FS values (**Table 2b**, p=0.0002, 0.0004 and 0.001, respectively). All *TTN* variants (except rs2303838 in relation to FS) sustained Bonferroni correction for multiple testing. The *TTN* genotypes also remained significant in a multiple linear regression model when controlling for non-genetic covariates (**Table 3a and 3b**).

The analysis of functionally predicted rare variants related to the doxorubicin pathway **(Table 4** and **Supplemental Table 2**) led to the detection of an association between LVEF M-Mode and rare variants enrichment in *NOS1* and *ABCG2* genes (p=0.0013 and p=0.0026, respectively) in the full PETALE cohort. Following stratification according to the risk groups, the additional association of the same outcome was detected with candidate genes *CBR1* in the SR group and *ABCC5* in the HR group, (p=0.0015 and p=0.0008, respectively), but *NOS1* and *ABCG2* associations were not detected in either stratum. The same association of *CBR1* and *ABCC5* genes

in a risk dependent manner was noted with FS. Additionally, in SR patients, the association between FS and LVEF 2D outcomes and rare variants in the *ABCG2* gene was observed (p=0.0027 and p=0.0019, respectively); as well as between LVEF 2D and rare variants enrichment in *AKR1C3* gene (p=0.0067).

Using the collapsing approach, we further explored the most significant contribution of single variants or their combination (**Table 4**) to the association signal obtained by SKAT-O. The analyses identified rs76090928, rs199473672, and rs2835266 SNVs in *NOS1, ABCG2* and *CBR1* respectively; and significant combinations of two SNVs in *ABCC5* and *AKR1C3* genes. A risk effect was seen for *CBR1* rs2835266 polymorphism, while the presence of all other collapsed rare variants for *NOS1, ABCG2,* and *AKR1C3* was associated with a protective effect. The top-ranking individual rare loci or identified combinations remained significant in their respective multivariable models (data not shown). An illustrative example with the effect of the most contributing variants in the *NOS1* and *ABCG2* genes on LVEF in M-mode, in the presence of non-genetic covariates is given in **Supplemental Table 3** and **Supplemental Figure 2.**

We did not identify any significant common variant associations that would satisfy multiple testing adjustments with an FDR threshold of <5% through exome-wide association. In contrast, analysis of functionally predicted rare variants on the exome-wide level in the full PETALE cohort (**Table 5** and **Supplemental Table 4**) led to the detection of an association between FS and rare variants enrichment in the *NOD2* and *ZNF267* genes ($p=1.39x10^{-6}$ and $p=3.51x10^{-6}$, respectively). In addition, rare variant enrichment in the *NOD2* gene was also detected in relation to LVEF M-Mode ($p=2x10^{-6}$). Variant combinations that contributed most to the observed association signals are presented in **Table 5** using the collapsing approach**.**

5.4.1. Replication analysis

The number of statistically significant discovery variants proposed for SJLIFE replication can be found in **Supplemental Table 7**. All 3 *TTN* common variants were available in the SJLIFE whole-genome sequencing data. Of the 46 rare variants proposed for replication, only 26 passed SJLIFE whole-genome sequencing quality control, consistent with the variant information availability with gnomAD[56], of which only 8 were not monomorphic in SJLIFE. The results of the replication analyses are presented in **Supplemental Tables 8** and **9** for common variants in the *TTN* gene and non-monomorphic quality-controlled rare variants across several genes (most important contributors). None of the associated genes were replicated in the SJLIFE cohort. It is interesting to note nevertheless that the same direction of the effect was observed for the collapsed *NOS1* and *CBR1* rare variant signal, notably the carriers of the *NOS1* rare variants in the full population had LVEF M-mode 71% vs 65% for non-carriers, and 62% vs 67% for *CBR1* in the SR group (**Supplemental Table 9**).

5.5. DISCUSSION

Elevated risk of clinical heart failure has been reported in survivors of childhood cancer exposed to anthracycline treatment[47,57,58]. The increased risk of cardiotoxicity underlines the importance of understanding the mechanisms behind this complication and its predictors.

Several candidates gene and genome-wide association studies (GWAS) have reported genetic variants that are associated with long-term treatment-related cardiotoxicity[23,29,59-71]. Our analyses, using candidate genes and exome-wide approaches, identified 8 loci that can influence the treatment-related cardiotoxicity, notably common and rare variants in *TTN, NOS1, ABCG2, ABCC5, CBR1*, *AKR1C3, NOD2,* and *ZNF267* genes. It is worth mentioning that all of these genes, with the exception of *ZNF267*, have been previously associated with treatment-related cardiotoxicity[23,61,72-74] and/or studied in the context of cardiovascular morbidity[75,76].

5.5.1. Variants detected through candidate gene analyses.

5.5.1.1. Functioning of the cardiac system (TTN)

We have shown a cardioprotective effect of the three common independent (without a strong linkage disequilibrium, r2<0.8) *TTN* variants for both LVEF and FS M-mode outcomes. Titin, encoded by the *TTN* gene, is the largest known human protein (molecular mass of up to ~3800 kDa) and serves as the main scaffold of the sarcomere and participates in structural, sensory, and signaling functions of the heart[77,78]. Titin-truncating variants represent the most common genetic predisposition to the dilated cardiomyopathy[78-80]. Moreover, polymorphisms of the *TTN* gene have been associated with other clinical phenotypes, including peripartum cardiomyopathy[81,82] and skeletal myopathies (some of which are also characterized by cardiac impairment)[83-86]. Moreover, the *TTN* gene has been mentioned in the context of cardiac function during endurance training[87] and previously detected in the PETALE participants in association with cardiorespiratory fitness[88]. Furthermore, two recent studies reported the contribution of rare

variants in the *TTN* gene that lead to truncated isoforms of the protein, to the increased risk for chemotherapy-induced cardiomyopathy in children and adults[59,89]. Our results in contrast identified common *TTN* variant alleles that are associated with lower susceptibility to anthracycline-related cardiotoxicity. Interestingly, one of the variants described in our study, rs72648998, was found to be associated with prolonged QT interval in the GWAS of 31 cohorts of European origin[90]. Given demonstrated importance of the *TTN* gene for the impairment of cardiac function, we evaluated all nominally significant variants from WES data in this gene even if they did not pass the FDR threshold (details on all tested common *TTN* variants are presented in **Supplemental Table 10a.** Interestingly, we detected the rs35813871 variant with a risk-increasing effect on LVEF (M-mode) and FS (**Supplemental Table 11**). The rs35813871 *TTN* is tagging one of the *TTN* haplotypes (**Supplemental Table 10b**), which includes a non-synonymous SNV (missense variant) predicted to affect the protein function (NP_001254479.2: p.Thr811Ile), however, the evidence on its role in the context of cardiovascular disease is limited. It is important to note that due to the high frequency of *TTN* variants in the general population and their incomplete penetrance, interpretation of *TTN* variants may demand further investigation, including a comprehensive assessment of the clinical phenotypes, as well as functional studies.

5.5.1.2. Semiquinone Formation: One Electron Reduction (NOS1)

An association between LVEF M-Mode and rare variants enrichment in the *NOS1* gene coding for Neuronal nitric oxide synthase (nNOS) was detected through candidate gene association studies. The one-electron reduction of anthracyclines to generate a semiquinone radical in the cytosol is carried out by nitric oxide synthases (nNOS), Inducible nitric oxide synthase (iNOS), and Endothelial NOS (eNOS))[91]. Doxorubicin is involved in the inhibition of NOS activity as it binds to all three isoforms[92]. A cardioprotective effect of the common variant in *NOS3* gene (encoding the eNOS enzyme) has been previously reported in survivors of ALL[23].

Decreased bioavailability of nitric oxide (NO) in the endothelium, a key precursor for impaired vasodilation and hypertension, has been linked to cardiac hypertrophy, fibrosis, and myocardial infarction[93]. In the healthy heart, nNOS are localized in the cytosol, however, pathological stimuli may change its properties and location (nNOS translocate from sarcoplasmic reticulum to cytoplasmic membrane)[94], affecting downstream signaling[93].

In hypoxic conditions, the increase in mRNA and protein expressions as well as the activity of nNOS in the systematic and local arteries has been observed in an animal model and in a human aortic cell line[95]. Interestingly, rs76090928 (identified in the current study as the strongest contributor to the association signal in the set of rare variants), is located in exon 1; this exon contains a hypoxia-responsive promotor and is exclusively transcribed in hypoxic conditions[96], thus functionally affecting the vascular smooth muscle contraction[95].

5.5.1.3. Membrane transporters (ABCG2, ABCC5)

We report a protective effect of functionally predicted rare variants of the *ABCG2* (ATP-binding cassette subfamily G member 2) gene; ABCG2 (also known as breast cancer resistance protein BCRP) belongs to the G-subsection of the superfamily of ABC transporters[97]. Its expression has been detected in a variety of tissues including endothelial cells of the human heart[97]; besides, its variable cardiac expression could be induced by dilated and ischemic cardiomyopathies[97]. High expression of *ABCG2* has been observed in adult ALL patients[98]. The modulating effect of *ABCG2* on chemotherapy-related cardiotoxicity remains controversial. The variant allele of common rs2231142 polymorphism in *ABCG2* gene was associated with greater cardiac toxicity in adults treated for acute myeloid leukemia (AML)[99], whereas similar studies conducted in AML, childhood ALL and osteosarcoma did not demonstrate such an effect[65,73].

Following stratification according to risk groups, the analysis of rare variants led to the detection of a protective effect between LVEF and FS, and rare variants enrichment in the *ABCC5* gene in survivors treated according to the high-risk protocols. Several variants were previously identified as predictors of anthracycline-induced cardiotoxicity in the *ABCC* subfamily[65,67,68,100], in particular, the *ABCC5* rs7627754 TT genotype, was linked to significant reductions in EF and FS in survivors of childhood ALL[23].

5.5.1.4. Hydroxylation: Two Electron Reduction (CBR1, AKRC1)

We observed the higher risk of toxicity for the rare *CBR1* rs2835266 polymorphism in SR patients, which was the only risk-increasing effect identified in our study. This finding corresponds to the previously reported modifying effect of *CBR* common genotypes (CBR3 V244M) that was restricted only to low- to moderate-dose anthracycline regimens[61].

The *CBR1* gene is located adjacent to the related *CBR3* gene which encodes for a homolog with supposedly similar catalytic properties[101]. The polymorphisms in *CBR* genes have been investigated in several research studies and have been shown to affect the catalytic activity for anthracyclines[61,102]. Besides, the importance of the *CBR1* gene in doxorubicin reduction has been demonstrated in animal models, in which the inactivation of one Cbr1 allele protected murine cardiac tissue from the damage caused by doxorubicin[103]. Interindividual differences in the expression and activity of *CBR1* could be responsible for the variability of the reduction of doxorubicin and the formation of its metabolites in cancer patients. Moreover, the generation of anthracycline alcohol metabolites does not seem to be dependent on anthracycline dose. In the cases of high anthracycline doses, cardiotoxicity is mainly mediated by oxidative stress caused by an excess of unmetabolized anthracyclines[61] and is thus independent from anthracycline alcohol metabolites, which might explain the effect of *CBR1* variants seen in SR patients.

We also discovered an association between LVEF 2D and rare variants in the *AKR1C3* genes in standard-risk patients. The encoded enzyme is one of the main anthracycline-reducing cytosolic enzymes but may also play an important regulatory role in cell proliferation and differentiation[104,105]. Several *AKR1C3* variants, which alter the enzymatic activity, have been identified, in particular variants A106T, R170C, and P180S significantly reduced doxorubicin metabolism compared with the wild-type enzymes[74]. The protective effect of the *AKR1C3* gene variants identified in our study could be linked to reduced activity of the enzyme, a functional study is nevertheless needed to confirm this suggestion.

5.5.2. Variants detected through exome-wide analyses

5.5.2.1. NOD-like receptors (NOD2)

To further address the role of the genetic component in long-term chemotherapy-related cardiotoxicity we have used a hypothesis-free approach and detected an association between LVEF M-Mode and rare variants enrichment in the *NOD2* (*Nucleotide Binding Oligomerization Domain Containing 2,* also designated as *CARD15*) gene. NOD-like receptors play an important role in the innate immune response[106]. Current evidence has established a relationship between the immune response and cardiovascular disease[38,76,106-111]. In particular, several recent studies in the murine model have shown that NOD2 is one of the critical components of a signal transduction pathway that connects heart damage through exacerbation of inflammation[106]. Moreover, NOD2 is involved in the expansion of the lipid-rich necrotic core and promotes vascular inflammation in atherosclerosis in hypercholesterolemic mice[107]; NOD2 also debilitates cardiac hypertrophy and fibrosis (via regulation, among others, of MAPKs, and NF-κB, pathways)[108]. Genetic association studies of inflammation and immune response in the pathogenesis of atherosclerosis confirmed indeed that common polymorphisms in the *NOD2* gene may influence the risk of developing clinically evident coronary artery disease[76].

5.5.2.2. Zinc-finger proteins (ZNF267)

Another signal that reached an exome-wide significance level of association was detected within *ZNF267 (Zinc Finger Protein 267* also known as *human zinc finger 2* or *HZF2)* locus, suggesting a link between rare variants and higher values of FS. *ZNF267* belongs to the Kruppel-like (KLF) family of DNA binding transcription factors that are major regulators of tissue homeostasis[112]. In recent years, KLFs have been studied in the context of cardiovascular health and disease, and their critical function in cardiomyocyte remodeling has been proposed[112]. In particular, KLFs are involved in myocyte enlargement, fibrosis and changes in myocardial energy metabolism, and the pathogenesis of cardiac hypertrophy[112]. ZNF267 mRNA is up-regulated in response to the nitric oxide treatment in venous endothelial cells[113], whereas, studies of chronic liver disease, showed that ROS formation also increases ZNF267 mRNA expression[114]. Although the potential implication of the *ZNF267* gene in the oxidative stress response warrants further investigation, its role in chemotherapy-induced cardiotoxicity remains unknown.

5.6. CONCLUDING REMARKS

Special attention is currently paid to the contribution of rare and low-frequency variants to human traits and diseases. Furthermore, they could contribute significantly to the understanding of the role of the genetic component on the pathophysiology of cardiotoxicity induced by chemotherapy[115].

While several candidate genes and GWAS have reported the input of common genetic variants, the knowledge about the contribution of rare variants remains limited. In this study, using WES data and two complementary approaches of the genetic association studies, we demonstrated the contribution of rare genetic variants on long-term treatment-related cardiotoxicity in childhood ALL survivors.

We acknowledge that our study has certain limitations. For example, the limited sample size may affect the accuracy of the results, particularly in the context of the stratified analysis and different candidate gene groups. The association results obtained for rare variants should be taken with caution given their low number. None of the associations detected in the PETALE cohort were replicated in the SJLIFE cohort. This could be explained by several reasons. First of all, despite the application of similar inclusion/exclusion criteria and use of similar outcomes between the two cohorts, it is possible that small sample sizes in both cohorts, differences in treatment protocols (for example, in SJLIFE patients were not treated with dexrazoxane, and only 30% were treated with doxorubicin compared to 100% of PETALE participants) and/or time of ALL diagnosis[116] contributed to the observed discrepancies. Likewise, in some cases, only a limited number of discovery rare variants within a locus of interest, particularly those with extremely low minor allele frequencies, passed quality control and were not monomorphic in SJLIFE and were thus evaluable for replication. Nevertheless, we can not disregard the possibility that some of the associations observed in the discovery cohort could have been obtained by chance.

It is worth noting that even though identified genetic markers were not validated in the SJLIFE cohort, some of these loci have already been associated with treatment-related cardiotoxicity whereas others have been shown to play an important role in cardiovascular biology, suggesting that the current findings provide nevertheless further evidence of genetic contribution to the treatment-related cardiovascular outcomes in ALL patients. The identification of genetic markers associated with high or low risk of treatment-related cardiac complications (together with other known risk factors)[117], could enable their mitigation through individualized treatment adaptation, targeted therapies[118,119] or development of new prevention, intervention, and follow-up strategies.

5.7. SUMMARY POINTS

An important number of survivors of childhood acute lymphoblastic leukemia (ALL) suffer from treatment-related late adverse effects, with cardiovascular complications being one of the leading causes of treatment-related morbidity and mortality.

Significant associations derived through the candidate gene approach were detected among common variants rs72648998, rs3829747, and rs2303838 in the titin (*TTN*) gene, left ventricle ejection fraction (LVEF $p \leq 0.0005$), and fractional shortening (FS $p \leq 0.001$).

Rare variants enrichment in the *NOS1* and *ABCG2* of doxorubicin pathway ($p=0.001$ and $p=0.003$) was observed in relation to LVEF, whereas the exome-wide approach detected contribution of rare variants in *NOD2* to LVEF ($p=2x10\text{-}6$), and in *NOD2* and *ZNF267* genes to FS ($p=1.4x10\text{-}6$ and $p=3.5x10\text{-}6$, respectively).

Following stratification according to risk groups, the modulatory effect of rare variants was also found in the candidate *CBR1*, *ABCC5*, *ABCG2*, and *AKR1C3* genes ($p \leq 0.007$).

The association with common *TTN* variants in the independent St. Jude Lifetime Cohort (SJLIFE) was not significant.

Among 46 associated rare variants, 8 non-monomorphic rare variants were identified in SJLIFE and passed quality control but were not replicated.

Current findings may help to understand the influence of genetic factors on long-term treatment-related cardiovascular complications enabling their mitigation through individualized treatment approaches and preventive strategies.

Further studies are needed to confirm whether the described genetic markers may be useful in identifying patients at increased risk of these complications.

5.8. LIST OF ABBREVIATIONS

ALL: Acute lymphoblastic leukemia
PETALE: **P**révenir les **E**ffets tardifs des **T**raitements de la leucémie **A**iguë **L**ymphoblastique chez l'**E**nfant
SJUHC: Sainte-Justine University Health Center
DFCI: Dana-Faber Cancer Institute
LVEF M-MODE: Left Ventricular Ejection Fraction (M-mode; Teichholz)
LVEF 2D (Simpson): Left ventricular ejection fraction method of discs biplane (decimal)
LVFS M-MODE: Left ventricular fractional shortening (M-mode, decimal)
SNV: Single nucleotide variation
MAF: Minor allele frequency
SKAT-O test: Optimal Sequence Kernel Association Test
FDR: False discovery rate
SR: standard risk
HR: high risk
GWAS: Genome-wide association study
SJLIFE: St-Jude Lifetime Cohort Study

5.9. DECLARATIONS

Ethics approval and consent to participate

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of SJUHC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

Institute of Cancer Research (ICR) of the Canadian Institutes of Health Research (CIHR) ***grant number: 118694***, Cancer Research Society (CRS), Canadian Cancer Society Research Institute (CCSRI), Ontario Institute for Cancer Research (OICR), Pediatric Oncology Group of Ontario (POGO), Garron Family Cancer Centre at SickKids Hospital (Ontario), The Terry Fox foundation, FRQS Applied Medical Genetics Network and Sainte-Justine Hospital Foundation supported the PETALE study.

Kateryna Petrykey is a scholar of the Cole Foundation and the Université de Montréal.

Authors' contributions

D.S., M.K., G.U.A., and C.L. elaborated and supervised the study; S.D. and L.B. participated in its conception and coordination; D.S. supervised WES analyses; G.U.A. and M-J.R. contributed to the assessment of cardiovascular data; P.S.O. and P.B. contributed to bioinformatics data processing; A.R. contributed to patients' data processing; K.P. and M.L.G. executed computational and statistical

analysis; K.P. and M.K. performed interpretation of results; Y.Y., F.W, J.B. and M.H. performed the replication analysis; K.P. drafted the manuscript and all authors revised it critically.

Acknowledgements

The authors would like to thank all childhood ALL survivors and their parents for the participation in the PETALE study, as well as all study collaborators for their valuable contribution.

5.10. TABLES AND FIGURES

5.10.1. Table 1. PETALE cohort: patient demographics and clinical characteristics, N=233.

		N	%
Sex			
	Males	114	48.9
	Females	119	51.1
Prognostic risk group			
	Standard risk	106	45.5
	High risk*	127	54.5
DFCI protocol			
	(87-01)	18	7.73
	(91-01)	47	20.17
	(95-01)	68	29.18
	(00-01)	74	31.76
	(05-01)	26	11.16
Age at diagnosis - median (range)			
	4 (0-18)		
Time since the end of treatment			
	10 or more years	174	74.7
	Less than 10 years	59	25.3
	Median	**Min**	**Max**
	13	3	24
Dexrazoxane treatment in High-risk patients			
	Yes	68	53.5
	No	59	46.5
Dexrazoxane cumulative dose (mg/m2)			

Median	Max	Min
2961.9	3706.2	1433.8

Doxorubicin cumulative dose (mg/m2)

Median	Max	Min
209.3	472.9	41.3

Continuous outcomes				
	Mean	Median	Max	Min
FS	35.13	35.09	46.3	24
LVEF M-Mode (%)	60.05	59.87	77.3	47
LVEF 2D (%)	55.96	55.89	66.8	39
LVEDD (mm) z-score	0.649	0.675	4.28	-1.83

DFCI: Dana-Farber Cancer Institute; FS: fractional shortening; LVEF: left ventricular ejection fraction; LVEDD: Left Ventricular End Diastolic Diameter.

*Criteria for High-risk stratification were mainly attributed based on age, white blood cell count, immunophenotype (presence of T-cell markers) and combination of these factors; as well as central nervous system (CNS) status and Minimal residual disease at diagnosis.

5.10.2. Table 2. Top-ranking associations of the common variants' analysis, all cohort, N=233.

a. Associations detected between LVEF M-mode (%) and SNVs in the *TTN* gene.

Outcome	SNVs tested	Position (hg19)	P value	FDR*	Genotype	Genetic model			
					CC/CT/TT	CC/ CT+TT	Mean	SD	Beta** [CI, 0.95]
LVEF M-Mode (%)	rs72648998	chr2:179575511	0.0003	0.011	207/19/3	207/22	59.6/ 64.4	5.6/ 5.8	4.78 [2.3-7.3]
	rs3829747	chr2:179397561	0.0003	0.011	181/40/8	181/48	59.4/ 62.7	5.6/ 5.9	3.33 [1.5-5.2]
	rs2303838	chr2:179444939	0.0005	0.014	161/55/11	161/66	59.3/ 61.9	5.6/ 6.1	2.59 [0.9-4.2]

b. Associations detected between FS M-mode (%) and SNVs in the *TTN* gene.

Outcome	SNVs tested	Position (hg19)	P value	FDR*	Genotype	Genetic model			
					CC/CT/TT	CC/ CT+TT	Mean	SD	Beta** [CI, 0.95]
FS M-Mode (%)	rs72648998	chr2:179575511	0.0002	0.014	207/19/3	207/22	34.8/ 38.1	3.9/ 4.1	3.35 [1.6-5.1]
	rs3829747	chr2:179397561	0.0004	0.014	181/40/8	181/48	34.7/ 37.0	3.9/ 4.1	2.34 [1.1-3.6]
	rs2303838	chr2:179444939	0.001	0.03	161/55/11	161/66	34.7/ 36.3	3.9/ 4.3	1.68 [0.5-2.8]

SNV: single nucleotide variation; LVEF: Left ventricular ejection fraction; *TTN: titin*; FDR: false discovery rate; SD: standard deviation; FS: fractional shortening; OR: odds ratio; CI: confidence interval.
* All associations have FDR-BH (Benjamini-Hochberg false discovery rate) lower than 5%. Additionally, all associations have p value lower than 0.0006, except rs2303838 in relation to FS, which is Bonferroni cut-off value for the number of variants tested through the candidate gene approach.

**Unstandardized beta coefficient was obtained through the linear regression using the dominant genetic model.

5.10.3. Table 3. Multiple linear regression models including non-genetic covariates, *TTN* gene.

a. LVEF M-mode (%), all cohort, N=233.

SNV	TEST	BETA	L95	U95	P
rs72648998	*TTN*	0.24	0.12	0.37	**0.0002**
	Age at diagnosis	0.10	-0.04	0.24	0.1669
	Time end treatment	0.26	-0.10	0.63	0.1616
	Sex	0.02	-0.10	0.15	0.7182
	Protocol	0.36	-0.02	0.73	0.0637
	Risk-treatment*	-0.17	-0.32	-0.03	**0.0213**

b. FS M-mode (%), all cohort, N=233.

SNV	TEST	BETA	L95	U95	P
rs72648998	*TTN*	0.24	0.11	0.36	**0.0003**
	Age at diagnosis	0.09	-0.05	0.23	0.1957
	Time end treatment	0.05	-0.32	0.43	0.7747
	Sex	0.05	-0.08	0.17	0.4598
	Protocol	0.07	-0.31	0.45	0.7085
	Risk-treatment*	-0.18	-0.33	-0.03	**0.0162**

LVEF: left ventricular ejection fraction; SNV: single nucleotide variation; FS: fractional shortening.
TTN: *titin*.
Due to the partial linkage disequilibrium between reported *TTN* variants, their association with LVEF M-mode is not independent, however it is independent from the effect of several non-genetic covariables. The results for the variant with the strongest effect are presented in the table.
* This variable is combined based on risk category and dexrazoxane treatment, as defined in methodology (categorical variable).

5.10.4. Table 4. Top-ranking associations of the rare variants of doxorubicin pathway identified through SKAT-O test.

Outcome	Gene	SNVs*		P value SKAT-O**	Collapsed variants***				
		position	rs number		Genotype 11/12/22	11/ 12+22	Mean	SD	P value
LVEF M-mode (%)		**All cohort N=233**							
	NOS1	**chr12:117768154**	rs76090928	0.0013	227/4/0	227/4	59.9/ 71.4	5.6/ 2.3	0.00007
	ABCG2	**chr4:89052998**	rs199473672	0.0026	230/2/1	230/3	59.9/ 70.5	5.7/ 5.9	0.002
		SR group N=106							
	CBR1	**chr21:37444697**	rs2835266	0.0015	101/5/0	101/5	61.0/ 53.5	5.6/ 5.5	0.005
		HR group N=127							
	ABCC5	**chr3:183681203** **chr3:183681255**	. .	0.0008	125/2/0 119/8/0	117/10*	59.04/ 65.7	5.5/ 5.7	0.0003
FS M-mode (%)		**SR group N=106**							
	ABCG2	**chr4:89052998**	rs199473672	0.0027	104/1/1	104/2	35.4/ 43.9	3.7/ 3.1	0.002
	CBR1	**chr21:37444697**	rs2835266	0.0037	103/5/0	101/5	35.8/ 31.2	3.7/ 4.1	0.009
		HR group N=127							
	ABCC5	**chr3:183681203** **chr3:183681255**	. .	0.0014	125/2/0 119/8/0	117/10*	34.5/ 38.4	4.1/ 4.1	0.004
LVEF 2D (%)		**SR group N=106**							
	ABCG2	**chr4:89018670** **chr4:89052998** **chr4:89060981**	rs45605536 rs199473672 .	0.0019	104/1/0 103/1/1 103/1/0	100/4	55.7/ 63.4	4.5/ 3.4	0.001
	AKR1C3	**chr10:5139642** **chr10:5141609**	rs200981816 rs34186955	0.0067	104/1/0 104/1/0	103/2	55.9/ 65.4	4.5/ 0.99	0.003

SNV: single nucleotide variation; LVEF: left ventricular ejection fraction, HR: high risk; SR: standard risk; FS: fractional shortening; SD: standard deviation; MAF: minor allele frequency.
NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group); CBR1: Carbonyl Reductase 1; ABCC5: ATP Binding Cassette Subfamily C Member 5; AKR1C3: Aldo-Keto Reductase Family 1 Member C3.
*SNVs that are identified as the most important contributors to the association signal are presented.
** P value refers to the gene-level analysis with the SKAT-O test as reported in Supplemental table 2.
***Collapsed variants, carriers of at least one of the rare variants were counted as carriers in the model.

5.10.5. Table 5. Top-ranking associations of the rare variants identified through SKAT-O test in the exome-wide association study.

Outcome	Gene	All cohort N=233							
		SNVs*		P value SKAT-O**	Collapsed variants***				
		position	rs number		Genotype 11/12/22	11/ 12+22	Mean	SD	P value
FS M-mode (%)	*NOD2*	**chr16:50745114**	rs104895431	$1.39x10^{-6}$	228/3/0	215/7*	34.9/ 43.6	3.8/ 3.1	$7.6x10^{-9}$
		chr16:50746086	rs61747625		224/3/0				
		chr16:50746100	rs3813758		226/1/0				
		chr16:50746199	rs104895444		229/2/0				
	ZNF267	**chr16:31926312**	.	$3.51x10^{-6}$	232/1/0	216/7*	34.8/ 41.9	3.8/ 4.3	$2x10^{-6}$
		chr16:31927308	rs118056264		227/5/0				
		chr16:32077452	.		226/1/0				
LVEF M-mode (%)	*NOD2*	**chr16:50745114**	rs104895431	$2x10^{-6}$	228/3/0	215/7*	59.7/ 72.4	5.4/ 3.5	$3.6x10^{-9}$
		chr16:50746086	rs61747625		224/3/0				
		chr16:50746100	rs3813758		226/1/0				
		chr16:50746199	rs104895444		239/2/0				

SNV: single nucleotide variation; LVEF: Left ventricular ejection fraction; FS: fractional shortening; SD: standard deviation; MAF: minor allele frequency.
ZNF267: Zinc Finger Protein 267; NOD2: Nucleotide Binding Oligomerization Domain Containing 2.
*SNVs that are identified as the most important contributors to the association signal are presented.
** P value refers to the gene-level analysis with the SKAT-O test as reported in Supplemental table 4.
***Collapsed variants, carriers of at least one of the rare variants were counted as carriers in the model.

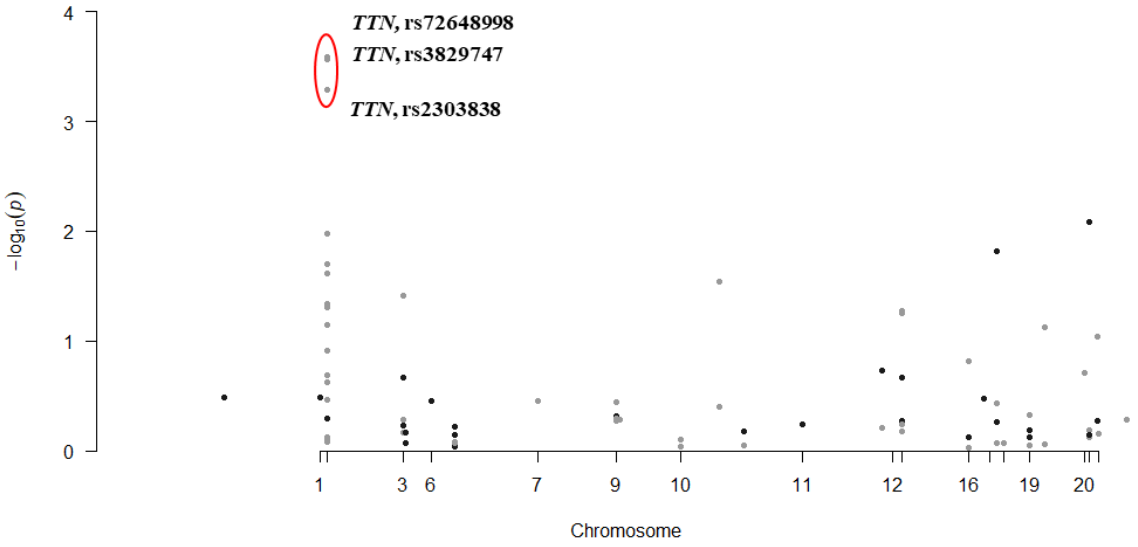


Figure 1. P value distribution of all common variants tested in the entire PETALE cohort for LVEF M-Mode (%) represented as a Manhattan plot.

Manhattan plot was created to display the statistical significance between all common genetic variants used in the candidate gene association study (data combined from three candidate genes datasets) and LVEF M-Mode (%) outcome and ranked according to their associated P value.

TTN: titin; LVEF: Left ventricular ejection fraction.

5.11. REFERENCES

1. Chow EJ, Leger KJ, Bhatt NS, Mulrooney DA, Ross CJ, Aggarwal S, et al. Paediatric cardio-oncology: epidemiology, screening, prevention, and treatment. Cardiovasc Res. 2019;115(5):922-34.
2. Fulbright JM. Review of cardiotoxicity in pediatric cancer patients: during and after therapy. Cardiol Res Pract. 2011;2011:942090. ***article of interest as discusses the late cardiac effects among long-term cancer survivors**
3. Hudson MM, Ness KK, Gurney JG, Mulrooney DA, Chemaitilly W, Krull KR, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. Jama. 2013;309(22):2371-81.
4. Adamiak M, Sahoo S. Exosomes in Myocardial Repair: Advances and Challenges in the Development of Next-Generation Therapeutics. Mol Ther. 2018;26(7):1635-43.
5. Haybar H, Shahrabi S, Rezaeeyan H, Shirzad R, Saki N. Endothelial Cells: From Dysfunction Mechanism to Pharmacological Effect in Cardiovascular Disease. Cardiovasc Toxicol. 2019;19(1):13-22.
6. Sardu C, Paolisso G, Marfella R. Inflammatory Related Cardiovascular Diseases: From Molecular Mechanisms to Therapeutic Targets. Curr Pharm Des. 2020;26(22):2565-73.
7. Shah MS, Brownlee M. Molecular and Cellular Mechanisms of Cardiovascular Disorders in Diabetes. Circulation research. 2016;118(11):1808-29.
8. Mitry MA, Edwards JG. Doxorubicin induced heart failure: Phenotype and molecular mechanisms. IJC Heart & Vasculature. 2016;10:17-24.
9. Yu H, Wang Z. Cardiomyocyte-Derived Exosomes: Biological Functions and Potential Therapeutic Implications. Frontiers in Physiology. 2019;10(1049).
10. Zadeh FJ, Ghasemi Y, Bagheri S, Maleknia M, Davari N, Rezaeeyan H. Do exosomes play role in cardiovascular disease development in hematological malignancy? Molecular Biology Reports. 2020;47(7):5487-93.
11. Farmakis D, Parissis J, Filippatos G. Insights into onco-cardiology: atrial fibrillation in cancer. J Am Coll Cardiol. 2014;63(10):945-53.
12. Farmakis D, Mantzourani M, Filippatos G. Anthracycline-induced cardiomyopathy: secrets and lies. European Journal of Heart Failure. 2018;20(5):907-9.
13. Octavia Y, Tocchetti CG, Gabrielson KL, Janssens S, Crijns HJ, Moens AL. Doxorubicin-induced cardiomyopathy: from molecular mechanisms to therapeutic strategies. J Mol Cell Cardiol. 2012;52(6):1213-25.
14. Olson RD, Mushlin PS. Doxorubicin cardiotoxicity: analysis of prevailing hypotheses. FASEB J. 1990;4(13):3076-86.
15. Chow EJ, Antal Z, Constine LS, Gardner R, Wallace WH, Weil BR, et al. New Agents, Emerging Late Effects, and the Development of Precision Survivorship. J Clin Oncol. 2018;36(21):2231-40.
16. Armenian SH, Lacchetti C, Lenihan D. Prevention and Monitoring of Cardiac Dysfunction in Survivors of Adult Cancers: American Society of Clinical Oncology Clinical Practice Guideline Summary. Journal of Oncology Practice. 2017;13(4):270-5.
17. Bansal N, Adams MJ, Ganatra S, Colan SD, Aggarwal S, Steiner R, et al. Strategies to prevent anthracycline-induced cardiotoxicity in cancer survivors. Cardiooncology. 2019;5:18-.
18. Lipshultz SE, Alvarez JA, Scully RE. Anthracycline associated cardiotoxicity in survivors of childhood cancer. Heart. 2008;94(4):525-33. ***article of interest as describes mechanisms of anthracycline-induced cardiotoxicity**
19. Armenian S, Bhatia S. Predicting and Preventing Anthracycline-Related Cardiotoxicity. American Society of Clinical Oncology Educational Book. 2018(38):3-12.
20. Marcoux S, Drouin S, Laverdiere C, Alos N, Andelfinger GU, Bertout L, et al. The PETALE study: Late adverse effects and biomarkers in childhood acute lymphoblastic leukemia survivors. Pediatr

Blood Cancer. 2017;64(6). **article of considerable interest as explains in detail the whole study design of PETALE cohort
21. Petrykey K, Lippe S, Robaey P, Sultan S, Laniel J, Drouin S, et al. Influence of genetic factors on long-term treatment related neurocognitive complications, and on anxiety and depression in survivors of childhood acute lymphoblastic leukemia: The Petale study. PLoS One. 2019;14(6):e0217314.
22. Schiller NB, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, et al. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two-Dimensional Echocardiograms. J Am Soc Echocardiogr. 1989;2(5):358-67.
23. Krajinovic M, Elbared J, Drouin S, Bertout L, Rezgui A, Ansari M, et al. Polymorphisms of ABCC5 and NOS3 genes influence doxorubicin cardiotoxicity in survivors of childhood acute lymphoblastic leukemia. Pharmacogenomics J. 2016;16(6):530-5. **article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity
24. Dolgin M FA, Gorlin R, Levin RI, New York Heart Association. Criteria Committee. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. . 1994.
25. Kampmann C, Wiethoff CM, Wenzel A, Stolz G, Betancor M, Wippermann CF, et al. Normal values of M mode echocardiographic measurements of more than 2000 healthy infants and children in central Europe. Heart (British Cardiac Society). 2000;83(6):667-72.
26. Pfaffenberger S, Bartko P, Graf A, Pernicka E, Babayev J, Lolic E, et al. Size matters! Impact of age, sex, height, and weight on the normal heart size. Circ Cardiovasc Imaging. 2013;6(6):1073-9.
27. Spinella JF, Healy J, Saillour V, Richer C, Cassart P, Ouimet M, et al. Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. BMC Cancer. 2015;15:539.
28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
29. Picard Tools - By Broad Institute: Code ZIP File; 2018 [Available from: https://broadinstitute.github.io/picard/.
30. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
31. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
32. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.
33. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet. 2006;7:61-80.
34. The 1000 Genomes Browsers | 1000 Genomes 2018 [Available from: http://www.internationalgenome.org/1000-genomes-browsers/.
35. Exome Variant Server 2018 [Available from: http://evs.gs.washington.edu/EVS/.
36. Longley DB, Johnston PG. Molecular mechanisms of drug resistance. J Pathol. 2005;205(2):275-92.
37. Tacar O, Sriamornsak P, Dass CR. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. J Pharm Pharmacol. 2013;65(2):157-70.
38. Genotype-Tissue Expression (GTEx) [Internet]. Available from: https://gtexportal.org/home/.
39. Gewirtz DA. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. Biochem Pharmacol. 1999;57(7):727-41.
40. Thorn CF, Oshiro C, Marsh S, Hernandez-Boussard T, McLeod H, Klein TE, et al. Doxorubicin pathways: pharmacodynamics and adverse effects. Pharmacogenetics and genomics. 2011;21(7):440-6.
41. Bray AW, Ballinger SW. Mitochondrial DNA mutations and cardiovascular disease. Curr Opin Cardiol. 2017;32(3):267-74.

42. Madamanchi NR, Vendrov A, Runge MS. Oxidative stress and vascular disease. Arterioscler Thromb Vasc Biol. 2005;25(1):29-38.
43. KEGG: Kyoto Encyclopedia of Genes and Genomes 2018 [Available from: https://www.genome.jp/kegg/.
44. Whole genome association analysis toolset [Available from: http://zzz.bwh.harvard.edu/plink/index.shtml.
45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
46. Barry E, Alvarez JA, Scully RE, Miller TL, Lipshultz SE. Anthracycline-induced cardiotoxicity: course, pathophysiology, prevention and management. Expert Opinion on Pharmacotherapy. 2007;8(8):1039-58. ***article of interest as describes mechanisms of anthracycline-induced cardiotoxicity**
47. Lipshultz SE, Colan SD, Gelber RD, Perez-Atayde AR, Sallan SE, Sanders SP. Late cardiac effects of doxorubicin therapy for acute lymphoblastic leukemia in childhood. N Engl J Med. 1991;324(12):808-15. *** article of interest as discusses late cardiac effects among long-term cancer survivors**
48. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279-84.
49. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. Genetics. 2005;171(2):783-90.
50. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. American journal of human genetics. 2012;91(2):224-37.
51. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. American journal of human genetics. 2010;86(6):929-42.
52. CRAN - Package SKAT 2019 [Available from: https://cran.r-project.org/web/packages/SKAT/index.html.
53. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. Genet Epidemiol. 2011;35 Suppl 1:S12-7.
54. Sung YJ, Korthauer KD, Swartz MD, Engelman CD. Methods for collapsing multiple rare variants in whole-genome sequence data. Genet Epidemiol. 2014;38 Suppl 1:S13-20.
55. Dixon SB, Chen Y, Yasui Y, Pui CH, Hunger SP, Silverman LB, et al. Reduced Morbidity and Mortality in Survivors of Childhood Acute Lymphoblastic Leukemia: A Report From the Childhood Cancer Survivor Study. J Clin Oncol. 2020;38(29):3418-29.
56. The Genome Aggregation Database (gnomAD).
57. Lipshultz SE, Karnik R, Sambatakos P, Franco VI, Ross SW, Miller TL. Anthracycline-related cardiotoxicity in childhood cancer survivors. Curr Opin Cardiol. 2014;29(1):103-12.
58. Mulrooney DA, Yeazel MW, Kawashima T, Mertens AC, Mitby P, Stovall M, et al. Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort. Bmj. 2009;339:b4606.
59. Garcia-Pavia P, Kim Y, Restrepo-Cordoba MA, Lunde IG, Wakimoto H, Smith AM, et al. Genetic Variants Associated With Cancer Therapy–Induced Cardiomyopathy. Circulation. 2019;140(1):31-41. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
60. Aminkeng F, Bhavsar AP, Visscher H, Rassekh SR, Li Y, Lee JW, et al. A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. Nat Genet. 2015;47(9):1079-84. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**

61. Blanco JG, Sun C-L, Landier W, Chen L, Esparza-Duran D, Leisenring W, et al. Anthracycline-related cardiomyopathy after childhood cancer: role of polymorphisms in carbonyl reductase genes--a report from the Children's Oncology Group. J Clin Oncol. 2012;30(13):1415-21. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
62. Hildebrandt MAT, Reyes M, Wu X, Pu X, Thompson KA, Ma J, et al. Hypertension Susceptibility Loci are Associated with Anthracycline-related Cardiotoxicity in Long-term Childhood Cancer Survivors. Scientific Reports. 2017;7(1):9698. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
63. Lipshultz SE, Lipsitz SR, Kutok JL, Miller TL, Colan SD, Neuberg DS, et al. Impact of hemochromatosis gene mutations on cardiac status in doxorubicin-treated survivors of childhood high-risk leukemia. Cancer. 2013;119(19):3555-62. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
64. Rajić V, Aplenc R, Debeljak M, Prestor VV, Karas-Kuželicki N, MlinariČ-RašČan I, et al. Influence of the polymorphism in candidate genes on late cardiac damage in patients treated due to acute leukemia in childhood. Leukemia & Lymphoma. 2009;50(10):1693-8. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
65. Sági JC, Egyed B, Kelemen A, Kutszegi N, Hegyi M, Gézsi A, et al. Possible roles of genetic variations in chemotherapy related cardiotoxicity in pediatric acute lymphoblastic leukemia and osteosarcoma. BMC Cancer. 2018;18(1):704. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
66. Visscher H, Rassekh SR, Sandor GS, Caron HN, van Dalen EC, Kremer LC, et al. Genetic variants in SLC22A17 and SLC22A7 are associated with anthracycline-induced cardiotoxicity in children. Pharmacogenomics. 2015;16(10):1065-76. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
67. Visscher H, Ross CJD, Rassekh SR, Barhdadi A, Dubé M-P, Al-Saloos H, et al. Pharmacogenomic Prediction of Anthracycline-Induced Cardiotoxicity in Children. Journal of Clinical Oncology. 2012;30(13):1422-8. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
68. Visscher H, Ross CJD, Rassekh SR, Sandor GSS, Caron HN, van Dalen EC, et al. Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. Pediatric Blood & Cancer. 2013;60(8):1375-81. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
69. Wang X, Liu W, Sun CL, Armenian SH, Hakonarson H, Hageman L, et al. Hyaluronan synthase 3 variant and anthracycline-related cardiomyopathy: a report from the children's oncology group. J Clin Oncol. 2014;32(7):647-53. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
70. Wang X, Sun CL, Quinones-Lombrana A, Singh P, Landier W, Hageman L, et al. CELF4 Variant and Anthracycline-Related Cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study. J Clin Oncol. 2016;34(8):863-70. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
71. Windsor RE, Strauss SJ, Kallis C, Wood NE, Whelan JS. Germline genetic polymorphisms may influence chemotherapy response and disease outcome in osteosarcoma: a pilot study. Cancer. 2012;118(7):1856-67. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
72. Fohner AE, Brackman DJ, Giacomini KM, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for ABCG2. Pharmacogenet Genomics. 2017;27(11):420-7.
73. Lubieniecka JM, Graham J, Heffner D, Mottus R, Reid R, Hogge D, et al. A discovery study of daunorubicin induced cardiotoxicity in a sample of acute myeloid leukemia patients prioritizes P450 oxidoreductase polymorphisms as a potential risk factor. Front Genet. 2013;4:231.

74. Bains OS, Grigliatti TA, Reid RE, Riggs KW. Naturally occurring variants of human aldo-keto reductases with reduced in vitro metabolism of daunorubicin and doxorubicin. J Pharmacol Exp Ther. 2010;335(3):533-45.
75. El Mokhtari NE, Ott SJ, Nebel A, Schäfer A, Rosenstiel P, Förster M, et al. Role of NOD2/CARD15 in coronary heart disease. BMC Genet. 2007;8:76.
76. Galluzzo S, Patti G, Dicuonzo G, Di Sciascio G, Tonini G, Ferraro E, et al. Association between NOD2/CARD15 polymorphisms and coronary artery disease: a case-control study. Hum Immunol. 2011;72(8):636-40.
77. Gerull B. Between Disease-Causing and an Innocent Bystander: The Role of Titin as a Modifier in Hypertrophic Cardiomyopathy. Canadian Journal of Cardiology. 2017;33(10):1217-20.
78. Ware JS, Cook SA. Role of titin in cardiomyopathy: from DNA variants to patient stratification. Nature Reviews Cardiology. 2018;15(4):241-52.
79. Ng D, Johnston JJ, Teer JK, Singh LN, Peller LC, Wynter JS, et al. Interpreting secondary cardiac disease variants in an exome cohort. Circ Cardiovasc Genet. 2013;6(4):337-46.
80. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of Titin Causing Dilated Cardiomyopathy. New England Journal of Medicine. 2012;366(7):619-28.
81. van Spaendonck-Zwarts KY, Posafalvi A, van den Berg MP, Hilfiker-Kleiner D, Bollen IA, Sliwa K, et al. Titin gene mutations are common in families with both peripartum cardiomyopathy and dilated cardiomyopathy. Eur Heart J. 2014;35(32):2165-73.
82. Ware JS, Li J, Mazaika E, Yasso CM, DeSouza T, Cappola TP, et al. Shared Genetic Predisposition in Peripartum and Dilated Cardiomyopathies. N Engl J Med. 2016;374(3):233-41.
83. Carmignac V, Salih MA, Quijano-Roy S, Marchand S, Al Rayess MM, Mukhtar MM, et al. C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. Ann Neurol. 2007;61(4):340-51.
84. Ceyhan-Birsoy O, Agrawal PB, Hidalgo C, Schmitz-Abe K, DeChene ET, Swanson LC, et al. Recessive truncating titin gene, TTN, mutations presenting as centronuclear myopathy. Neurology. 2013;81(14):1205-14.
85. Chauveau C, Bonnemann CG, Julien C, Kho AL, Marks H, Talim B, et al. Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. Hum Mol Genet. 2014;23(4):980-91.
86. Fernández-Marmiesse A, Carrascosa-Romero MC, Alfaro Ponce B, Nascimento A, Ortez C, Romero N, et al. Homozygous truncating mutation in prenatally expressed skeletal isoform of TTN gene results in arthrogryposis multiplex congenita and myopathy without cardiac involvement. Neuromuscul Disord. 2017;27(2):188-92.
87. Rankinen T, Rice T, Boudreau A, Leon AS, Skinner JS, Wilmore JH, et al. Titin is a candidate gene for stroke volume response to endurance training: the HERITAGE Family Study. Physiological Genomics. 2003;15(1):27-33.
88. Caru M, Petrykey K, Drouin S, Beaulieu P, St-Onge P, Lemay V, et al. Identification of genetic association between cardiorespiratory fitness and the trainability genes in childhood acute lymphoblastic leukemia survivors. BMC cancer. 2019;19(1):443-.
89. Linschoten M, Teske AJ, Baas AF, Vink A, Dooijes D, Baars HF, et al. Truncating Titin (TTN) Variants in Chemotherapy-Induced Cardiomyopathy. J Card Fail. 2017;23(6):476-9. ****article of considerable interest as describes genetic variants related to anthracycline-induced cardiotoxicity**
90. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nature genetics. 2014;46(8):826-36.
91. Megias-Vericat JE, Martinez-Cuadron D, Herrero MJ, Alino SF, Poveda JL, Sanz MA, et al. Pharmacogenetics of Metabolic Genes of Anthracyclines in Acute Myeloid Leukemia. Curr Drug Metab. 2018;19(1):55-74. ***article of interest as describes mechanisms of anthracycline-induced cardiotoxicity**

92. Garner AP, Paine MJ, Rodriguez-Crespo I, Chinje EC, Ortiz De Montellano P, Stratford IJ, et al. Nitric oxide synthases catalyze the activation of redox cycling and bioreductive anticancer agents. Cancer research. 1999;59(8):1929-34.
93. Zhang YH. Neuronal nitric oxide synthase in hypertension - an update. Clin Hypertens. 2016;22:20-.
94. Damy T, Ratajczak P, Shah AM, Camors E, Marty I, Hasenfuss G, et al. Increased neuronal nitric oxide synthase-derived NO production in the failing human heart. The Lancet. 2004;363(9418):1365-7.
95. Ward ME, Toporsian M, Scott JA, Teoh H, Govindaraju V, Quan A, et al. Hypoxia induces a functionally significant and translationally efficient neuronal NO synthase mRNA variant. J Clin Invest. 2005;115(11):3128-39.
96. Torgerson DG, Giri T, Druley TE, Zheng J, Huntsman S, Seibold MA, et al. Pooled Sequencing of Candidate Genes Implicates Rare Variants in the Development of Asthma Following Severe RSV Bronchiolitis in Infancy. PLoS One. 2015;10(11):e0142649.
97. Meissner K, Heydrich B, Jedlitschky G, Meyer Zu Schwabedissen H, Mosyagin I, Dazert P, et al. The ATP-binding cassette transporter ABCG2 (BCRP), a marker for side population stem cells, is expressed in human heart. J Histochem Cytochem. 2006;54(2):215-21.
98. Olarte Carrillo I, Ramos Peñafiel C, Miranda Peralta E, Rozen Fuller E, Kassack Ipiña JJ, Centeno Cruz F, et al. Clinical significance of the ABCB1 and ABCG2 gene expression levels in acute lymphoblastic leukemia. Hematology. 2017;22(5):286-91.
99. Megias-Vericat JE, Montesinos P, Herrero MJ, Moscardo F, Boso V, Rojas L, et al. Impact of ABC single nucleotide polymorphisms upon the efficacy and toxicity of induction chemotherapy in acute myeloid leukemia. Leuk Lymphoma. 2017;58(5):1197-206.
100. Semsei AF, Erdelyi DJ, Ungvari I, Csagoly E, Hegyi MZ, Kiszel PS, et al. ABCC1 polymorphisms in anthracycline-induced cardiotoxicity in childhood acute lymphoblastic leukaemia. Cell Biol Int. 2012;36(1):79-86.
101. Oppermann U. Carbonyl Reductases: The Complex Relationships of Mammalian Carbonyl- and Quinone-Reducing Enzymes and Their Role in Physiology. Annual Review of Pharmacology and Toxicology. 2007;47(1):293-322.
102. Blanco JG, Leisenring WM, Gonzalez-Covarrubias VM, Kawashima TI, Davies SM, Relling MV, et al. Genetic polymorphisms in the carbonyl reductase 3 gene CBR3 and the NAD(P)H:quinone oxidoreductase 1 gene NQO1 in patients who developed anthracycline-related congestive heart failure after childhood cancer. Cancer. 2008;112(12):2789-95.
103. Olson LE, Bedja D, Alvey SJ, Cardounel AJ, Gabrielson KL, Reeves RH. Protection from doxorubicin-induced cardiac toxicity in mice with a null allele of carbonyl reductase 1. Cancer research. 2003;63(20):6602-6.
104. Novotná E, Büküm N, Hofman J, Flaxová M, Kouklíková E, Louvarová D, et al. Aldo-keto reductase 1C3 (AKR1C3): a missing piece of the puzzle in the dinaciclib interaction profile. Archives of Toxicology. 2018;92(9):2845-57.
105. Birtwistle J, Hayden RE, Khanim FL, Green RM, Pearce C, Davies NJ, et al. The aldo-keto reductase AKR1C3 contributes to 7,12-dimethylbenz(a)anthracene-3,4-dihydrodiol mediated oxidative DNA damage in myeloid cells: implications for leukemogenesis. Mutat Res. 2009;662(1-2):67-74.
106. Li X, Li F, Chu Y, Wang X, Zhang H, Hu Y, et al. NOD2 Deficiency Protects against Cardiac Remodeling after Myocardial Infarction in Mice. Cellular Physiology and Biochemistry. 2013;32(6):1857-66.
107. Johansson ME, Zhang XY, Edfeldt K, Lundberg AM, Levin MC, Borén J, et al. Innate immune receptor NOD2 promotes vascular inflammation and formation of lipid-rich necrotic cores in hypercholesterolemic mice. Eur J Immunol. 2014;44(10):3081-92.
108. Zong J, Salim M, Zhou H, Bian ZY, Dai J, Yuan Y, et al. NOD2 deletion promotes cardiac hypertrophy and fibrosis induced by pressure overload. Lab Invest. 2013;93(10):1128-36.
109. Nishio H, Kanno S, Onoyama S, Ikeda K, Tanaka T, Kusuhara K, et al. Nod1 ligands induce site-specific vascular inflammation. Arterioscler Thromb Vasc Biol. 2011;31(5):1093-9.

110. Fernández-Velasco M, Prieto P, Terrón V, Benito G, Flores JM, Delgado C, et al. NOD1 activation induces cardiac dysfunction and modulates cardiac fibrosis and cardiomyocyte apoptosis. PLoS One. 2012;7(9):e45260.
111. Frantz S, Ertl G, Bauersachs J. Mechanisms of disease: Toll-like receptors in cardiovascular disease. Nat Clin Pract Cardiovasc Med. 2007;4(8):444-54.
112. Prosdocimo DA, Sabeh MK, Jain MK. Kruppel-like factors in muscle health and disease. Trends Cardiovasc Med. 2015;25(4):278-87.
113. Schäfer U, Schneider A, Neugebauer E. Identification of a nitric oxide-regulated zinc finger containing transcription factor using motif-directed differential display. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2000;1494(3):269-76.
114. Schnabl B, Czech B, Valletta D, Weiss TS, Kirovski G, Hellerbrand C. Increased expression of zinc finger protein 267 in non-alcoholic fatty liver disease. Int J Clin Exp Pathol. 2011;4(7):661-6.
115. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biology. 2017;18(1):77.
116. Hudson MM, Ness KK, Nolan VG, Armstrong GT, Green DM, Morris EB, et al. Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. Pediatric blood & cancer. 2011;56(5):825-36. ****article of considerable interest as explains in detail the whole study design of St. Jude Lifetime Cohort**
117. Aminkeng F, Ross CJ, Rassekh SR, Hwang S, Rieder MJ, Bhavsar AP, et al. Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. Br J Clin Pharmacol. 2016;82(3):683-95.
118. Mullighan CG, Jeha S, Pei D, Payne-Turner D, Coustan-Smith E, Roberts KG, et al. Outcome of children with hypodiploid ALL treated with risk-directed therapy based on MRD levels. Blood. 2015;126(26):2896-9.
119. Shahjahani M, Abroun A, Saki N, Bagher Mohammadi SM, Rezaeeyan H. STAT5: From Pathogenesis Mechanism to Therapeutic Approach in Acute Leukemia. Lab Med. 2020;51(4):345-51.

5.12. SUPPLEMENTAL MATERIAL

5.12.1. Supplemental Table 1. Complete list of candidate genes and SNVs.

Common variants

Cardiovascular function data set		
Gene	**SNV position**	**MAF (PETALE cohort)**
BAG3	chr10:121436286-121436286	0.085
CACNB2	chr10:18828635-18828635	0.157
COX15	chr10:101473218-101473218	0.115
CRYAB	chr11:111781047-111781047	0.295
DSG2	chr18:29104714-29104714	0.070
DSG2	chr18:29122618-29122618	0.075
EYA4	chr6:133789728-133789728	0.375
FKTN	chr9:108366734-108366734	0.292
JAG1	chr20:10628721-10628721	0.085
JAG1	chr20:10622501-10622501	0.066
JPH2	chr20:42747247-42747247	0.129
KCNH2	chr7:150645534-150645534	0.236
LAMA4	chr6:112522852-112522852	0.057
LAMA4	chr6:112457383-112457383	0.275
LMNA	chr1:156096387-156096387	0.057
LMNA	chr1:156099669-156099669	0.090
MYBPC3	chr11:47371598-47371598	0.082
MYH6	chr14:23859610-23859610	0.133
MYH6	chr14:23876267-23876267	0.129
MYOM1	chr18:3188976-3188976	0.465
MYOM1	chr18:3086065-3086065	0.056
MYPN	chr10:69934258-69934258	0.466
MYPN	chr10:69959242-69959242	0.447
NDUFAF1	chr15:41679685-41679685	0.178
NEBL	chr10:21139389-21139389	0.097
NEXN	chr1:78392446-78392446	0.113
SGCD	chr5:155935708-155935708	0.072
SYNE1	chr6:152453291-152453291	0.087
SYNE1	chr6:152772264-152772264	0.439
SYNE1	chr6:152443744-152443744	0.122
TBX1	chr22:19766782-19766782	0.206
TBX5	chr12:114793240-114793240	0.322
TMPO	chr12:98927830-98927830	0.093
TTN	chr2:179397561-179397561	0.125

TTN	chr2:179444939-179444939	0.172
TTN	chr2:179575511-179575511	0.054
TTN	chr2:179558366-179558366	0.234
TTN	chr2:179464527-179464527	0.233
TTN	chr2:179650408-179650408	0.252
TTN	chr2:179659912-179659912	0.057
TTN	chr2:179582537-179582537	0.210
TTN	chr2:179644855-179644855	0.136
TTN	chr2:179498042-179498042	0.104
TTN	chr2:179554305-179554305	0.351
TTN	chr2:179542464-179542464	0.063
TTN	chr2:179439880-179439880	0.059
TTN	chr2:179406294-179406294	0.084
TTN	chr2:179432185-179432185	0.294
TTN	chr2:179590256-179590256	0.087
TTN	chr2:179606538-179606538	0.071
TXNRD2	chr22:19907099-19907099	0.398
Doxorubicin pathway data set		
ABCB1	chr7:87160618-87160618	0.400
ABCB5	chr7:20687604-20687604	0.129
ABCB5	chr7:20691047-20691047	0.210
ABCB8	chr7:150732812-150732812	0.051
ABCC1	chr16:16173232-16173232	0.064
ABCC2	chr10:101553324-101553324	0.091
ABCC3	chr17:48761053-48761053	0.062
ABCC3	chr17:48712711-48712711	0.364
ABCC4	chr13:95859035-95859035	0.057
CBR1	chr21:37444120-37444120	0.128
CBR3	chr21:37518706-37518706	0.341
ERBB2	chr17:37884037-37884037	0.400
NOS2	chr17:26096597-26096597	0.182
NQO1	chr16:69748869-69748869	0.066
NQO1	chr16:69745145-69745145	0.172
SLC22A16	chr6:110763875-110763875	0.087
Mitochondrion function data set		
COX17	chr3:119395799-119395799	0.285
CRTC1	chr19:18876309-18876309	0.135
NCOA3	chr20:46264888-46264888	0.104
NCOA3	chr20:46256424-46256424	0.053
NDUFA10	chr2:240946766-240946766	0.364
NDUFA10	chr2:240923050-240923050	0.425
NDUFA11	chr19:5893058-5893058	0.091
NDUFA11	chr19:5892954-5892954	0.346

NDUFA6	chr22:42486723-42486723	0.317
NDUFS2	chr1:161182208-161182208	0.102
NDUFV2	chr18:9122611-9122611	0.053
NDUFV3	chr21:44324365-44324365	0.447
SIRT3	chr11:236091-236091	0.105

Rare variants

Cardiovascular function data set		
Gene	**SNV position**	**MAF (PETALE cohort)**
ABCC9	chr12:22005346-22005346	0.002
ABCC9	chr12:22040812-22040812	0.004
ABCC9	chr12:22063777-22063777	0.002
ABCC9	chr12:22078391-22078391	0.003
ACTN2	chr1:236907958-236907958	0.002
ACTN2	chr1:236908011-236908011	0.006
ACTN2	chr1:236917271-236917271	0.006
AKAP9	chr7:91603056-91603056	0.002
AKAP9	chr7:91622199-91622199	0.002
AKAP9	chr7:91631177-91631177	0.002
AKAP9	chr7:91659251-91659251	0.004
AKAP9	chr7:91670154-91670154	0.002
AKAP9	chr7:91672039-91672039	0.002
AKAP9	chr7:91674339-91674339	0.002
AKAP9	chr7:91674405-91674405	0.002
AKAP9	chr7:91682142-91682142	0.004
AKAP9	chr7:91690677-91690677	0.002
AKAP9	chr7:91694604-91694604	0.002
AKAP9	chr7:91695782-91695782	0.004
AKAP9	chr7:91708459-91708459	0.002
AKAP9	chr7:91714884-91714884	0.002
AKAP9	chr7:91714925-91714925	0.002
AKAP9	chr7:91726475-91726475	0.004
AKAP9	chr7:91726481-91726481	0.038
AKAP9	chr7:91726527-91726527	0.002
AKAP9	chr7:91731917-91731917	0.002
AKAP9	chr7:91737827-91737827	0.002
ANK2	chr4:114067061-114067061	0.003
ANK2	chr4:114171003-114171003	0.017
ANK2	chr4:114203904-114203904	0.002
ANK2	chr4:114203975-114203975	0.002
ANK2	chr4:114251586-114251586	0.006
ANK2	chr4:114264279-114264279	0.004
ANK2	chr4:114267122-114267122	0.002

ANK2	chr4:114274210-114274210	0.002
ANK2	chr4:114274366-114274366	0.002
ANK2	chr4:114274428-114274428	0.002
ANK2	chr4:114275548-114275548	0.013
ANK2	chr4:114275802-114275802	0.004
ANK2	chr4:114275980-114275980	0.002
ANK2	chr4:114276520-114276520	0.002
ANK2	chr4:114279057-114279057	0.002
ANK2	chr4:114279058-114279058	0.002
ANK2	chr4:114279628-114279628	0.011
ANK2	chr4:114279674-114279674	0.034
ANK2	chr4:114288920-114288920	0.004
ANK2	chr4:114293725-114293725	0.010
ANK2	chr4:114294462-114294462	0.011
ANK2	chr4:114294537-114294537	0.002
ANKRD1	chr10:92675322-92675322	0.004
ANKRD1	chr10:92675343-92675343	0.008
ANKRD1	chr10:92675944-92675944	0.002
ANKRD1	chr10:92678728-92678728	0.002
CACNA1C	chr12:2224449-2224449	0.004
CACNA1C	chr12:2602367-2602367	0.002
CACNA1C	chr12:2602428-2602428	0.004
CACNA1C	chr12:2694651-2694651	0.002
CACNA1C	chr12:2702450-2702450	0.009
CACNA1C	chr12:2702451-2702451	0.009
CACNA1C	chr12:2705099-2705099	0.002
CACNA1C	chr12:2711126-2711126	0.002
CACNA1C	chr12:2743475-2743475	0.002
CACNA1C	chr12:2788733-2788733	0.007
CACNA1C	chr12:2795368-2795368	0.004
CACNA2D1	chr7:81591256-81591256	0.002
CACNA2D1	chr7:81596943-81596943	0.004
CACNA2D1	chr7:81598203-81598203	0.002
CACNA2D1	chr7:81635140-81635140	0.004
CACNA2D1	chr7:81679904-81679904	0.002
CACNA2D1	chr7:81689805-81689805	0.002
CACNB2	chr10:18823110-18823110	0.004
CACNB2	chr10:18828181-18828181	0.004
CACNB2	chr10:18828309-18828309	0.002
CACNB2	chr10:18828486-18828486	0.034

CACNB2	chr10:18828598-18828598	0.011
CASQ2	chr1:116247824-116247824	0.002
CASQ2	chr1:116310934-116310934	0.002
CASQ2	chr1:116310937-116310937	0.006
COX15	chr10:101478161-101478161	0.002
COX15	chr10:101478205-101478205	0.015
COX15	chr10:101480780-101480780	0.004
COX15	chr10:101486775-101486775	0.002
COX15	chr10:101491735-101491735	0.004
CSRP3	chr11:19204274-19204274	0.004
CSRP3	chr11:19204275-19204275	0.004
CSRP3	chr11:19213980-19213980	0.002
CSRP3	chr11:19213986-19213986	0.006
DES	chr2:220284873-220284873	0.002
DES	chr2:220284876-220284876	0.015
DES	chr2:220284998-220284998	0.002
DES	chr2:220285637-220285637	0.002
DES	chr2:220290449-220290449	0.002
DSG2	chr18:29099844-29099844	0.002
DSG2	chr18:29099850-29099850	0.002
DSG2	chr18:29101156-29101156	0.006
DSG2	chr18:29104840-29104840	0.002
DSP	chr6:7558439-7558439	0.002
DSP	chr6:7565727-7565727	0.029
DSP	chr6:7576619-7576619	0.002
DSP	chr6:7580126-7580126	0.002
DSP	chr6:7580958-7580958	0.044
DSP	chr6:7581032-7581032	0.011
DSP	chr6:7581546-7581546	0.034
DSP	chr6:7581641-7581641	0.002
DSP	chr6:7581774-7581774	0.004
DSP	chr6:7582993-7582993	0.013
DSP	chr6:7583703-7583703	0.017
DSP	chr6:7584376-7584376	0.002
DSP	chr6:7584708-7584708	0.002
DSP	chr6:7585411-7585411	0.002
DSP	chr6:7585419-7585419	0.004
DSP	chr6:7585489-7585489	0.004
DSP	chr6:7585804-7585804	0.002
DTNA	chr18:32374109-32374109	0.002

DTNA	chr18:32400878-32400878	0.002
DTNA	chr18:32407586-32407586	0.002
DTNA	chr18:32418088-32418088	0.002
DTNA	chr18:32418749-32418749	0.004
DTNA	chr18:32418752-32418752	0.006
DTNA	chr18:32455297-32455297	0.002
DTNA	chr18:32459599-32459599	0.006
DTNA	chr18:32459679-32459679	0.004
DTNA	chr18:32462088-32462088	0.002
EYA4	chr6:133785956-133785956	0.036
EYA4	chr6:133789765-133789765	0.004
EYA4	chr6:133833868-133833868	0.002
FKTN	chr9:108363426-108363426	0.031
FKTN	chr9:108366499-108366499	0.019
FKTN	chr9:108397495-108397495	0.017
FKTN	chr9:108397496-108397496	0.002
GAA	chr17:78078656-78078656	0.038
GAA	chr17:78078846-78078846	0.002
GAA	chr17:78081504-78081504	0.002
GAA	chr17:78084553-78084553	0.002
GAA	chr17:78085871-78085871	0.004
GAA	chr17:78086777-78086777	0.005
GAA	chr17:78092031-78092031	0.002
GAA	chr17:78092585-78092585	0.004
GATA4	chr8:11615875-11615875	0.002
GATA4	chr8:11615928-11615928	0.002
GATA4	chr8:11615936-11615936	0.002
GATA4	chr8:11615955-11615955	0.002
HCN4	chr15:73614857-73614857	0.005
HCN4	chr15:73616505-73616505	0.002
HCN4	chr15:73616506-73616506	0.002
HCN4	chr15:73616507-73616507	0.002
HCN4	chr15:73622035-73622035	0.004
JAG1	chr20:10625561-10625561	0.002
JAG1	chr20:10625592-10625592	0.015
JAG1	chr20:10629711-10629711	0.004
JAG1	chr20:10639244-10639244	0.002
JPH2	chr20:42815290-42815290	0.002
JPH2	chr20:42815294-42815294	0.004
JPH2	chr20:42815301-42815301	0.006

JUP	chr17:39914695-39914695	0.002
JUP	chr17:39915108-39915108	0.004
JUP	chr17:39921270-39921270	0.006
JUP	chr17:39925713-39925713	0.030
JUP	chr17:39927995-39927995	0.006
KCNA5	chr12:5153825-5153825	0.002
KCNA5	chr12:5154058-5154058	0.006
KCNA5	chr12:5154064-5154064	0.006
KCNA5	chr12:5154437-5154437	0.006
KCNA5	chr12:5154747-5154747	0.004
KCNE2	chr21:35742799-35742799	0.011
KCNE2	chr21:35742814-35742814	0.004
KCNE2	chr21:35742955-35742955	0.002
KCNH2	chr7:150645550-150645550	0.002
KCNH2	chr7:150648643-150648643	0.002
KCNH2	chr7:150648644-150648644	0.002
KCNH2	chr7:150649923-150649923	0.002
KCNJ5	chr11:128781289-128781289	0.002
KCNJ5	chr11:128781805-128781805	0.002
KCNJ5	chr11:128786525-128786525	0.002
LAMA4	chr6:112430719-112430719	0.004
LAMA4	chr6:112430720-112430720	0.004
LAMA4	chr6:112438956-112438956	0.002
LAMA4	chr6:112439053-112439053	0.004
LAMA4	chr6:112441506-112441506	0.002
LAMA4	chr6:112453998-112453998	0.008
LAMA4	chr6:112460365-112460365	0.025
LAMA4	chr6:112460461-112460461	0.004
LAMA4	chr6:112476135-112476135	0.002
LAMA4	chr6:112496640-112496640	0.004
LAMA4	chr6:112496641-112496641	0.004
LAMA4	chr6:112506487-112506487	0.002
LAMA4	chr6:112508770-112508770	0.004
LDB3	chr10:88428502-88428502	0.002
LDB3	chr10:88439193-88439193	0.002
LDB3	chr10:88441539-88441539	0.002
LDB3	chr10:88451715-88451715	0.002
LDB3	chr10:88466320-88466320	0.002
LDB3	chr10:88469801-88469801	0.002
LDB3	chr10:88476354-88476354	0.002

LDB3	chr10:88476387-88476387	0.002
LDB3	chr10:88485932-88485932	0.002
MYBPC3	chr11:47354367-47354367	0.002
MYBPC3	chr11:47357481-47357481	0.002
MYBPC3	chr11:47357511-47357511	0.002
MYBPC3	chr11:47359273-47359273	0.002
MYBPC3	chr11:47367848-47367848	0.002
MYH6	chr14:23851683-23851683	0.002
MYH6	chr14:23852473-23852473	0.004
MYH6	chr14:23853925-23853925	0.002
MYH6	chr14:23857474-23857474	0.002
MYH6	chr14:23858697-23858697	0.011
MYH6	chr14:23862646-23862646	0.004
MYH6	chr14:23863351-23863351	0.002
MYH6	chr14:23865523-23865523	0.002
MYH6	chr14:23865624-23865624	0.004
MYH6	chr14:23868072-23868072	0.004
MYH6	chr14:23874459-23874459	0.004
MYH6	chr14:23876339-23876339	0.011
MYH7	chr14:23886375-23886375	0.036
MYH7	chr14:23886409-23886409	0.019
MYH7	chr14:23886470-23886470	0.002
MYH7	chr14:23890168-23890168	0.002
MYH7	chr14:23890169-23890169	0.005
MYH7	chr14:23894192-23894192	0.002
MYH7	chr14:23894196-23894196	0.011
MYH7	chr14:23896053-23896053	0.004
MYH7	chr14:23896932-23896932	0.017
MYH7	chr14:23900998-23900998	0.002
MYH7	chr14:23902385-23902385	0.002
MYH7	chr14:23902802-23902802	0.002
MYH7	chr14:23902818-23902818	0.004
MYL2	chr12:111348976-111348976	0.004
MYL2	chr12:111356937-111356937	0.006
MYL2	chr12:111356938-111356938	0.013
MYL2	chr12:111356947-111356947	0.006
MYLK2	chr20:30407962-30407962	0.002
MYLK2	chr20:30408306-30408306	0.016
MYLK2	chr20:30414706-30414706	0.004
MYLK2	chr20:30419886-30419886	0.002

MYLK2	chr20:30419932-30419932	0.002
MYOM1	chr18:3071878-3071878	0.002
MYOM1	chr18:3135569-3135569	0.002
MYOM1	chr18:3135644-3135644	0.002
MYOM1	chr18:3135692-3135692	0.004
MYOM1	chr18:3141984-3141984	0.002
MYOM1	chr18:3151736-3151736	0.002
MYOM1	chr18:3151799-3151799	0.002
MYOM1	chr18:3154949-3154949	0.004
MYOM1	chr18:3155022-3155022	0.002
MYOM1	chr18:3174118-3174118	0.002
MYOM1	chr18:3188778-3188778	0.004
MYOM1	chr18:3188873-3188873	0.048
MYOM1	chr18:3193928-3193928	0.011
MYOM1	chr18:3193931-3193931	0.026
MYOM1	chr18:3193932-3193932	0.026
MYOM1	chr18:3215126-3215126	0.004
MYPN	chr10:69905278-69905278	0.002
MYPN	chr10:69905283-69905283	0.002
MYPN	chr10:69934259-69934259	0.040
MYPN	chr10:69957185-69957185	0.002
NDUFAF1	chr15:41679701-41679701	0.002
NDUFAF1	chr15:41688732-41688732	0.031
NDUFAF1	chr15:41688890-41688890	0.002
NEBL	chr10:21097546-21097546	0.006
NEBL	chr10:21101809-21101809	0.002
NEBL	chr10:21157621-21157621	0.002
NEBL	chr10:21178852-21178852	0.004
NEBL	chr10:21309081-21309081	0.002
NEBL	chr10:21358886-21358886	0.005
NEBL	chr10:21461352-21461352	0.008
NEBL	chr10:21461353-21461353	0.008
NEBL	chr10:21461354-21461354	0.008
NEXN	chr1:78383296-78383296	0.002
NEXN	chr1:78383297-78383297	0.002
NEXN	chr1:78401514-78401514	0.006
NEXN	chr1:78401632-78401632	0.002
NEXN	chr1:78401664-78401664	0.002
NEXN	chr1:78408491-78408491	0.002
PDLIM3	chr4:186427735-186427735	0.023

PDLIM3	chr4:186435461-186435461	0.004
PDLIM3	chr4:186435476-186435476	0.011
PDLIM3	chr4:186435906-186435906	0.002
PDLIM3	chr4:186435975-186435975	0.002
PKP2	chr12:32945612-32945612	0.002
PKP2	chr12:32955415-32955415	0.034
PKP2	chr12:32975454-32975454	0.002
PKP2	chr12:32977026-32977026	0.006
PKP2	chr12:33031395-33031395	0.002
PRKAG2	chr7:151262467-151262467	0.002
PRKAG2	chr7:151478406-151478406	0.002
PRKAG2	chr7:151573675-151573675	0.004
PSEN2	chr1:227068398-227068398	0.002
PSEN2	chr1:227071475-227071475	0.007
PSEN2	chr1:227071518-227071518	0.002
PSEN2	chr1:227073271-227073271	0.004
PSEN2	chr1:227077798-227077798	0.002
PSEN2	chr1:227079504-227079504	0.002
RAF1	chr3:12627230-12627230	0.002
RAF1	chr3:12632305-12632305	0.002
RAF1	chr3:12653509-12653509	0.006
RAF1	chr3:12660096-12660096	0.004
RAF1	chr3:12660099-12660099	0.002
RBM20	chr10:112540785-112540785	0.004
RBM20	chr10:112540897-112540897	0.002
RBM20	chr10:112541062-112541062	0.003
RBM20	chr10:112541271-112541271	0.002
RBM20	chr10:112541481-112541481	0.002
RBM20	chr10:112543134-112543134	0.006
RBM20	chr10:112544134-112544134	0.003
RBM20	chr10:112557338-112557338	0.009
RBM20	chr10:112583294-112583294	0.004
RBM20	chr10:112595695-112595695	0.003
RYR2	chr1:237659888-237659888	0.002
RYR2	chr1:237659966-237659966	0.002
RYR2	chr1:237664127-237664127	0.002
RYR2	chr1:237674997-237674997	0.002
RYR2	chr1:237711862-237711862	0.002
RYR2	chr1:237753175-237753175	0.002
RYR2	chr1:237754138-237754138	0.002

RYR2	chr1:237777706-237777706	0.002
RYR2	chr1:237811912-237811912	0.002
RYR2	chr1:237823367-237823367	0.002
RYR2	chr1:237824227-237824227	0.002
RYR2	chr1:237863717-237863717	0.004
RYR2	chr1:237870285-237870285	0.009
RYR2	chr1:237872268-237872268	0.004
RYR2	chr1:237872280-237872280	0.002
RYR2	chr1:237880510-237880510	0.004
RYR2	chr1:237919639-237919639	0.002
RYR2	chr1:237919642-237919642	0.006
RYR2	chr1:237951416-237951416	0.004
SCN5A	chr3:38592174-38592174	0.004
SCN5A	chr3:38592795-38592795	0.006
SCN5A	chr3:38603922-38603922	0.002
SCN5A	chr3:38607917-38607917	0.004
SCN5A	chr3:38622478-38622478	0.002
SCN5A	chr3:38628974-38628974	0.006
SCN5A	chr3:38663951-38663951	0.002
SOS1	chr2:39213083-39213083	0.008
SOS1	chr2:39250011-39250011	0.006
SOS1	chr2:39250012-39250012	0.006
SOS1	chr2:39250104-39250104	0.006
SOS1	chr2:39281837-39281837	0.006
SOS1	chr2:39281948-39281948	0.002
SOS1	chr2:39283929-39283929	0.004
SOS1	chr2:39285912-39285912	0.013
SOS1	chr2:39285913-39285913	0.013
SYNE1	chr6:152454556-152454556	0.004
SYNE1	chr6:152456292-152456292	0.002
SYNE1	chr6:152464870-152464870	0.013
SYNE1	chr6:152469204-152469204	0.006
SYNE1	chr6:152469354-152469354	0.002
SYNE1	chr6:152469432-152469432	0.002
SYNE1	chr6:152470620-152470620	0.002
SYNE1	chr6:152470698-152470698	0.002
SYNE1	chr6:152472711-152472711	0.006
SYNE1	chr6:152473212-152473212	0.002
SYNE1	chr6:152473266-152473266	0.004
SYNE1	chr6:152501317-152501317	0.002

SYNE1	chr6:152501409-152501409	0.013
SYNE1	chr6:152501416-152501416	0.008
SYNE1	chr6:152529141-152529141	0.004
SYNE1	chr6:152532702-152532702	0.021
SYNE1	chr6:152534850-152534850	0.002
SYNE1	chr6:152536156-152536156	0.002
SYNE1	chr6:152542681-152542681	0.002
SYNE1	chr6:152551777-152551777	0.002
SYNE1	chr6:152570346-152570346	0.002
SYNE1	chr6:152583258-152583258	0.002
SYNE1	chr6:152605279-152605279	0.002
SYNE1	chr6:152631566-152631566	0.004
SYNE1	chr6:152631852-152631852	0.002
SYNE1	chr6:152642918-152642918	0.002
SYNE1	chr6:152647746-152647746	0.002
SYNE1	chr6:152650973-152650973	0.002
SYNE1	chr6:152651024-152651024	0.002
SYNE1	chr6:152651196-152651196	0.002
SYNE1	chr6:152651557-152651557	0.002
SYNE1	chr6:152652505-152652505	0.002
SYNE1	chr6:152658053-152658053	0.011
SYNE1	chr6:152658056-152658056	0.021
SYNE1	chr6:152658062-152658062	0.024
SYNE1	chr6:152660405-152660405	0.002
SYNE1	chr6:152665298-152665298	0.011
SYNE1	chr6:152665299-152665299	0.002
SYNE1	chr6:152668211-152668211	0.015
SYNE1	chr6:152674445-152674445	0.002
SYNE1	chr6:152679518-152679518	0.011
SYNE1	chr6:152683324-152683324	0.002
SYNE1	chr6:152686067-152686067	0.004
SYNE1	chr6:152688257-152688257	0.004
SYNE1	chr6:152688276-152688276	0.002
SYNE1	chr6:152690150-152690150	0.002
SYNE1	chr6:152690199-152690199	0.002
SYNE1	chr6:152697692-152697692	0.002
SYNE1	chr6:152708334-152708334	0.002
SYNE1	chr6:152712505-152712505	0.002
SYNE1	chr6:152718102-152718102	0.002
SYNE1	chr6:152722413-152722413	0.002

SYNE1	chr6:152728282-152728282	0.004
SYNE1	chr6:152730273-152730273	0.004
SYNE1	chr6:152757224-152757224	0.002
SYNE1	chr6:152768726-152768726	0.002
SYNE1	chr6:152770696-152770696	0.002
SYNE1	chr6:152771884-152771884	0.002
SYNE1	chr6:152771956-152771956	0.004
SYNE1	chr6:152774626-152774626	0.004
SYNE1	chr6:152776571-152776571	0.002
SYNE1	chr6:152776583-152776583	0.002
SYNE1	chr6:152779933-152779933	0.030
SYNE1	chr6:152784601-152784601	0.002
SYNE1	chr6:152784621-152784621	0.007
SYNE1	chr6:152786497-152786497	0.004
SYNE1	chr6:152786498-152786498	0.004
SYNE1	chr6:152804227-152804227	0.002
SYNE1	chr6:152806014-152806014	0.002
SYNE1	chr6:152823872-152823872	0.006
SYNE1	chr6:152826378-152826378	0.002
SYNE1	chr6:152861132-152861132	0.002
TMEM43	chr3:14170981-14170981	0.002
TMEM43	chr3:14180731-14180731	0.004
TMEM43	chr3:14183204-14183204	0.002
TMPO	chr12:98921742-98921742	0.002
TMPO	chr12:98926805-98926805	0.002
TMPO	chr12:98927272-98927272	0.006
TMPO	chr12:98927312-98927312	0.008
TMPO	chr12:98927695-98927695	0.002
TMPO	chr12:98938295-98938295	0.011
TMPO	chr12:98941467-98941467	0.002
TTN	chr2:179393111-179393111	0.039
TTN	chr2:179393691-179393691	0.004
TTN	chr2:179393859-179393859	0.022
TTN	chr2:179395508-179395508	0.002
TTN	chr2:179395554-179395554	0.009
TTN	chr2:179395555-179395555	0.009
TTN	chr2:179395560-179395560	0.037
TTN	chr2:179395573-179395573	0.011
TTN	chr2:179396162-179396162	0.030
TTN	chr2:179396320-179396320	0.004

TTN	chr2:179396329-179396329	0.024
TTN	chr2:179396446-179396446	0.004
TTN	chr2:179396682-179396682	0.002
TTN	chr2:179396766-179396766	0.019
TTN	chr2:179396782-179396782	0.004
TTN	chr2:179396928-179396928	0.002
TTN	chr2:179396961-179396961	0.002
TTN	chr2:179396965-179396965	0.002
TTN	chr2:179397342-179397342	0.002
TTN	chr2:179398465-179398465	0.002
TTN	chr2:179398509-179398509	0.037
TTN	chr2:179398914-179398914	0.002
TTN	chr2:179398944-179398944	0.002
TTN	chr2:179399539-179399539	0.002
TTN	chr2:179399576-179399576	0.017
TTN	chr2:179399677-179399677	0.006
TTN	chr2:179399748-179399748	0.024
TTN	chr2:179399755-179399755	0.011
TTN	chr2:179400130-179400130	0.015
TTN	chr2:179400433-179400433	0.004
TTN	chr2:179400541-179400541	0.002
TTN	chr2:179400895-179400895	0.030
TTN	chr2:179401740-179401740	0.028
TTN	chr2:179401742-179401742	0.013
TTN	chr2:179403702-179403702	0.004
TTN	chr2:179403703-179403703	0.004
TTN	chr2:179403750-179403750	0.026
TTN	chr2:179404393-179404393	0.002
TTN	chr2:179404402-179404402	0.004
TTN	chr2:179404498-179404498	0.006
TTN	chr2:179404501-179404501	0.004
TTN	chr2:179404550-179404550	0.002
TTN	chr2:179404628-179404628	0.030
TTN	chr2:179404661-179404661	0.002
TTN	chr2:179404872-179404872	0.002
TTN	chr2:179406044-179406044	0.006
TTN	chr2:179406266-179406266	0.006
TTN	chr2:179407159-179407159	0.002
TTN	chr2:179407236-179407236	0.011
TTN	chr2:179407482-179407482	0.002

TTN	chr2:179408585-179408585	0.002
TTN	chr2:179408713-179408713	0.028
TTN	chr2:179409096-179409096	0.006
TTN	chr2:179410184-179410184	0.002
TTN	chr2:179410282-179410282	0.022
TTN	chr2:179410298-179410298	0.004
TTN	chr2:179410397-179410397	0.002
TTN	chr2:179410607-179410607	0.004
TTN	chr2:179410954-179410954	0.002
TTN	chr2:179411011-179411011	0.028
TTN	chr2:179412307-179412307	0.002
TTN	chr2:179412452-179412452	0.002
TTN	chr2:179413223-179413223	0.002
TTN	chr2:179413483-179413483	0.002
TTN	chr2:179413571-179413571	0.002
TTN	chr2:179413601-179413601	0.002
TTN	chr2:179414062-179414062	0.002
TTN	chr2:179414162-179414162	0.004
TTN	chr2:179414177-179414177	0.002
TTN	chr2:179414318-179414318	0.028
TTN	chr2:179414512-179414512	0.002
TTN	chr2:179414944-179414944	0.002
TTN	chr2:179414992-179414992	0.011
TTN	chr2:179415710-179415710	0.004
TTN	chr2:179415776-179415776	0.004
TTN	chr2:179415859-179415859	0.002
TTN	chr2:179416659-179416659	0.002
TTN	chr2:179416924-179416924	0.006
TTN	chr2:179416925-179416925	0.006
TTN	chr2:179417104-179417104	0.002
TTN	chr2:179417919-179417919	0.002
TTN	chr2:179417925-179417925	0.002
TTN	chr2:179418507-179418507	0.009
TTN	chr2:179419203-179419203	0.004
TTN	chr2:179419338-179419338	0.002
TTN	chr2:179419353-179419353	0.002
TTN	chr2:179419464-179419464	0.002
TTN	chr2:179419792-179419792	0.002
TTN	chr2:179422470-179422470	0.004
TTN	chr2:179422669-179422669	0.002

TTN	chr2:179422882-179422882	0.004
TTN	chr2:179424514-179424514	0.002
TTN	chr2:179425021-179425021	0.002
TTN	chr2:179425136-179425136	0.002
TTN	chr2:179425213-179425213	0.002
TTN	chr2:179425387-179425387	0.002
TTN	chr2:179425819-179425819	0.006
TTN	chr2:179426711-179426711	0.002
TTN	chr2:179427560-179427560	0.002
TTN	chr2:179427754-179427754	0.002
TTN	chr2:179427796-179427796	0.002
TTN	chr2:179428061-179428061	0.002
TTN	chr2:179428168-179428168	0.002
TTN	chr2:179428595-179428595	0.002
TTN	chr2:179428678-179428678	0.002
TTN	chr2:179428769-179428769	0.002
TTN	chr2:179429387-179429387	0.002
TTN	chr2:179430124-179430124	0.004
TTN	chr2:179430142-179430142	0.002
TTN	chr2:179430224-179430224	0.002
TTN	chr2:179430305-179430305	0.002
TTN	chr2:179430521-179430521	0.002
TTN	chr2:179430815-179430815	0.004
TTN	chr2:179430960-179430960	0.002
TTN	chr2:179431076-179431076	0.019
TTN	chr2:179431207-179431207	0.006
TTN	chr2:179431214-179431214	0.004
TTN	chr2:179431444-179431444	0.002
TTN	chr2:179431544-179431544	0.002
TTN	chr2:179431594-179431594	0.028
TTN	chr2:179431879-179431879	0.002
TTN	chr2:179433046-179433046	0.002
TTN	chr2:179433221-179433221	0.028
TTN	chr2:179433317-179433317	0.002
TTN	chr2:179433580-179433580	0.028
TTN	chr2:179434120-179434120	0.002
TTN	chr2:179434139-179434139	0.028
TTN	chr2:179434160-179434160	0.002
TTN	chr2:179434508-179434508	0.004
TTN	chr2:179434516-179434516	0.037

TTN	chr2:179435027-179435027	0.002
TTN	chr2:179435250-179435250	0.004
TTN	chr2:179435887-179435887	0.002
TTN	chr2:179435964-179435964	0.002
TTN	chr2:179436320-179436320	0.002
TTN	chr2:179437034-179437034	0.004
TTN	chr2:179437255-179437255	0.002
TTN	chr2:179437928-179437928	0.006
TTN	chr2:179438390-179438390	0.002
TTN	chr2:179438528-179438528	0.002
TTN	chr2:179438560-179438560	0.002
TTN	chr2:179438866-179438866	0.026
TTN	chr2:179439154-179439154	0.002
TTN	chr2:179439491-179439491	0.002
TTN	chr2:179439710-179439710	0.011
TTN	chr2:179439823-179439823	0.006
TTN	chr2:179439877-179439877	0.022
TTN	chr2:179439884-179439884	0.013
TTN	chr2:179439974-179439974	0.002
TTN	chr2:179440163-179440163	0.006
TTN	chr2:179440885-179440885	0.002
TTN	chr2:179441119-179441119	0.002
TTN	chr2:179441295-179441295	0.004
TTN	chr2:179441513-179441513	0.004
TTN	chr2:179441917-179441917	0.002
TTN	chr2:179441932-179441932	0.011
TTN	chr2:179442198-179442198	0.002
TTN	chr2:179442825-179442825	0.002
TTN	chr2:179444768-179444768	0.002
TTN	chr2:179444915-179444915	0.002
TTN	chr2:179445129-179445129	0.002
TTN	chr2:179445189-179445189	0.002
TTN	chr2:179446381-179446381	0.015
TTN	chr2:179446764-179446764	0.002
TTN	chr2:179447787-179447787	0.002
TTN	chr2:179447927-179447927	0.002
TTN	chr2:179448393-179448393	0.002
TTN	chr2:179448493-179448493	0.002
TTN	chr2:179449131-179449131	0.030
TTN	chr2:179449186-179449186	0.006

TTN	chr2:179452459-179452459	0.006
TTN	chr2:179452902-179452902	0.004
TTN	chr2:179452911-179452911	0.002
TTN	chr2:179453355-179453355	0.002
TTN	chr2:179453429-179453429	0.011
TTN	chr2:179453918-179453918	0.002
TTN	chr2:179454020-179454020	0.004
TTN	chr2:179454299-179454299	0.002
TTN	chr2:179454392-179454392	0.002
TTN	chr2:179454393-179454393	0.002
TTN	chr2:179454530-179454530	0.006
TTN	chr2:179455086-179455086	0.007
TTN	chr2:179455128-179455128	0.002
TTN	chr2:179455352-179455352	0.004
TTN	chr2:179456911-179456911	0.006
TTN	chr2:179456929-179456929	0.002
TTN	chr2:179457370-179457370	0.002
TTN	chr2:179458020-179458020	0.006
TTN	chr2:179458396-179458396	0.002
TTN	chr2:179458426-179458426	0.002
TTN	chr2:179460495-179460495	0.004
TTN	chr2:179460521-179460521	0.002
TTN	chr2:179462345-179462345	0.002
TTN	chr2:179462367-179462367	0.002
TTN	chr2:179462477-179462477	0.002
TTN	chr2:179463490-179463490	0.002
TTN	chr2:179463987-179463987	0.002
TTN	chr2:179464030-179464030	0.030
TTN	chr2:179464119-179464119	0.007
TTN	chr2:179464572-179464572	0.002
TTN	chr2:179465827-179465827	0.002
TTN	chr2:179468704-179468704	0.002
TTN	chr2:179468729-179468729	0.002
TTN	chr2:179468804-179468804	0.002
TTN	chr2:179468833-179468833	0.002
TTN	chr2:179469468-179469468	0.002
TTN	chr2:179469738-179469738	0.002
TTN	chr2:179469873-179469873	0.004
TTN	chr2:179471766-179471766	0.004
TTN	chr2:179472223-179472223	0.004

TTN	chr2:179472292-179472292	0.030
TTN	chr2:179472293-179472293	0.002
TTN	chr2:179472319-179472319	0.002
TTN	chr2:179472610-179472610	0.002
TTN	chr2:179472662-179472662	0.002
TTN	chr2:179474216-179474216	0.002
TTN	chr2:179474668-179474668	0.015
TTN	chr2:179476557-179476557	0.002
TTN	chr2:179476584-179476584	0.002
TTN	chr2:179476841-179476841	0.002
TTN	chr2:179477267-179477267	0.011
TTN	chr2:179477678-179477678	0.002
TTN	chr2:179479288-179479288	0.002
TTN	chr2:179481337-179481337	0.011
TTN	chr2:179481488-179481488	0.011
TTN	chr2:179481489-179481489	0.009
TTN	chr2:179481885-179481885	0.004
TTN	chr2:179481892-179481892	0.026
TTN	chr2:179481894-179481894	0.009
TTN	chr2:179482089-179482089	0.002
TTN	chr2:179482937-179482937	0.009
TTN	chr2:179482994-179482994	0.006
TTN	chr2:179483108-179483108	0.002
TTN	chr2:179483349-179483349	0.002
TTN	chr2:179484965-179484965	0.002
TTN	chr2:179485125-179485125	0.002
TTN	chr2:179485946-179485946	0.002
TTN	chr2:179486037-179486037	0.004
TTN	chr2:179486223-179486223	0.004
TTN	chr2:179486283-179486283	0.002
TTN	chr2:179486431-179486431	0.002
TTN	chr2:179495039-179495039	0.002
TTN	chr2:179497316-179497316	0.002
TTN	chr2:179498789-179498789	0.006
TTN	chr2:179499530-179499530	0.013
TTN	chr2:179499914-179499914	0.002
TTN	chr2:179501209-179501209	0.015
TTN	chr2:179501444-179501444	0.002
TTN	chr2:179515483-179515483	0.006
TTN	chr2:179517220-179517220	0.002

TTN	chr2:179528603-179528603	0.002
TTN	chr2:179537200-179537200	0.002
TTN	chr2:179547465-179547465	0.006
TTN	chr2:179547573-179547573	0.002
TTN	chr2:179549131-179549131	0.024
TTN	chr2:179549404-179549404	0.005
TTN	chr2:179549407-179549407	0.017
TTN	chr2:179553849-179553849	0.002
TTN	chr2:179554549-179554549	0.002
TTN	chr2:179556748-179556748	0.002
TTN	chr2:179557242-179557242	0.002
TTN	chr2:179558686-179558686	0.009
TTN	chr2:179559353-179559353	0.002
TTN	chr2:179560789-179560789	0.002
TTN	chr2:179560847-179560847	0.002
TTN	chr2:179560942-179560942	0.002
TTN	chr2:179563606-179563606	0.002
TTN	chr2:179567340-179567340	0.002
TTN	chr2:179569098-179569098	0.002
TTN	chr2:179569387-179569387	0.028
TTN	chr2:179571448-179571448	0.024
TTN	chr2:179571595-179571595	0.002
TTN	chr2:179571618-179571618	0.004
TTN	chr2:179572274-179572274	0.011
TTN	chr2:179572519-179572519	0.002
TTN	chr2:179575869-179575869	0.002
TTN	chr2:179576757-179576757	0.002
TTN	chr2:179578704-179578704	0.024
TTN	chr2:179578891-179578891	0.002
TTN	chr2:179579172-179579172	0.002
TTN	chr2:179579273-179579273	0.011
TTN	chr2:179579977-179579977	0.022
TTN	chr2:179579989-179579989	0.002
TTN	chr2:179579990-179579990	0.002
TTN	chr2:179581835-179581835	0.032
TTN	chr2:179582760-179582760	0.004
TTN	chr2:179582853-179582853	0.037
TTN	chr2:179584430-179584430	0.006
TTN	chr2:179585257-179585257	0.004
TTN	chr2:179585312-179585312	0.004

TTN	chr2:179586604-179586604	0.028
TTN	chr2:179587094-179587094	0.002
TTN	chr2:179587097-179587097	0.004
TTN	chr2:179587955-179587955	0.002
TTN	chr2:179588279-179588279	0.002
TTN	chr2:179589058-179589058	0.004
TTN	chr2:179589241-179589241	0.015
TTN	chr2:179590329-179590329	0.022
TTN	chr2:179590714-179590714	0.002
TTN	chr2:179592096-179592096	0.004
TTN	chr2:179593352-179593352	0.019
TTN	chr2:179593503-179593503	0.002
TTN	chr2:179593664-179593664	0.004
TTN	chr2:179595241-179595241	0.004
TTN	chr2:179595277-179595277	0.009
TTN	chr2:179595372-179595372	0.002
TTN	chr2:179596049-179596049	0.004
TTN	chr2:179596191-179596191	0.004
TTN	chr2:179596554-179596554	0.006
TTN	chr2:179596838-179596838	0.002
TTN	chr2:179596920-179596920	0.002
TTN	chr2:179597259-179597259	0.002
TTN	chr2:179597276-179597276	0.004
TTN	chr2:179597443-179597443	0.004
TTN	chr2:179597444-179597444	0.004
TTN	chr2:179597593-179597593	0.004
TTN	chr2:179597600-179597600	0.009
TTN	chr2:179598004-179598004	0.002
TTN	chr2:179598041-179598041	0.002
TTN	chr2:179598553-179598553	0.004
TTN	chr2:179599069-179599069	0.004
TTN	chr2:179599074-179599074	0.002
TTN	chr2:179599088-179599088	0.006
TTN	chr2:179599144-179599144	0.004
TTN	chr2:179599473-179599473	0.002
TTN	chr2:179599491-179599491	0.002
TTN	chr2:179599667-179599667	0.009
TTN	chr2:179600300-179600300	0.002
TTN	chr2:179600303-179600303	0.002
TTN	chr2:179600640-179600640	0.002

TTN	chr2:179600648-179600648	0.048
TTN	chr2:179603970-179603970	0.002
TTN	chr2:179603994-179603994	0.011
TTN	chr2:179604042-179604042	0.004
TTN	chr2:179604160-179604160	0.045
TTN	chr2:179604221-179604221	0.004
TTN	chr2:179604222-179604222	0.004
TTN	chr2:179604366-179604366	0.045
TTN	chr2:179605073-179605073	0.002
TTN	chr2:179605380-179605380	0.009
TTN	chr2:179605658-179605658	0.002
TTN	chr2:179605706-179605706	0.015
TTN	chr2:179605707-179605707	0.015
TTN	chr2:179605725-179605725	0.030
TTN	chr2:179605991-179605991	0.009
TTN	chr2:179606288-179606288	0.002
TTN	chr2:179610320-179610320	0.005
TTN	chr2:179610967-179610967	0.006
TTN	chr2:179611711-179611711	0.044
TTN	chr2:179613191-179613191	0.004
TTN	chr2:179614454-179614454	0.002
TTN	chr2:179615278-179615278	0.002
TTN	chr2:179615654-179615654	0.002
TTN	chr2:179615931-179615931	0.004
TTN	chr2:179616446-179616446	0.004
TTN	chr2:179616458-179616458	0.009
TTN	chr2:179621309-179621309	0.002
TTN	chr2:179621353-179621353	0.002
TTN	chr2:179621503-179621503	0.011
TTN	chr2:179622464-179622464	0.006
TTN	chr2:179622616-179622616	0.002
TTN	chr2:179623887-179623887	0.004
TTN	chr2:179628918-179628918	0.008
TTN	chr2:179628946-179628946	0.002
TTN	chr2:179629367-179629367	0.002
TTN	chr2:179631128-179631128	0.002
TTN	chr2:179631233-179631233	0.006
TTN	chr2:179632496-179632496	0.019
TTN	chr2:179632598-179632598	0.002
TTN	chr2:179633472-179633472	0.006

TTN	chr2:179634936-179634936	0.044
TTN	chr2:179634961-179634961	0.025
TTN	chr2:179635985-179635985	0.002
TTN	chr2:179637861-179637861	0.041
TTN	chr2:179638651-179638651	0.002
TTN	chr2:179638721-179638721	0.017
TTN	chr2:179638834-179638834	0.002
TTN	chr2:179639725-179639725	0.002
TTN	chr2:179640399-179640399	0.002
TTN	chr2:179640923-179640923	0.002
TTN	chr2:179640946-179640946	0.002
TTN	chr2:179641283-179641283	0.002
TTN	chr2:179641901-179641901	0.002
TTN	chr2:179641975-179641975	0.044
TTN	chr2:179642153-179642153	0.002
TTN	chr2:179642440-179642440	0.002
TTN	chr2:179642441-179642441	0.002
TTN	chr2:179642589-179642589	0.015
TTN	chr2:179643656-179643656	0.002
TTN	chr2:179643775-179643775	0.006
TTN	chr2:179644012-179644012	0.009
TTN	chr2:179644035-179644035	0.031
TTN	chr2:179645962-179645962	0.004
TTN	chr2:179647782-179647782	0.019
TTN	chr2:179647783-179647783	0.017
TTN	chr2:179649074-179649074	0.004
TTN	chr2:179649075-179649075	0.004
TTN	chr2:179654142-179654142	0.002
TTN	chr2:179658175-179658175	0.028
TTN	chr2:179659236-179659236	0.002
TTN	chr2:179659795-179659795	0.002
TTN	chr2:179659806-179659806	0.002
TTN	chr2:179659911-179659911	0.002
TTN	chr2:179664292-179664292	0.004
TTN	chr2:179665163-179665163	0.002
TTN	chr2:179666894-179666894	0.004
TTN	chr2:179666979-179666979	0.002
TTN	chr2:179666982-179666982	0.015
TTN	chr2:179666988-179666988	0.006
TTN	chr2:179669295-179669295	0.004

TTR	chr18:29178549-29178549	0.002
TTR	chr18:29178562-29178562	0.004
TTR	chr18:29178610-29178610	0.002
TXNRD2	chr22:19882691-19882691	0.002
TXNRD2	chr22:19883021-19883021	0.002
TXNRD2	chr22:19918615-19918615	0.004
VCL	chr10:75849898-75849898	0.002
VCL	chr10:75855491-75855491	0.004
VCL	chr10:75860740-75860740	0.002
VCL	chr10:75864961-75864961	0.002
VCL	chr10:75865066-75865066	0.002
Doxorubicin pathway data set		
ABCB1	chr7:87138760-87138760	0.002
ABCB1	chr7:87145953-87145953	0.015
ABCB1	chr7:87168609-87168609	0.002
ABCB1	chr7:87175288-87175288	0.002
ABCB1	chr7:87179256-87179256	0.002
ABCB1	chr7:87179813-87179813	0.025
ABCB1	chr7:87196161-87196161	0.011
ABCB1	chr7:87196164-87196164	0.008
ABCB1	chr7:87214975-87214975	0.006
ABCB5	chr7:20682884-20682884	0.026
ABCB5	chr7:20683170-20683170	0.011
ABCB5	chr7:20685484-20685484	0.011
ABCB5	chr7:20685488-20685488	0.006
ABCB5	chr7:20685493-20685493	0.002
ABCB5	chr7:20691167-20691167	0.004
ABCB5	chr7:20691184-20691184	0.002
ABCB5	chr7:20691185-20691185	0.002
ABCB5	chr7:20698170-20698170	0.002
ABCB5	chr7:20721260-20721260	0.002
ABCB5	chr7:20739510-20739510	0.004
ABCB5	chr7:20739703-20739703	0.002
ABCB5	chr7:20767954-20767954	0.019
ABCB5	chr7:20768024-20768024	0.004
ABCB5	chr7:20782528-20782528	0.002
ABCB5	chr7:20784910-20784910	0.009
ABCC1	chr16:16142079-16142079	0.011
ABCC1	chr16:16162142-16162142	0.004
ABCC1	chr16:16165569-16165569	0.006

ABCC1	chr16:16208889-16208889	0.006
ABCC2	chr10:101557063-101557063	0.006
ABCC2	chr10:101559093-101559093	0.004
ABCC2	chr10:101564013-101564013	0.004
ABCC2	chr10:101577182-101577182	0.002
ABCC2	chr10:101578952-101578952	0.011
ABCC2	chr10:101591737-101591737	0.015
ABCC2	chr10:101591866-101591866	0.002
ABCC2	chr10:101596043-101596043	0.002
ABCC2	chr10:101604107-101604107	0.002
ABCC3	chr17:48742541-48742541	0.002
ABCC3	chr17:48744940-48744940	0.004
ABCC3	chr17:48744966-48744966	0.002
ABCC3	chr17:48746251-48746251	0.012
ABCC3	chr17:48746739-48746739	0.002
ABCC3	chr17:48753045-48753045	0.003
ABCC3	chr17:48755145-48755145	0.002
ABCC3	chr17:48755178-48755178	0.006
ABCC3	chr17:48757208-48757208	0.006
ABCC3	chr17:48761062-48761062	0.006
ABCC3	chr17:48762190-48762190	0.002
ABCC3	chr17:48762223-48762223	0.002
ABCC3	chr17:48765024-48765024	0.002
ABCC3	chr17:48765063-48765063	0.002
ABCC4	chr13:95696575-95696575	0.004
ABCC4	chr13:95705380-95705380	0.013
ABCC4	chr13:95705424-95705424	0.006
ABCC4	chr13:95735520-95735520	0.006
ABCC4	chr13:95818570-95818570	0.002
ABCC4	chr13:95829966-95829966	0.006
ABCC4	chr13:95830300-95830300	0.002
ABCC4	chr13:95860061-95860061	0.002
ABCC4	chr13:95863008-95863008	0.044
ABCC5	chr3:183669271-183669271	0.006
ABCC5	chr3:183677620-183677620	0.017
ABCC5	chr3:183677621-183677621	0.017
ABCC5	chr3:183679412-183679412	0.009
ABCC5	chr3:183681203-183681203	0.004
ABCC5	chr3:183681255-183681255	0.017
ABCC5	chr3:183689395-183689395	0.002

ABCC5	chr3:183695316-183695316	0.002
ABCC5	chr3:183700648-183700648	0.002
ABCG2	chr4:89016692-89016692	0.002
ABCG2	chr4:89016693-89016693	0.002
ABCG2	chr4:89016694-89016694	0.002
ABCG2	chr4:89018670-89018670	0.002
ABCG2	chr4:89052317-89052317	0.002
ABCG2	chr4:89052998-89052998	0.008
ABCG2	chr4:89060981-89060981	0.002
ACO1	chr9:32405589-32405589	0.002
ACO1	chr9:32408612-32408612	0.006
ACO1	chr9:32418473-32418473	0.002
ACO1	chr9:32427386-32427386	0.002
ACO1	chr9:32430509-32430509	0.002
ACO1	chr9:32430510-32430510	0.002
ACO1	chr9:32433782-32433782	0.008
ACO1	chr9:32434610-32434610	0.002
ACO1	chr9:32436107-32436107	0.019
AKR1A1	chr1:46032311-46032311	0.042
AKR1A1	chr1:46035584-46035584	0.004
AKR1A1	chr1:46035585-46035585	0.006
AKR1C3	chr10:5139642-5139642	0.002
AKR1C3	chr10:5141583-5141583	0.002
AKR1C3	chr10:5141609-5141609	0.013
CAT	chr11:34470833-34470833	0.002
CAT	chr11:34477621-34477621	0.002
CAT	chr11:34477698-34477698	0.002
CBR1	chr21:37443349-37443349	0.002
CBR1	chr21:37444697-37444697	0.034
CBR1	chr21:37444971-37444971	0.002
CBR1	chr21:37445131-37445131	0.002
CBR3	chr21:37507719-37507719	0.002
CBR3	chr21:37510224-37510224	0.004
CBR3	chr21:37518728-37518728	0.004
ERBB2	chr17:37866395-37866395	0.004
ERBB2	chr17:37872115-37872115	0.002
ERBB2	chr17:37872806-37872806	0.002
ERBB2	chr17:37873672-37873672	0.048
ERBB2	chr17:37882019-37882019	0.002
ERBB2	chr17:37884176-37884176	0.002

ERBB2	chr17:37884217-37884217	0.002
ERBB2	chr17:37884285-37884285	0.002
ERCC2	chr19:45855769-45855769	0.002
ERCC2	chr19:45860581-45860581	0.002
ERCC2	chr19:45860928-45860928	0.002
ERCC2	chr19:45868386-45868386	0.002
HAS3	chr16:69143315-69143315	0.002
HAS3	chr16:69143510-69143510	0.002
HAS3	chr16:69152282-69152282	0.002
MLH1	chr3:37056007-37056007	0.004
MLH1	chr3:37061893-37061893	0.002
MLH1	chr3:37061929-37061929	0.004
MLH1	chr3:37083821-37083821	0.002
MLH1	chr3:37089098-37089098	0.002
MLH1	chr3:37089130-37089130	0.008
MLH1	chr3:37089131-37089131	0.008
MLH1	chr3:37090087-37090087	0.004
MLH1	chr3:37090417-37090417	0.002
MLH1	chr3:37092019-37092019	0.006
MLH1	chr3:37092025-37092025	0.002
MSH2	chr2:47637301-47637301	0.002
MSH2	chr2:47643457-47643457	0.008
MSH2	chr2:47702358-47702358	0.002
MSH2	chr2:47705493-47705493	0.002
NFKB1	chr4:103432092-103432092	0.020
NFKB1	chr4:103432093-103432093	0.020
NFKB1	chr4:103514626-103514626	0.006
NFKB1	chr4:103516106-103516106	0.002
NFKB1	chr4:103518700-103518700	0.002
NFKB1	chr4:103537719-103537719	0.002
NOS1	chr12:117693787-117693787	0.002
NOS1	chr12:117693822-117693822	0.002
NOS1	chr12:117710246-117710246	0.004
NOS1	chr12:117725904-117725904	0.034
NOS1	chr12:117768154-117768154	0.009
NOS2	chr17:26093543-26093543	0.002
NOS2	chr17:26096131-26096131	0.002
NOS2	chr17:26108154-26108154	0.002
NOS2	chr17:26109102-26109102	0.011
NOS2	chr17:26110047-26110047	0.017

NOS2	chr17:26110055-26110055	0.002
NOS3	chr7:150693634-150693634	0.009
NOS3	chr7:150693636-150693636	0.009
NOS3	chr7:150695677-150695677	0.002
NOS3	chr7:150695728-150695728	0.002
NOS3	chr7:150696347-150696347	0.009
NOS3	chr7:150696348-150696348	0.006
NOS3	chr7:150696369-150696369	0.002
NOS3	chr7:150696378-150696378	0.004
NOS3	chr7:150696436-150696436	0.002
NOS3	chr7:150704236-150704236	0.002
NOS3	chr7:150709519-150709519	0.002
NR1I2	chr3:119526176-119526176	0.002
NR1I2	chr3:119526203-119526203	0.025
NR1I2	chr3:119526265-119526265	0.007
NR1I2	chr3:119535974-119535974	0.002
NR1I2	chr3:119536041-119536041	0.009
PTGS2	chr1:186643727-186643727	0.002
PTGS2	chr1:186645242-186645242	0.002
PTGS2	chr1:186646011-186646011	0.002
SLC22A16	chr6:110763506-110763506	0.004
SLC22A16	chr6:110763890-110763890	0.002
SLC22A16	chr6:110763935-110763935	0.028
SLC22A16	chr6:110768128-110768128	0.015
SLC22A16	chr6:110778099-110778099	0.017
SLC28A3	chr9:86900369-86900369	0.042
SLC28A3	chr9:86912182-86912182	0.002
SLC28A3	chr9:86912183-86912183	0.002
SLC28A3	chr9:86917301-86917301	0.040
TOP1MT	chr8:144398198-144398198	0.002
TOP1MT	chr8:144399938-144399938	0.007
TOP1MT	chr8:144408466-144408466	0.004
TOP1MT	chr8:144411609-144411609	0.002
TOP1MT	chr8:144413430-144413430	0.002
TOP2A	chr17:38545824-38545824	0.037
TOP2A	chr17:38546384-38546384	0.013
TOP2A	chr17:38559269-38559269	0.004
TOP2A	chr17:38564238-38564238	0.017
TOP2A	chr17:38567619-38567619	0.002
TOP2B	chr3:25639821-25639821	0.004

TOP2B	chr3:25646268-25646268	0.004
TOP2B	chr3:25659978-25659978	0.002
TOP2B	chr3:25660150-25660150	0.009
TOP2B	chr3:25660151-25660151	0.009
TOP2B	chr3:25666192-25666192	0.002
TOP2B	chr3:25666268-25666268	0.002
TOP2B	chr3:25668084-25668084	0.002
TOP2B	chr3:25674005-25674005	0.004
TOP2B	chr3:25674035-25674035	0.004
UGT1A1	chr2:234677027-234677027	0.004
UGT1A1	chr2:234680917-234680917	0.004
UGT1A1	chr2:234680986-234680986	0.002
UGT1A6	chr2:234601816-234601816	0.002
UGT1A6	chr2:234601817-234601817	0.002
UGT1A6	chr2:234602507-234602507	0.002
XDH	chr2:31560536-31560536	0.002
XDH	chr2:31560572-31560572	0.021
XDH	chr2:31562390-31562390	0.002
XDH	chr2:31565072-31565072	0.004
XDH	chr2:31570459-31570459	0.004
XDH	chr2:31588857-31588857	0.002
XDH	chr2:31588938-31588938	0.002
XDH	chr2:31590859-31590859	0.002
XDH	chr2:31590917-31590917	0.038
XDH	chr2:31600072-31600072	0.006
XDH	chr2:31605901-31605901	0.002
XDH	chr2:31609376-31609376	0.002
XDH	chr2:31610711-31610711	0.002
Mitochondrion function data set		
AIFM2	chr10:71874714-71874714	0.002
AIFM2	chr10:71876524-71876524	0.004
AIFM2	chr10:71883836-71883836	0.002
AIFM3	chr22:21328571-21328571	0.002
AIFM3	chr22:21330954-21330954	0.002
AIFM3	chr22:21331032-21331032	0.003
AIFM3	chr22:21333954-21333954	0.002
COX10	chr17:13977645-13977645	0.004
COX10	chr17:13977646-13977646	0.004
COX10	chr17:13980176-13980176	0.002
COX10	chr17:13980356-13980356	0.006

COX11	chr17:53040128-53040128	0.004
COX11	chr17:53040261-53040261	0.006
COX11	chr17:53045929-53045929	0.002
KAT2A	chr17:40265734-40265734	0.002
KAT2A	chr17:40266241-40266241	0.009
KAT2A	chr17:40267842-40267842	0.002
NCOA3	chr20:46252801-46252801	0.002
NCOA3	chr20:46256321-46256321	0.002
NCOA3	chr20:46264171-46264171	0.013
NCOA3	chr20:46268708-46268708	0.002
NCOA3	chr20:46268732-46268732	0.002
NCOA3	chr20:46279753-46279753	0.002
NCOA3	chr20:46279918-46279918	0.008
NDUFA10	chr2:240913012-240913012	0.009
NDUFA10	chr2:240944661-240944661	0.004
NDUFA10	chr2:240951057-240951057	0.002
NDUFA10	chr2:240951071-240951071	0.013
NDUFA10	chr2:240957978-240957978	0.002
NDUFA10	chr2:240960659-240960659	0.006
NDUFA10	chr2:240960712-240960712	0.002
NDUFA5	chr7:123196827-123196827	0.002
NDUFA5	chr7:123196884-123196884	0.002
NDUFA5	chr7:123196936-123196936	0.004
NDUFA5	chr7:123196955-123196955	0.026
NDUFA5	chr7:123197120-123197120	0.011
NDUFA5	chr7:123197142-123197142	0.002
NDUFA9	chr12:4763994-4763994	0.007
NDUFA9	chr12:4777682-4777682	0.002
NDUFA9	chr12:4791451-4791451	0.013
NDUFA9	chr12:4796219-4796219	0.002
NDUFS2	chr1:161179085-161179085	0.002
NDUFS2	chr1:161180482-161180482	0.004
NDUFS2	chr1:161183446-161183446	0.002
NDUFV1	chr11:67375944-67375944	0.002
NDUFV1	chr11:67377928-67377928	0.019
NDUFV1	chr11:67378519-67378519	0.002
NDUFV3	chr21:44317065-44317065	0.002
NDUFV3	chr21:44317156-44317156	0.015
NDUFV3	chr21:44323856-44323856	0.002
PPARGC1A	chr4:23803919-23803919	0.002

PPARGC1A	chr4:23814701-23814701	0.004
PPARGC1A	chr4:23814707-23814707	0.044
PPARGC1A	chr4:23816049-23816049	0.002
PPARGC1A	chr4:23826104-23826104	0.002
PPARGC1A	chr4:23831108-23831108	0.002
PPARGC1A	chr4:23886388-23886388	0.002
PPARGC1A	chr4:23886457-23886457	0.028
SIRT3	chr11:218791-218791	0.004
SIRT3	chr11:219008-219008	0.003
SIRT3	chr11:224194-224194	0.002
SIRT3	chr11:230474-230474	0.015
SIRT3	chr11:233037-233037	0.002

SNV: single nucleotide variation; MAF: minor allele frequency.

5.12.2. Supplemental Table 2. Complete results of the SKAT-O test, the candidate genes association study.

Doxorubicin pathway, All cohort N=233						
Outcome	**Gene**	**SNVs tested**		**MAF**	**SKAT-O**	
		Complete position	**rs number**		**P value**	**FDR**
LVEF M-mode (%)	***NOS1***	chr12:117693787-117693787		0.002	0.0013	0.020
		chr12:117693822-117693822	rs200972861	0.002		
		chr12:117710246-117710246	rs41356652	0.004		
		chr12:117725904-117725904		0.03		
		chr12:117768154-117768154*	rs76090928	0.009		
	ABCG2	chr4:89016692-89016692		0.002	0.0026	0.020
		chr4:89016693-89016693		0.002		
		chr4:89016694-89016694		0.002		
		chr4:89018670-89018670	rs45605536	0.002		
		chr4:89052317-89052317		0.002		
		chr4:89052998-89052998*	rs199473672	0.008		
		chr4:89060981-89060981		0.002		
Doxorubicin pathway, SR group N=106						
LVEF M-mode (%)	***CBR1***	chr21:37443349-37443349	rs41557318	0.005	0.0015	0.037
		chr21:37444697-37444697*	rs2835266	0.023		
		chr21:37444971-37444971		0.005		
		chr21:37445131-37445131		0.005		
Doxorubicin pathway, HR group N=127						
LVEF M-mode (%)	***ABCC5***	chr3:183669271-183669271		0.01	0.0008	0.016
		chr3:183677620-183677620		0.02		
		chr3:183677621-183677621		0.02		
		chr3:183679412-183679412		0.01		
		chr3:183681203-183681203*		0.01		
		chr3:183681255-183681255*		0.03		
		chr3:183689395-183689395	rs368112213	0.004		
		chr3:183695316-183695316		0.004		
Doxorubicin pathway, SR group N=106						
FS M-mode (%)	***ABCG2***	chr4:89018670-89018670	rs45605536	0.005	0.0027	0.040
		chr4:89052998-89052998*	rs199473672	0.014		
		chr4:89060981-89060981		0.005		
	CBR1	chr21:37443349-37443349	rs41557318	0.005	0.0037	0.040
		chr21:37444697-37444697*	rs2835266	0.023		
		chr21:37444971-37444971		0.005		

		chr21:37445131-37445131		0.005		
Doxorubicin pathway, HR group N=128						
FS M-mode (%)	***ABCC5***	chr3:183669271-183669271		0.01	0.0014	0.035
		chr3:183677620-183677620		0.02		
		chr3:183677621-183677621		0.02		
		chr3:183679412-183679412		0.01		
		chr3:183681203-183681203*		0.01		
		chr3:183681255-183681255*		0.03		
		chr3:183689395-183689395	rs368112213	0.004		
		chr3:183695316-183695316		0.004		
Doxorubicin pathway, SR group N=108						
LVEF 2D (%)	***ABCG2***	**chr4:89018670-89018670***	rs45605536	0.005	0.0019	0.009
		chr4:89052998-89052998*	rs199473672	0.014		
		chr4:89060981-89060981*		0.005		
	AKR1C3	**chr10:5139642-5139642***	rs200981816	0.005	0.007	0.017
		chr10:5141609-5141609*	rs34186955	0.005		

SNV: single nucleotide variation; LVEF: left ventricular ejection fraction; FS: Fractional shortening**;** HR: high risk; SR: standard risk; MAF: minor allele frequency; FDR: false discovery rate.
NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group); CBR1: Carbonyl Reductase 1; ABCC5: ATP Binding Cassette Subfamily C Member 5; AKR1C3: Aldo-Keto Reductase Family 1 Member C3.
*SNVs that are identified as the most important contributors to the association signal and presented in more details in Table 4.

5.12.3. Supplemental Table 3. Multiple linear regression model that included genetic variant combination of the two most contributing rare variant associations with LVEF M-mode (%), N=236.

TEST	BETA	L95	U95	P
NOS1* chr12:117768154	**0.23**	**0.11**	**0.35**	**0.00024**
ABCG2* chr4:89052998	**0.25**	**0.13**	**0.37**	**0.00006**
Age at diagnosis	0.09	-0.04	0.23	0.1584
Time end treatment	0.30	-0.07	0.63	0.1159
Sex	-0.02	-0.14	0.10	0.7184
Protocol	0.37	0.02	0.73	**0.0417**
Risk-treatment **	-0.16	-0.30	-0.02	**0.0217**

LVEF: left ventricular ejection fraction; *NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group).*
*SNVs that are identified as the most important contributors to the association signal are presented.
** Risk-treatment combined variable with the following scores: SR patients =1, HR patients that received a cardioprotective drug (dexrazoxane)=2, HR patients that did not receive a cardioprotective drug=3 (categorical variable).

5.12.4. Supplemental Table 4. Complete results of the SKAT-O test, the exome-wide association study.

Exome-wide association, All cohort, N=233						
Outcome	**Gene**	**SNVs tested**		**MAF**	**SKAT-O**	
		position	**rs number**		**P value**	**FDR**
FS M-mode (%)	***NOD2***	chr16:50731162-50731162		0.009	1.39×10^{-6}	0.014
		chr16:50733823-50733823		0.002		
		chr16:50741791-50741791	rs61755182	0.008		
		chr16:50744753-50744753	rs104895427	0.002		
		chr16:50744927-50744927		0.006		
		chr16:50745099-50745099		0.004		
		chr16:50745114-50745114*	rs104895431	0.006		
		chr16:50745233-50745233	rs1078327	0.002		
		chr16:50746086-50746086*	rs61747625	0.007		
		chr16:50746100-50746100*	rs3813758	0.002		
		chr16:50746199-50746199*	rs104895444	0.004		
		chr16:50756540-50756540	rs2066845	0.025		
		chr16:50757286-50757286		0.004		
		chr16:50759443-50759443	rs148561632	0.002		
	ZNF267	**chr16:31926312-31926312***	.	0.002	3.5×10^{-6}	0.018
		chr16:31927308-31927308*	rs118056264	0.011		
		chr16:31927353-31927353	rs79859029	0.002		
		chr16:32077452-32077452*	.	0.002		
		chr16:32077650-32077650	rs200744708	0.003		
		chr16:32077657-32077657	rs144576359	0.005		
Exome-wide association, All cohort, N=233						
LVEF M-mode (%)	***NOD2***	chr16:50731162-50731162		0.009	2×10^{-6}	0.025
		chr16:50733823-50733823		0.002		
		chr16:50741791-50741791	rs61755182	0.008		
		chr16:50744753-50744753	rs104895427	0.002		
		chr16:50744927-50744927		0.006		
		chr16:50745099-50745099		0.004		
		chr16:50745114-50745114*	rs104895431	0.006		
		chr16:50745233-50745233	rs1078327	0.002		
		chr16:50746086-50746086*	rs61747625	0.007		
		chr16:50746100-50746100*	rs3813758	0.002		
		chr16:50746199-50746199*	rs104895444	0.004		
		chr16:50756540-50756540	rs2066845	0.025		
		chr16:50757286-50757286		0.004		
		chr16:50759443-50759443	rs148561632	0.002		

SNV: single nucleotide variation; LVEF: left ventricular ejection fraction, MAF: minor allele frequency; FDR: false discovery rate; *ZNF267: Zinc Finger Protein 267; NOD2: Nucleotide Binding Oligomerization Domain Containing 2*.

5.12.5. Supplemental Table 5. Demographic characteristics of SJLIFE participants included for replication analysis.

Replication Analysis	Outcome	Group	N	Age at diagnosis (min/median/max)	Years since the end of treatment (min/median/max)	Sex (n female/n male)
Candidate genes,	LVEF-2D	SR	102	0.2/4.2/17.4	17.9/28.6/50.0	49/53
total number=149	LVEF-2D	HR	46	1.9/4.8/15.2	24.7/33.5/38.0	23/23
	LVEF-2D	All	148	0.2/4.6/17.4	17.9/29.7/50.0	72/76
	LVEF M-Mode	SR	56	0.9/4.3/17.4	17.9/23.0/45.3	31/25
	LVEF M-Mode	HR	32	2.4/5.1/15.2	25.1/28.8/38.0	15/17
	LVEF M-Mode	All	88	0.9/4.8/17.4	17.9/26.2/45.3	46/42
	FS M-Mode	SR	56	0.9/4.3/17.4	17.9/23.0/45.3	31/25
	FS M-Mode	HR	32	2.4/5.1/15.2	25.1/28.8/38.0	15/17
	FS M-Mode	All	88	0.9/4.8/17.4	17.9/26.2/45.3	46/42
total	At least one measurement for one of the above outcomes	SR	103	0.2/4.3/17.4		50/53
		HR	46	1.9/4.8/15.2	varies by outcome	23/23
		All	149	0.2/4.6/17.4		73/76
Exome-wide association study,	LVEF-2D	SR	109	0.2/4.5/17.4	17.9/28.5/50.0	51/58
total number=158	LVEF-2D	HR	48	1.9/4.7/15.2	24.7/33.3/38.0	24/24
	LVEF-2D	All	157	0.2/4.6/17.4	17.9/29.7/50.0	75/82
	LVEF M-Mode	SR	58	0.9/4.3/17.4	17.9/23.0/45.3	31/27
	LVEF M-Mode	HR	32	2.4/5.1/15.2	25.1/28.8/38.0	15/17
	LVEF M-Mode	All	90	0.9/4.8/17.4	17.9/26.2/45.3	46/44
	FS M-Mode	SR	58	0.9/4.3/17.4	17.9/23.0/45.3	31/27
	FS M-Mode	HR	32	2.4/5.1/15.2	25.1/28.8/38.0	15/17
	FS M-Mode	All	90	0.9/4.8/17.4	17.9/26.2/45.3	46/44
total	At least one measurement for one of the above outcomes	SR	110	0.2/4.6/17.4		52/58
		HR	48	1.9/4.7/15.2	varies by outcome	24/24
		All	158	0.2/4.7/17.4		76/82

LVEF: Left ventricular ejection fraction; FS: fractional shortening; SR: standard risk; HR: high risk; All refers to all patients tested independent of the risk group (HR+SR).

5.12.6. Supplemental Table 6. 1980s SR and HR childhood ALL risk group definitions from the Childhood Cancer Survivor Study (48).

Exposure/Event	1980s SR-like	1980s HR-like
Cranial Radiation	0<CRT≤20 Gy	>0 Gy
Dexamethasone	No	No
Anthracycline cumulative dose	≤120 mg/m^2	>120 mg/m^2
Cyclophosphamide	Yes or No	Yes or No
Cytarabine, IV*	Yes or No	Yes
Relapse	No	No
Transplant	No	No

SR: standard risk; HR: high risk; ALL: childhood Acute Lymphoblastic Leukemia, IV: intravenous.
* Is exposure to IV cytarabine/high dose cytarabine or IT cytarabine. Only a handful of ALL survivors had IT but not IV cytarabine exposure.
Exposures/events are only considered if they occurred within five years of the ALL primary diagnosis date.

5.12.7. Supplemental Table 7. Summary of non-monomorphic variant genotypes passing quality control in SJLIFE replication cohort.

Gene	Nb. of Variants in the Discovery cohort	Nb. of Discovery Variants Passing SJLIFE QC (Nb. of Discovery Variants available* in gnomAD)	Nb. of SJLIFE QC-passing Variants Monomorphic in SJLIFE	Nb. of Variants Analyzed in SJLIFE
TTN	3	3 (3)	0	3
NOS1	5	3 (3)	1	2
ABCG2	7	3 (3)	2	1
CBR1	4	2 (2)	1	1
AKR1C3	2	2 (2)	1	1
ZNF267	6	3 (3)	1	2
NOD2	14	10 (10)	9	1
ABCC5	8	0 (0)	-	0
Total	49	26 (26)	15	11

SJLIFE: St. Jude Lifetime Cohort; QC: quality control, *TTN*: *titin*; *NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group); CBR1: Carbonyl Reductase 1; AKR1C3: Aldo-Keto Reductase Family 1 Member C3; ZNF267: Zinc Finger Protein 267; NOD2: Nucleotide Binding Oligomerization Domain Containing 2; ABCC5: ATP Binding Cassette Subfamily C Member 5*.

*gnomAD availability means there was a variant at the locus in gnomAD, the variant passed gnomAD QC, and the frequency of the variant was >0.015% (0.015% would mean approximately one variant carrier expected in SJLIFE).

5.12.8. Supplemental Table 8. Replication analysis of the *TTN* common variants, SJLIFE cohort.

Outcome (Group)	Variant ID in the proposal (hg19)	ID in SJLIFE WGS (hg38)	Quality Control			Linear Regression					
			MISS	MAF	HWE	N [a]	11 [b]	12+22 [b]	Beta	Se	*P* [c]
LVEF	chr2:179397561	chr2.178532834.C.T	0.00	0.15	0.10	62/26	0.65(0.08)	0.66(0.06)	0.014	0.017	0.413
M-Mode	chr2:179444939	chr2.178580212.C.T	0.00	0.18	0.17	57/31	0.65(0.08)	0.65(0.07)	0.002	0.017	0.894
(All)	chr2:179575511	chr2.178710784.C.T	0.00	0.07	0.45	75/13	0.65(0.08)	0.66(0.06)	0.007	0.022	0.740
FS	chr2:179397561	chr2.178532834.C.T	0.00	0.15	0.10	62/26	0.36(0.06)	0.37(0.04)	0.010	0.013	0.433
M-Mode	chr2:179444939	chr2.178580212.C.T	0.00	0.18	0.17	57/31	0.36(0.06)	0.36(0.05)	0.001	0.013	0.926
(All)	chr2:179575511	chr2.178710784.C.T	0.00	0.07	0.45	75/13	0.36(0.06)	0.36(0.04)	0.004	0.017	0.818

TTN: *titin*; LVEF: Left ventricular ejection fraction; FS: fractional shortening; MISS: missing genotyping rate; MAF: minor allele frequency; HWE: p-value of Hardy-Weinberg Equilibrium test.
All refers to all patients tested independent of the risk group.
N, genotype counts: 11 / 12+22.
Allele coding for all three variants: 1=C, 2=T.
P, p-value from linear regression adjusting for age at diagnosis, sex, time since end of treatment, and risk-treatment factor.

5.12.9. Supplemental Table 9. Replication analysis of rare variants, SJLIFE cohort.

Outcome (Group)	Gene	SKAT-O test			Analysis of the most important contributor through Linear Regression						
		N1	n	P1	Variants tested	N2	11	12+22	Beta	Se	P2
LVEF M-Mode (All)	*NOS1*	88	2	0.402	chr12:117768154*	84/4	0.65(0.07)	0.71(0.06)	0.06	0.04	0.100
LVEF M-Mode (SR)	*CBR1*	56	1	.	chr21:37444697*	49/7	0.67(0.070)	0.62(0.09)	-0.04	0.03	0.241
LVEF-2D (SR)	*ABCG2*	102	1	.	chr4:89052998*	101/1*	0.63(0.05)	0.59(NA)	-0.03	0.05	0.488
	AKR1C3	102	1	.	chr10:5139642*	101/1*	0.63(0.05)	0.6(NA)	-0.02	0.05	0.603
FS M-Mode (SR)	*CBR1*	56	1	.	chr21:37444697*	49/7	0.37(0.05)	0.34(0.07)	-0.02	0.02	0.292
FS M-Mode (All)	*ZNF267*	90	2	0.373	chr16:31927308*	87/3*	0.36(0.05)	0.35(0.05)	-0.01	0.03	0.671
	NOD2	90	1	.	chr16:50756540	87/3	0.36(0.05)	0.37(0.08)	0.01	0.03	0.729
LVEF M-Mode (SR)	*NOD2*	90	1	.	chr16:50756540	87/3	0.65(0.07)	0.66(0.10)	0.01	0.04	0.826

LVEF: Left ventricular ejection fraction; FS: Fractional Shortening; *NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2 (Junior Blood Group); CBR1: Carbonyl Reductase 1; AKR1C3: Aldo-Keto Reductase Family 1 Member C3; ZNF267: Zinc Finger Protein 267; NOD2: Nucleotide Binding Oligomerization Domain Containing 2; ABCC5: ATP Binding Cassette Subfamily C Member 5*.
All refers to all patients tested independent of the risk group.
N1, number of samples used the SKAT-O test.
n, number of rare variants used in the SKAT-O test.
P1, p-value of SKAT-O test.
N2, genotype counts: 11/12+22.
* the most significant rare contributors requested for replication within specific outcome.
P2, p-value from linear regression adjusting for age at diagnosis, sex, time since end of treatment, and risk-treatment factor.
The analyses are performed as SKAT-O (combination of available variants in the gene) and as collapsing analyses of the most important contributor(s) detected in the discovery cohort. Please note that the results (p values) for SKAT-O analyses are not included if only one rare variant was detected in the gene. Genes in which all discovery variants were monomorphic in SJLIFE for a particular phenotype included *ABCC5* for all outcomes and the *ABCG2* gene, except for LVEF 2D mode for the latter.

5.12.10. Supplemental Table 10. *TTN* common variants available in the PETALE whole-exome sequencing data.

(A) *TTN* common variants details.

Gene	Position (hg19)	Ref	Var	rs number	Amino Acid change	Function	MAF (PETALE)
TTN	**chr2:179397561**	**C**	**T**	**rs3829747**	**TTN:NM_003319:exon186: c.G76586A:p.R25529H**	**nonsynonymous SNV**	0.125
	chr2:179444939	**C**	**T**	**rs2303838**	**TTN:NM_003319:exon146: c.G39880A:p.V13294I**	**nonsynonymous SNV**	0.172
	chr2:179464527	T	C	rs1001238	TTN:NM_003319:exon117: c.A28906G:p.N9636D	nonsynonymous SNV	0.233
	chr2:179558366	T	C	rs2042995	TTN:NM_133378:exon116: c.A27832G:p.I9278V	nonsynonymous SNV	0.234
	chr2:179575511	**C**	**T**	**rs72648998**	**TTN:NM_133378:exon95: c.G24581A:p.R8194Q**	**nonsynonymous SNV**	0.054
	chr2:179582537	G	T	rs2627043	TTN:NM_133378:exon84: c.C21332A:p.A7111E	nonsynonymous SNV	0.21
	chr2:179650408	*G*	*A*	*rs35813871*	*TTN:NM_003319:exon14: c.C2294T:p.T765I NP_001254479.2:p.Thr811Ile*	*nonsynonymous SNV*	*0.252*
	chr2:179659912	G	A	rs16866538	TTN:NM_001256850:exon7: c.C982T:p.R328C	nonsynonymous SNV	0.057

(B) Haplotype that is tagged by *rs35813871.*

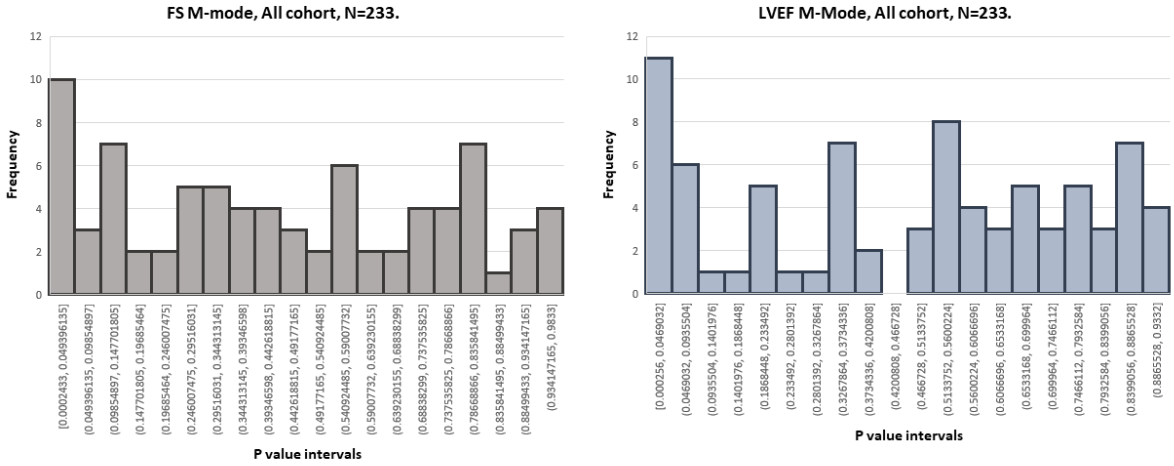
Order of SNVs	Haplotype sequence*	Haplotype frequency (SD)**
rs3829747	0	0.225 (0.004)
rs2303838	0	
rs1001238	0	
rs2042995	0	
rs72648998	0	
rs2627043	0	
rs35813871	1	
rs16866538	0	

TTN: *titin*; SNV: Single Nucleotide Variation; MAF: minor allele frequency; Ref: reference allele; Var: variant allele, SD: standard deviation.
The SNVs identified as top-ranking associations (FDR<=5%) with protective effect are highlighted, the risk-increasing variant is represented in italics.
*0 corresponds to reference allele, 1 corresponds to variant allele.
**Estimation of the haplotype frequency for the entire PETALE cohort.

5.12.11. Supplemental Table 11. Association results for the *TTN* common risk-increasing variant.

All cohort, N=236				Genetic models							
Outcome	rs number	P value*	FDR	Add	Mean	SD	P value**	Rec	Mean	SD	P value ***
LVEF M-mode	rs35813871	0.024	0.24	130/89 /14	60.5/59.97 /55.99	5.97/ 5.5/4.2	0.017	219 /14	60.3/ 55.99	5.8/ 4.2	0.009
FS M-mode	rs35813871	0.021	0.24	130/89 /14	35.5/34.9 /32.69	4.1/ 3.99/2.97	0.024	219 /14	35.3/ 32.7	4.1/ 2.97	0.035

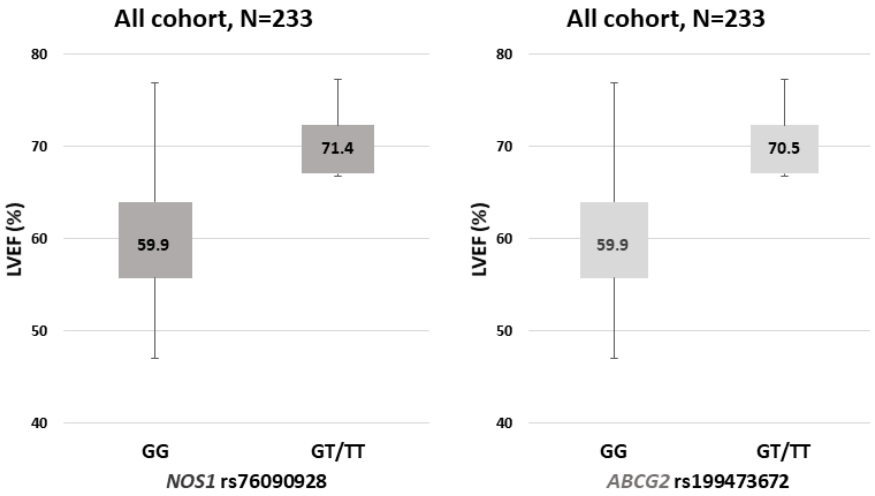
LVEF: Left ventricular ejection fraction; FS: Fractional Shortening; FDR: false discovery rate; SD: standard deviation; Add: additive model; Rec: recessive model.
*P value was obtained using Quantitative trait association and allelic model based on Wald statistic implemented in PLINK.
**P value for additive genetic model was obtained through linear regression implemented in SPSS, multiple regression model included the above-described adjustment variables: age at the time of diagnosis; time since the end of treatment; sex; DFCI Protocol; and a risk-treatment combined variable.
***P value for recessive genetic model was obtained through linear regression implemented in SPSS, multiple regression model included the above-described adjustment variables: age at the time of diagnosis; time since the end of treatment; sex; DFCI Protocol; and a risk-treatment combined variable.



5.12.12. Supplemental Figure 1. P value distribution of the common variants tested through candidate genes approach for FS M-Mode (%) and LVEF M-Mode (%) represented in histograms.

The P value distribution of all common variants (N=80) tested for each outcome (FS M-Mode (%) and LVEF M-Mode (%)) was shown in the histograms. The P values are distributed between 0 and 1. There is an apparent overabundance of low P values, suggesting the existence of genes that have a potential modulating effect on LVEF and FS.

LVEF: Left ventricular ejection fraction; FS: fractional shortening.



5.12.13. Supplemental Figure 2. The top-ranking rare loci associated with LVEF M-mode (%) identified in the complete cohort (N=236) through the candidate genes approach.

A box plot represents the distribution of the mean values of the LVEF M-mode (%) between genotypes, genotypes were recoded according to the dominant model.

LVEF: left ventricular ejection fraction; *NOS1: Nitric Oxide Synthase 1 (Neuronal); ABCG2: ATP Binding Cassette Subfamily G Member 2.*

Section B

Chapter 6

Genetic factors in anthracycline-induced cardiotoxicity in patients treated for pediatric cancer.

Expert Opin Drug Metab Toxicol. 2020 Oct;16(10):865-883. doi: 10.1080/17425255.2020.1807937. Epub 2020 Sep 24.

PMID: 32772754 DOI: 10.1080/17425255.2020.1807937

Genetic factors in anthracycline-induced cardiotoxicity in patients treated for pediatric cancer.

Kateryna Petrykey[1,2], Gregor U. Andelfinger[1,3], Caroline Laverdière[1,3], Daniel Sinnett[1,3] and Maja Krajinovic[1,2,3]

1. Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada.
2. Department of Pharmacology and Physiology, Université de Montréal (Quebec), Canada.
3. Department of Pediatrics, Université de Montréal (Quebec), Canada.

Corresponding author:

Maja KRAJINOVIC
Sainte-Justine University Health Center (SJUHC)
3175 Chemin de la Côte-Sainte-Catherine, Montreal (Quebec), Canada, H3T 1C5
E-mail: maja.krajinovic@umontreal.ca
Fax: 514-345-4731
Phone: 514-345-4931 #6259

6.1. ABSTRACT

Introduction. Cardiovascular diseases are the main cause of treatment-related morbidity and mortality in childhood cancer patients. Anthracyclines, one of the most common chemotherapeutic agents in treatment regimens, are implicated in chemotherapy-induced cardiotoxicity.

Areas covered: This review describes the pharmacogenomic markers related to anthracycline-induced cardiotoxicity affecting childhood cancer patients. We also included a brief overview of the applicability of reported findings to the well established PETALE cohort of childhood acute lymphoblastic leukemia survivors of the Sainte-Justine University Health Center (Canada).

Expert opinion: The wide variation in interindividual susceptibility to anthracycline-induced cardiotoxicity along with a multitude of genetic variants detected through association studies suggests that genetic contributions could be essential for the design of new individualized preventive approaches.

Keywords

anthracycline-induced cardiotoxicity, genetic association studies, pharmacogenomic markers, doxorubicin, childhood cancer.

6.2. ARTICLE HIGHLIGHTS

Anthracyclines, highly effective anticancer agents widely used to treat a variety of childhood malignancies, are linked to chemotherapy-induced cardiotoxicity.

Known clinical and treatment-related risk factors do not adequately explain variability in response to anthracyclines, suggesting the contribution of genetic determinants.

Gene-treatment interactions in the context of specific treatment protocols are important for the implementation of pharmacogenetics findings.

6.3. INTRODUCTION

Remarkable improvement in cure rates for childhood cancer has been achieved over the past few decades due to recent advances in cancer research and diagnosis, combined with the introduction of multi-agent treatment protocols and the optimization of dosing regimens. The use of anthracyclines, highly effective anticancer agents, contributed significantly to the increase of 5-year survival rates for several childhood cancers[1-4]. However, despite the fact that approximately 80% of children with cancer are currently recovering from their illness[5,6], many of them are facing long-term morbidity. Over 60% of survivors suffer from treatment-related effects, which may include, among others, growth issues, cardiac and metabolic problems, neurocognitive and motor impairments, neurological and endocrine disorders, emotional difficulties, and fertility problems[1,7-9]. Cardiovascular diseases are the main cause of secondary morbidity and mortality in childhood cancer patients[1,10,11]. These patients are eight times more prone to death from cardiovascular diseases than the normal population[12,13]. Indeed, manifested as asymptomatic cardiac dysfunction in up to 57% cancer survivors[14], cardiomyopathy may lead to subsequent congestive heart failure in up to 16% of cases[15].

Anthracyclines are an essential component of childhood cancer therapy, with *doxorubicin* being the most used agent administered in a wide spectrum of hematological and solid malignancies including leukemia, lymphoma, and sarcomas[16-18]. Approximately 50–60% of childhood cancer survivors have been treated with an anthracycline regimen[2,10,18,19]. The clinical presentation of anthracycline-induced cardiotoxicity (ACT) may manifest as decreased ability to perform physical exercise, chest pain and shortness of breath[4,10,20]. Additional attention is required for asymptomatic patients, their identification is important as the effects of cardiac toxicity could be reversible if detected early[19,21].

ACT is defined as a more than 10% decrease in left ventricular ejection fraction (LVEF) to levels of less than 50-55% compared with pre-treatment values and, depending on the population and duration of follow-up, it is associated with heart failure (HF) in 5% to 26% cases[22-26]. Based on the time of manifestation, ACT can be categorized as acute, or ACT with early-and late-onset. In acute ACT, symptoms manifest within hours or days of administration, early-onset ACT occurs within 1 year of treatment, and late-onset develops ≥ 1 year after treatment [19]. Early- and late-onset ACT are defined by progressive LV dysfunction, which, in some cases, leads to subsequent congestive HF[3,17,19].

Higher cumulative doses are an independent risk factor for ACT[27,28]; thus, explaining the fact that up to 60% of patients exposed to high doses (> 250 mg/m^2) of anthracyclines will subsequently develop cardiac abnormalities[24,27]. However, some patients may have subclinical cardiotoxicity even at low doses[29-31]. Known clinical and treatment-related risk factors (such as young age, female gender, cumulative anthracycline dose, and radiotherapy involving the cardiac region)[8,32] do not adequately explain variability in response to anthracyclines, thereby suggesting the contribution of other determinants. Indeed, several research groups explored the potential of genetics in predicting ACT and improving the risk stratification. Identification of such markers could allow identifying individuals who may benefit from additional monitoring, cardioprotective medications, and early initiation of treatment.

The purpose of this review is to summarize the current knowledge of genetic variants that contribute to ACT in patients treated for childhood cancer. It should be mentioned that some reported genetic markers have been identified through studies that evaluated children with cancer treated with anthracyclines for both early and late cardiotoxicity; while the other markers have been reported in pediatric cancer survivors years after treatment.

6.4. PATHOPHYSIOLOGY OF ACT

Chemotherapy related cardiotoxicity can be characterized as type 1 or type 2 depending on the effect of the drug on cardiomyocytes[25]. Type I cardiotoxicity implies the death of cardiomyocytes either through necrosis or apoptosis and is irreversible, whereas type II is due to cardiomyocyte dysfunction rather than cell death, and may be reversible[21,22]. Long-term cardiotoxicity caused by anthracyclines includes the permanent cardiac damage and subsequent death of cardiomyocytes and, therefore, represents the type I toxicity[21,33].

The current inability to predict and prevent the cardiotoxicity of anthracyclines is partly due to the fact that the underlying molecular mechanisms are complex and not fully understood[34].

Based on studies in animal and in vitro models, several hypotheses have been proposed to explain the mechanisms of damage to cardiomyocytes by anthracyclines. The suggested mechanisms are summarized in **Figure 1** and are described only briefly (more details can be found elsewhere[8,19,35-39], as an introduction for the selection of candidate genes for association studies. Some of these mechanisms are detailed further depending on the finding of the described studies.

6.4.1. Oxidative stress

Oxidative stress is one of the most studied and commonly accepted cellular mechanism related to ACT[3,4]. There is no doubt that the biochemical transformations of anthracyclines lead to the formation of reactive oxygen species (ROS). However, the potential involvement of the formation of ROS in the cytotoxicity of anthracyclines (both in context of antitumor activity and cardiotoxicity) is complex and subject to much discussion[40-42]. The quinone structure allows anthracyclines to act as electron acceptors in reactions mediated by oxidoreductive enzymes, including cytochrome P450 reductase, NADH dehydrogenase, and xanthine oxidase[43,44]. Conversion of quinone to semiquinone is associated with ROS generation, which can cause the DNA damage inducing the activation of apoptotic pathways and cell death[21].

One of the reasons for the high sensitivity of cardiomyocytes to oxidative stress caused by anthracyclines could be explained by their high dependence on the metabolism of the oxidative substrate (due to high volume of mitochondria in cardiomyocytes compared to glycolytic tumor cells). Indeed, the oxidative damage induced by doxorubicin in tumor cells was observed only at very high concentrations of the drug[37]. Moreover, the heart has a reduced ability to block the formation of the ROS by various scavengers such as superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase[45,46] which may also sensitise cardiac myocytes to ROS induced injury[2,19].

While the formation of ROS is induced by the quinone component of anthracyclines, oxidative stress can also arise through the induction of nitric oxide synthase, which leads to the formation of nitric oxide and peroxynitrite[47]. This mechanism has been associated with nitration and inactivation of key enzymes in the heart, including myofibrillar creatine kinase[34].

6.4.2. DNA damage through interference with topoisomerase II

Topoisomerase II (TOP2) is an adenosine triphosphate (ATP)-dependent enzyme with several crucial functions such as involvement in DNA replication, transcription and chromosome segregation[35,48]; it is expressed in the form of TOP2α and TOP2β isoenzymes[35] . TOP2α is highly expressed in proliferating cells, whereas TOP2β is the major form in the heart muscle[41]. The formation of a ternary cleavage complex (DNA-TOP2-Anthracyclines) leads to double-stranded DNA breaks[35,49]. While in the proliferating malignant cells this complex inhibits DNA replication and induces apoptosis as intended, the damage to non-proliferating cardiomyocytes will lead to adverse effects[35]. For example, doxorubicin disrupts the normal catalytic cycle of TOP2β, causing DNA double-stranded breaks that in turn result in p53 activation[50]. This further changes the transcriptome, leading to defective mitochondrial biogenesis and increase in ROS[41,51]. Consequently, cardiomyocytes presented myofibrillar disarray and vacuolization[51].

6.4.3. Accumulation of drug and cardiotoxic metabolites

The individual differences in the rate of conversion of anthracyclines into active metabolites as well as their absorption/elimination by transmembrane carriers may affect the risk to develop ACT [33]. For example, anthracycline alcohol metabolites form a reservoir in cardiomyocytes[32,52] and impair contractility by inhibiting the activity of the Ca2+ and Na+/K+ pumps[32,37]. Interindividual variability in the formation of these metabolites may affect the risk of cardiomyopathy[53]. The synthesis of cardiotoxic alcohol metabolites is catalyzed by myocardial cytosolic carbonyl reductases (CBRs)[52,53] which are considered as the main metabolizing enzymes of anthracyclines in the human heart[37].

Several transporters can modify anthracyclines uptake or efflux changing their plasmatic and cellular concentration. The solute carriers (SLC) are known to encode membrane proteins that can transport a large number of endogenous compounds and drugs[54]. Moreover, the expression of these transporters in the myocardium may significantly contribute to anthracycline induced cardiotoxicity[55]. Adenosine triphosphate-binding cassette subfamily C (ABC) transporters include several efflux pumps that influence the level of anthracyclines in leukemia cells and other tissues, including myocardium[56-60].

6.4.4. Disruption of iron metabolism

Iron is supposed to play a crucial role in the ACT due to its important role in the generation of ROS [35]. More precisely, formation of doxorubicin-Fe2+ complexes may catalyse a Fenton reaction (Fe2+-catalysed transformation of hydrogen peroxide to hydroxyl radical) resulting in the generation of ROS[35,61]. Several studies have indicated that iron loading aggravates the toxic effects of anthracyclines. For example, patients with higher concentrations of iron in tissues (individuals with hereditary disorders of iron metabolism) have an increased risk of heart damage because the accumulation of iron can enhance the response of ROS to the effects of anthracyclines.

Interestingly, the only FDA-approved cardioprotective agent – dexrazoxane[62], has two biological activities: it may chelate iron, which prevents anthracycline-induced oxidative damage[63] and can change the configuration of TOP2, and thus its binding to anthracyclines. This can subsequently improve anti-oxidant gene expression preventing mitochondrial dysfunction and cardiomyocyte death[62].

6.5. GENETIC VARIATIONS ASSOCIATED WITH ANTHRACYCLINE-INDUCED CARDIOTOXICITY

A total of 45 SNPs located in 34 genes were reported related to ACT in patients treated for childhood cancer. The summary of these studies with all details, including authors, research design, anthracycline type, and significant associations, is reported in **Table 1.** Studies are described below in the sections 3.1 to 3.7 focusing on the genes coding for transporters, metabolizing and antioxidative enzymes as well as proteins involved in cardiac function. Most of the studies used a candidate gene approach[32,64-74], two variants were nevertheless identified through genome-wide association studies (GWAS)[74,75]

6.5.1. Membrane transporters.

ATP Binding Cassette (ABC)

ABCB1, ABCB4, ABCC1, ABCC2, ABCC5

Polymorphisms in genes coding for ABC transporters were correlated with higher incidence of ACT in childhood cancer patients treated with anthracyclines[54,65,70]. In addition to its role as efflux transporters, the ABCC subfamily is involved in endothelial homeostasis (for example, in the regulation of vascular tone and systemic blood pressure) and may have important effects in the myocardium through the efflux of cyclic nucleotides, in particular through the depletion of cellular cyclic adenosine monophosphate[76].

Eight variants in five genes *(ABCB1, ABCB4, ABCC1, ABCC2, ABCC5)* were identified as predictors of ACT[65,69,70,72]. Some studies reported validation in internal replication cohorts. For example, *Krajinovic et al.* reported that survivors of childhood acute lymphoblastic leukemia (ALL) with *ABCC5* rs7627754 TT genotype had significant reductions in ejection fraction (EF) and fractional shortening (FS)[65]; *Visscher et al.* reported two gender-dependent associations in *ABCB4* gene (rs1149222 and rs4148808) that appear to be significant only among females[77]; in addition, variants in *ABCB1* (rs2235047) and *ABCC1* (rs4148350) were also associated with higher

risk of ACT in patients with various pediatric cancer[72,77]. Two other associations in *ABCC1* gene (rs3743527 and rs246221) were identified by *Semsei et al*. in relation to anthracycline-induced LV dysfunction in childhood ALL patients[70]. Likewise, *ABCC2* rs3740066 GG genotype had lower FS during both acute phase of therapy and 5–10 years after treatment in pediatric ALL and sarcoma patients[69]. The last two studies did not have an internal replication cohort and further verifications are necessary.

The *ABCC1* is one of the most studied transporter-coding genes in the context of cancer treatment and cardiotoxicity. It was the first described in a doxorubicin-resistant cell line and is expressed in different tissues including myocardium[59,78]. The reduced activity may exacerbate an intracellular anthracyclines' accumulation and subsequently results in cellular toxicity[79], while higher *ABCC1* activity could then explain lower drug levels and reduced toxicity. Furthermore, human *ABCC1* is expressed not only on the plasma membrane but is also found in the cytoplasm. It has been hypothesized that such expression can allow the sequestration of *doxorubicin* in lysosomes, without even reaching its target, the nucleus[54,78].

Interestingly, rs4148350 variant in *ABCC1* gene associated with ACT in pediatric patients[72,75,77] is located within the same haplotype block as nonsynonymous variant Gly671Val reported in adults with non-Hodgkin lymphoma[54]; suggesting the linkage disequilibrium (LD) between these SNPs and similar involvement of some of the variants in ACT following treatment of both pediatric and adult cancers[72].

Solute carriers (SLC)

SLC10A2, SLC22A2, SLC22A6, SLC22A7, SLC22A17, SLC28A1, SLC28A3

Given essential role the SLCs play in the absorption and excretion of drugs, considerable knowledge has been accumulated on SLC subfamilies, mainly expressed in the kidneys, liver, and intestine[80,81]. The association with ACT was reported for SNPs located in several SLCs. The

protective role of minor alleles of rs7853758 and rs4877847 in *SLC28A3* gene was identified and replicated in at least 2 pediatric cancer cohorts[71,72,77]. Interestingly, the association of rs4877847 remained significant after conditioning on rs7853758, suggesting an independent effect of the former[72]. Moreover, rs7853758 was recently recommended for pharmacogenomic testing in childhood cancer patients that received doxorubicin or daunorubicin (Level B – moderate)[82]. The association between *SLC28A3* rs7853758 and ACT seems to be more specific for childhood cancer patients since it was not replicated in adult cohorts of breast cancer[83] and aggressive B-cell lymphoma[84]. However, it is worth noting that certain inconsistencies were reported for rs4877847, since minor A allele of, was found more frequent by *Ságiet al.* among ALL and osteosarcoma patients with ACT[69].

Regarding other SLC genes, *Visscher et al.* identified and replicated rs4982753 and rs4149178 SNPs in *SLC22A17* and *SLC22A7* genes, respectively, as potential protective markers of ACT in childhood cancer patients[71]. The same group reported through different studies additional protective variants in *SLC10A2* (rs9514091 and rs7319981*)* and *SLC22A2* (rs316019); interestingly, the effect of the latter was age-related and was observed in younger patients (younger than 5 years)[72,77]. Both risk and protective variant (rs2305364 and rs2290271, respectively) were also reported in *SLC28A1* gene[72,75,77]. Finally, the *SLC22A6* rs6591722 rare AA genotype was associated with lower mean FS 5–10 years after the diagnosis in patients treated for childhood ALL[69].

The *SLC28A1* gene encodes a pyrimidine-preferred nucleoside transporter and is considered to be the best candidate to mediate uptake of a variety of fluoropyrimidines currently used in the treatment of solid tumors[85,86], while *SLC28A3* gene encodes a carrier with a wider range of

substrates[87]. Remarkably, in human cultured leukemia cells (HL60), nucleoside transporters are involved in the uptake of anthracyclines, including doxorubicin, but not daunorubicin[88].

The *SLC22A7* transporter has been increasingly recognized in terms of its role in drug disposition[89]. The well-characterized drug substrates of *SLC22A7* include various antibiotics[89], antivirals[90] and antimetabolites (5-fluoro-uracil and methotrexate)[91]. These transporters are widely expressed in different tissues, including myocardium[89], therefore it would be fair to assume that the proteins encoded by this gene could transport anthracyclines into the cell and that variable intake may either protect against cardiotoxicity or lead to undesirable effects[71].

The *SLC22A6* gene is involved in renal excretion of endogenous and exogenous organic anions and mediate excretion of many drugs including methotrexate[92]. In an in vitro study, adverse cardiac effects of indoxyl sulfate (non-dialysable uremic toxin, that has direct pro-hypertrophic and profibrotic effects on cardiac myocytes and fibroblasts[93]) were attenuated by *SLC22A6*[94]. This gene was for the first time mentioned in the context of anthracycline-induced cardiotoxicity or cardiac function by *Sagi et al.*[69].

The specific role of other transporters such as *SLC22A17, SLC10A2 and SLC22A2* (mainly expressed in the brain[95], in the liver and intestines[96] and kidney[97], respectively) in relation to ACT requires further investigation.

6.5.2. Anthracyclines metabolizing enzymes.

Hydroxylation

Carbonyl reductases (CBR)

CBR3

Several studies confirmed that CBR-mediated two-electron reduction of the *C-13* carbonyl group in the anthracycline side chain is one of the main metabolic pathways involved in the pathogenesis of cardiotoxicity because it results in the formation of cardiotoxic alcohol metabolites

(doxorubicinol, daunorubicinol) in the myocardium[32,37,98]. *Blanco et al.,* examined modifying effect of two functional polymorphisms, *CBR1* 1066G>A and *CBR3* V244M, on the dose dependent risk of ACT in patients diagnosed with various childhood cancers[32]. They found that among homozygotes for the minor G allele of *CBR3* rs1056892 (V244M), exposure to low-to-moderate doses of anthracyclines increased the risk of ACT by three times compared to other genotypes[32]. However, the other groups that studied the effect of *CBR3* rs1056892 could not replicate its association with ACT[71,72].

One Electron Reduction

Nitric Oxide Synthases (NOS)

NOS3

NOS3 is an endothelial enzyme that catalyzes the production of nitric oxide (NO). This enzyme is involved in the regulation of vascular tone, cellular proliferation, leukocyte adhesion, and platelet aggregation[99]. In addition, *NOS3* participates in the one-electron reduction of anthracyclines[32]. The rs1799983 SNP in *NOS3* gene results in an amino-acid substitution that alters susceptibility to cleavage and decreases endothelial NO [65]. A cardioprotective effect of this polymorphism was observed in high-risk ALL patients who received a higher cumulative dose of doxorubicin [65].

NADPH Dehydrogenase

NQO1

NQO1 is implicated in both one- and two-electron reduction of anthracyclines. *NQO1* reduces quinone to hydroquinone conversion and prevents oxidative stress, potentially playing a role in defense against pro-oxidant drugs[53,100]. *Sági et al.* reported an association between rs1043470 polymorphism in this gene and reduced cardiac function in childhood ALL and osteosarcoma patients[69]. The evidence of the *NQO1* role in ACT risk is nevertheless controversial[32,101], and requires additional validation.

6.5.3. Other enzymes with potential effect on the metabolism of anthracyclines

Alcohol Dehydrogenase (ADH)

ADH7

Alcohol dehydrogenase (ADH) catalyzes the interconversion between ethanol and acetaldehyde, the latter is metabolized to acetate by aldehyde dehydrogenase (ALDH)[102]. The genetic variation of ADH and ALDH can modify the kinetic properties of enzymes, creating abnormal accumulation of acetaldehyde[103], which in turn, can generate adducts with DNA and proteins[102,103]. The ADH polymorphisms are potentially implicated in induction of oxidative stress and apoptosis through the activation of stress signaling[102,104,105], which may induce myocardial hypertrophy[106]. Moreover, it was suggested that mitochondrial aldehyde dehydrogenase (ALDH2) exerts a protective effect against doxorubicin-induced cardiotoxicity[107]. *Visscher et al.*[72] reported a variant rs729147 in ADH7 gene linked to an increased risk of ACT.

Cytochrome P450 (CYP)

CYP3A5 and CYP4F11

Cytochrome P450 (CYP) enzymes located in the heart can influence both drug metabolism and endogenous cellular function[108]. The importance of CYPs in cardiovascular physiology stems from their ability to metabolize arachidonic acid (starting material in the synthesis of essential substances such as prostaglandins and leukotrienes[109]) into epoxyeicosatrienoic acids (EETs)[110]. The cardioprotective effect of EETs has been demonstrated in ischemia/reperfusion injury[111], cardiac hypertrophy and in cardiotoxicity induced by doxorubicin[112].

Among several investigated CYP polymorphisms, only rs2108623 in *CYP4F11* gene was reported associated with an increased risk of ACT in two independent cohorts of childhood cancer survivors[72,77]. The risk of pathological FS was significantly affected by rs4646450 in the *CYP3A5* gene in childhood ALL and osteosarcoma patients[69], this association needs nevertheless further validation.

6.5.4. Other medication neutralizing enzymes

Glutathione S-transferase (GST)

GSTP1

Glutathione S-transferases *(GSTs)* catalyze the detoxification of many endogenous or exogenous compounds (including anthracyclines). They also reduce the oxidative damage produced by these compounds; their expression is induced by the ROS generation[113]. The polymorphisms of *GSTP1* (rs1138272 and rs1695) have been correlated with a less active enzyme, increasing exposure to the drug and reducing the clearance of anthracyclines[114]. The association between rs1695 *GSTP1* and cardiotoxicity, measured by reduced ejection fraction was reported by *Windsor et al.*[115].

Sulfotransferase (SULT)

SULT2B1

Other enzymes related to oxidative stress include sulfotransferase (SULT) responsible for sulfate conjugation of anthracyclines. Polymorphisms in this gene were correlated to ACT in pediatric cancer patients[72,75,77]. Gender specific protective effect, seen only in males, was reported for intronic variant rs10426377 in *SULT2B1* gene[77].

UDP-glucuronosyltransferase (UGT)

UGT1A6

UGT1A6 is an important player in drug detoxification (including anthracyclines metabolites) via glucuronidation pathway[77]. Noteworthy, a synonymous SNP rs17863783 has been associated with an increased risk of ACT in three independent cohorts of pediatric cancer patients with leukemia and lymphoma, sarcoma, and other malignancies[71,72,77,82]. Moreover, this functional variant was recommended for pharmacogenomic testing in childhood cancer patients treated with doxorubicin or daunorubicin therapy[82]. Additionally, an increased risk of ACT associated with functional variant rs6759892 in *UGT1A6* was also shown[72,77].

6.5.5. Other antioxidative Enzymes, Catalase (CAT)

CAT

There is only one study conducted in survivors of childhood ALL patients that showed an association between intronic rs10836235 variant in the *CAT* gene and the risk of developing cardiac toxicity in patients exposed to anthracyclines[67]. Interestingly, other study conducted in childhood cancer patients[72] examined 19 other variants of this gene, but did not confirm their role in ACT development.

6.5.6. Genes involved in cardiac function

RARG, HAS3, CELF4, PLCE1, ATP2B1, HNMT, GPR35

A recent genome-wide association study (GWAS), conducted by *Aminkeng et al.*, identified a new gene associated with ACT in survivors of childhood cancer composed of ALL, Ewing's sarcoma, and rhabdomyosarcoma[75]. The study design included a discovery phase followed by replication in two independent survivor cohorts. A meta-analysis of all study populations using logistic regression adjusted for age, dose, radiotherapy to the heart and type of malignancy consistently reproduce the association between the non-synonymous coding variant rs2229774 in the *RARG* (Retinoic Acid Receptor Gamma) gene and the risk of ACT[75,82]. Overall, carriers for the minor A allele of rs2229774 had significantly increased odds of developing ACT compared to non-carriers[75]. The functional studies conducted by the same team confirmed that the presence of the non-synonymous rs2229774 variant (p.Ser427Leu) alters the function of *RARG*, which in turn affects the repression of Top2b[37,41]. This resulted into higher Top2b levels, conferring increased susceptibility of the cells to the cytotoxic effect of anthracyclines[37,75]. This association is considered as one of the most convincing and consistent across several cohorts. It is recommended as a marker for pharmacogenomic testing in childhood cancer patients receiving doxorubicin or daunorubicin therapy[82].

Wang et al, used a matched case-control study design, to explore the relation between ACT and polymorphisms in *Hyaluronic Acid Synthase 3 (HAS3)* gene[73]. They identified a common SNP, rs2232228, that exerts a modifying effect on anthracycline dose-dependent cardiomyopathy risk. Among pediatric cancer survivors with rs2232228 GG genotype, cardiomyopathy was infrequent and not dose-related, while in individuals exposed to high-dose anthracyclines, the rs2232228 AA genotype conferred an increased ACT risk. This finding was successfully replicated in an independent cohort of patients with ACT. In order to functionally explain the reported association this team suggested that Hyaluronan, a glycosaminoglycan abundantly found in the extracellular matrix plays an important role during tissue remodeling following injury[73,116]. Hyaluronan is also involved in reducing ROS–mediated cardiac toxicity[116]. Therefore, high risk rs2232228 AA genotype could be related to inappropriate remodeling and/or inadequate protection of the cardiomyocytes from ROS-induced injury during high-dose anthracyclines exposure[73].

While conducting a subsequent GWAS in childhood cancer survivors, the same research group (*Wang et al*) reported a dose-dependent association of rs1786814 in the *CELF4 (CUGBP Elav-Like Family Member 4)* gene[74]. The multivariable analyses adjusted for age at cancer diagnosis, sex, anthracyclines dose and chest radiation, disclosed that in carriers of A allele, ACT was not frequent and did not depend on dose, whereas carriers of GG genotype, exposed to higher doses of anthracyclines (greater than 300 mg/m2) were more prone to the risk of ACT. Recent evidence has emerged showing that variations in the *CELF4* gene are implicated in alternative splicing of *TNNT2*[74], the gene that encodes for cardiac troponin T (cTnT) and is known as a biomarker of myocardial injury[117].

The hypertension-susceptibility loci of the *PLCE1* (*Phospholipase C Epsilon 1*) and *ATP2B1* (*ATPase Plasma Membrane Ca2+ Transporting 1*) genes (rs9327264 and rs17249754

respectively) were reported by *Hildebrandt et al.*[64] to be significantly associated with ACT risk in a cohort of survivors of various pediatric cancers both conferring a protective effect. These findings require further validation and replication.

The *Histamine N-methyltransferase,* encoded by *HNMT* gene, catalyzes the N-methylation of histamine. Moreover, in an animal model of hypotension, central inhibition of *HNMT* activates the histaminergic system and induces the mobilization of compensatory cardiovascular mechanisms. The pathophysiological functions of cardiac histamine levels and related histamine receptors were previously investigated in the context of the development of chronic heart failure[118,119], suggesting the possible involvement of the *HNMT* gene as a modulator of cardiotoxicity[120]. An age-related effect of rs17583889 and rs17645700 in this gene was suggested, since the significant association with increased ACT was detected only in younger patients (younger than 5 years)[77]. Almost all associations previously reported in this review (except one in *SLC22A6* gene) focused on common polymorphisms with a frequency of minor alleles (MAF) >5%. Alternatively, *Ruiz-Pinto et al*. investigated the implication of low-frequency variants (MAF<5%) in the susceptibility to ACT in pediatric cancer patients[68]. They carried out a GWAS using the *Illumina HumanExome BeadChip* array, enriched with low-frequency coding variants. Through gene-based testing, they explored the combined effects of common and low-frequency variants and identified a novel significant association between *GPR35 (G Protein-Coupled Receptor 35)* gene and increased risk of ACT. A missense variant rs12468485[68] had the greatest contribution to the association signal. The *GPR35* gene belongs to the G protein-coupled receptor family of membrane proteins mediating a wide range of physiological processes with potential role in cardiac physiology and pathology[121]. This association lacks replication and will require future studies.

6.5.7. Iron metabolism

HFE

Lipshultz et al.[66] explored the frequency of hemochromatosis gene (*HFE*) mutations associated with hereditary hemochromatosis (a genetic disorder with symptoms of a pathological increase in iron levels) and their relationship with doxorubicin-associated ACT in survivors of childhood high-risk ALL, suggesting that carriers of rs1800562 *HFE* variant were more prone to heart injury[66]. Interestingly, *Armenian et al.* reported that the congestive heart failure in survivors of hematopoietic cell transplantation was associated with *HFE* rs1799945 variant, but not with rs1800562; these results were reported in combined cohort of pediatric and adult patients[122].

6.5.8. Other function

SPG7, FMO3, FMO2

Mitochondrial dysfunction plays an important role in the pathogenesis of multiple cardiac diseases[123]. The oxidative stress increased mitochondrial Ca(2+) and ROS generation act synergistically to produce the mitochondrial permeability transition pore (mPTP) and cell death[124,125]. Therefore, the inhibition of the mPTP appears to be protective in cardiac diseases, including cardiomyopathies[123,125]. It was recently confirmed that ubiquitously expressed inner mitochondrial membrane integral protein, SPG7 (encoded by *SPG7* gene), is a core component[124] or modulator of the mPTP[126].

The flavin-containing monooxygenases (FMO) are NADPH-dependent flavoenzymes catalyze the oxidation of numerous drugs and xenobiotics. Several recent studies have found a link between either FMO3 (the isoform 2 encoded by *FMO2* gene represents the truncated form without catalytic activity[127]) or its enzymatic product, trimethylamine *N*-oxide, with cardiovascular disease[128-130]. With the respect to ACT in childhood cancers, protective variants rs2019604, rs2020870 and rs1736557 in the *SPG7, FMO2* and *FMO3* genes are reported, respectively[72,77] .

6.6. MODIFYING EFFECT OF REPORTED PHARMACOGENES IN THE PETALE COHORT

To further explore the described ACT-associated genes, we pooled the information on genetic variants from these genes using whole exome sequencing (WES) data available from the well described PETALE cohort of childhood ALL survivors of the Sainte-Justine University Health Center (SJUHC)[9] and analyzed whether they correlate with ACT development. The analyses comprised a set of genes identified through reviewed studies (**Table 1**) that were also validated in an independent replication cohort.

The characteristics of the PETALE cohort were previously described[9,131,132]; briefly this cohort is composed of 236 patients diagnosed and treated for childhood ALL according to Dana Farber Cancer Institute (DFCI) ALL 87-01 to 05-01 protocols at SJUHC, Montreal, (Quebec), Canada. Eligible participants were younger than 19 years old at diagnosis, at least 5 years after diagnosis, without history of relapse or refractory ALL and had not received a hematopoietic stem cell transplant. The median age of patients at the time of diagnosis was 9 years and the time from end of treatment to evaluation ranged from 3-24 years with a median of 12 years. The analyses were performed in either all patients or subgroups that differ relative to doxorubicin treatment. These included patients assigned to standard (SR, 45.8%) and high risk (HR, 54.2%) groups based on prognostic factors, who received different cumulative doxorubicin doses[9] and HR patients who received or not cardioprotectant (53.1% and 46.9%, among HR patients, respectively)[9]. They were almost exclusively of reported French Canadian descent (>95%). Details are outlined in **Supplemental Table 1**.

The genotypes of 108 common and 381 rare variants of 22 genes were pooled from WES data[9,131,132] followed by association studies with cardiac complications, as defined by the left LVEF, FS and LVEDD (left ventricular end-diastolic diameter) evaluated by echocardiography[9] in PETALE participants.

Even if we did not necessarily test the same lead SNPs (unless the SNPs were present in WES data) as in original studies, almost all genes (21/22) had at least one of the variants significantly associated with tested outcomes. Common variants with risk effect were detected in *ABCB1, ABCC1, ABCC5, ADH7, CYP4F11, FMO2, FMO3, HNMT, NOS3, RARG, SLC10A2, SLC22A2, SLC22A7, SLC28A1, SPG7, SULT2B1* and *UGT1A6* genes (**Table 2**) and with protective effect in *ABCB1, ABCB4, ABCC1, ABCC5, ADH7, CYP4F11, FMO2, FMO3, NOS3, SLC10A2, SLC22A2, SLC28A1* and *SLC28A3* genes (**Table 3**). The significant effect of rare variants was noted in *ABCC5, ADH7, CELF4, HAS3, HNMT, NOS3, SLC10A2, SLC22A2* and *SLC28A1* genes (**Table 4**). These analyses confirmed the importance of the genes identifying through prior studies and, in some cases, pointed to potentially causal variants. It should be noted that, given many variants tested, only few of them had false discovery rate, FDR[133,134], lower or equal to 5%. Among such variants, are common SNPs in *ABCC1, SLC22A7* and *SPG7* genes. A risk effect of minor G allele of *ABCC1* rs246232 was observed in all patients, as based on LVEDD (p=0.0005) while the protective effect of minor A allele of *ABCC1* rs2230671 was seen in SR patients, as based on LVEF (p=0.0009). Likewise, carriers for the minor A allele of *SLC22A7* rs70953680 and SPG7 rs66845605 who were assigned to SR group, were at higher risk of having lower LVEF (OR=4.9; 95% CI, 1.99-11.8; p=0.0002 and OR=4.2; 95% CI, 1.7-10.3; p=0.001, respectively) (**Supplemental table 2**). Interestingly, the identified *SPG7* rs66845605 variant is in a LD with initially reported *SPG7* rs2019604 variant (r2=0.73, D'=0.94)[72,77].

The results obtained here are adding further weight to *ABCC1* gene, which seems to play an important role in ACT in pediatric cancer patients, as described in the section 3.1.1.[70,72,77]. Recent evidence suggests that *SLC22A7*, which is widely expressed in different tissues including cardiac

tissue[135,136], is involved not only in the transport of naturally occurring nucleosides[135], but is also linked to the transport of various drugs[80,91,136].

The *SPG7* gene is highly implicated in mitochondrial function, more precisely, it is one of the key regulators of the mitochondrial permeability transition pore (PTP), the opening of which leads to a bioenergetic crisis and to cell death[123,125].

The analysis of rare variants led to the detection of a risk effect between LVEF (%) and rare variants enrichment in *SLC10A2* (p=0.002) and *ADH7* (p=0.001) genes (**Supplemental table 3**). *SLC10A2* is expressed at tissue sites that are involved in the enterohepatic circulation of bile acids[137]. *Dawson et al.*[137] suggested that certain drugs and/or drug metabolites, may interact as inhibitors of *SLC10A2* [138], which in turn may explain some of the side effects associated with these drugs[137]. However, possible interaction between *SLC10A2* and anthracyclines remains unclear.

The polymorphisms of alcohol dehydrogenase (ADH) are potentially implicated in induction of oxidative stress and apoptosis through the activation of stress signaling[102,104,105], which may in turn induce myocardial hypertrophy[106]. Different alleles of *ADH7* gene were associated with the early stages of alcohol metabolism, alcohol-induced cardiac dysfunction, and insulin resistance[75,105,139]. Interestingly, the effect of minor alleles in *ADH1* gene was linked to lower rate of complete remission in AML patients[36]. Additional studies will be needed to elucidate the role of *ADH7* gene in the metabolism of anthracyclines and/or its role in ACT.

Several lead SNPs identified through described studies were also present in our dataset. Many of these markers were also associated with ACT in PETALE cohort (indicated in **Tables 2 and 3**), among them, are associations in *FMO2, FMO3, NOS3 and UGT1A6* genes. In addition, *SLC28A1* rs2290272 variant was identified in the PETALE cohort as a risk marker, which is in a strong LD with initially reported rs22900271.

Furthermore, in the PETALE cohort, minor allele of *RARG* rs2229774 was associated with an increased susceptibility to ACT in standard-risk patients (**Table 2**). This SNP is included in the pharmacogenomic panel for testing childhood cancer patients who receive doxorubicin or daunorubicin[82].

These associations provide further evidence of genetic contribution to the ACT in childhood ALL survivors.

6.7. CONCLUSION

This systematic review provides an insight into pharmacogenomic markers related to ACT in patients treated for childhood cancer. We also included a brief overview of the applicability of reported findings to the well established PETALE cohort.

An important number of genes have been identified through association studies of ACT. These findings may altogether lead to prediction models to identify patients who might be highly susceptible to ACT and require treatment adjustment or closer follow-up or have lower susceptibility to ACT, therefore may not need the cardio protectants and/or may not require additional follow-ups.

However, in some cases, replication analyses failed to validate initial findings. This discordance can be attributed to several factors, such as the heterogeneity between cohorts, the inconsistency in the definition of ACT, the differences among protocols, the type, and doses of anthracyclines, among many others.

Given certain limitations of the current pharmacogenomic evidence, we emphasize that new, well designed prospective studies on larger and well-defined populations are needed to accurately validate the predictive value of reported genetic biomarkers associated with ACT.

As more reliable data become available, healthcare providers will be able to integrate the results of pharmacogenomic testing into clinical therapeutic and follow-up decisions in order to detect the individuals that are at higher risk to have treatment-related cardiotoxicity, thus, to maximize treatment's efficacy and reduce its long-term effects in the particularly vulnerable population of pediatric patients.

6.8. EXPERT OPINION

Anthracycline-induced cardiotoxicity (ACT) occurs in almost 60% of treated patients and remains an important limitation of anthracycline-based chemotherapy[82]. The struggle to mitigate ACT has given rise to several new preventative strategies currently used in clinical practice, such as limitation of the dose exposure; encapsulation of anthracyclines in liposomes to reduce myocardial uptake[38]; modification of anthracycline structure in an effort to reduce myocardial toxicity; administration concurrently with the iron chelator dexrazoxane to reduce free iron-catalyzed ROS formation[8,34,35]. Despite advances in supportive and protective therapy for myocardial function, serious adverse effects of anthracyclines, like congestive heart failure are still causing major clinical problems.

The wide variation in interindividual susceptibility to ACT along with a multitude of genetic variants detected through association studies suggests that genetic contribution should be taken into consideration for the design of new individualized preventive approaches.

A major drawback of the current genetic studies is inconsistency across different populations and treatment protocols. Further validation is needed in the context of well-defined interacting and confounding factors such as, among others, treatment protocol, and disease. In that regard the gene-treatment interactions are important for the implementation of pharmacogenetics findings since the effect of polymorphisms could be related to the type of malignancy, the dose of the drug, and the duration of treatment, which can vary considerably across protocols. Other factors that are intrinsic to each study and which can affect the interpretation of the outcome include the type of drug, the definition of the patient risk group, definition of studied phenotypes, time of evaluation relative to the end of treatment, frequency of follow up, nutrition, life habits, and prophylactic treatment.

One of the recent publications on ACT includes evidence-based clinical practice recommendations for pharmacogenomic testing. *Aminkeng et al.*[82] pointed to the polymorphisms rs7853758, rs2229774 and rs17863783 in *SLC28A3, RARG and UGT1A6* genes, respectively, as genetic markers with the strongest association with ACT in patients with various childhood cancers. Interestingly, our results conducted in leukemia patients treated with doxorubicin replicated one of the findings, the risk variant rs2229774 in the *RARG* gene, in a subgroup of patients, suggesting that this variant warrant validation through a prospective study. Variability across studies demonstrates the challenge facing the clinical implementation of pharmacogenomics and the continuous need to critically analyze rapidly emerging data.

In conclusion, pharmacogenomics has a great potential to improve the use of medications to enhance efficacy and reduce toxicity by allowing for optimal treatment selection and/or dose personalization based on the genetic characteristics of individuals; it may also provide an improved ability to detect the likelihood of long-term adverse effects of the cancer treatments. It is worth mentioning that the establishment of optimal treatment approaches using genetic information in conjunction with data on non-genetic causes of inter-patient variability in drug responses could be much more challenging than the detection of genomic variability itself. Furthermore, the implementation of pharmacogenetic testing often relies on appropriate algorithms and adequate recommendations. Therefore, clinical implementation will need consensus guidelines that could help health professionals to interpret the results of testing and guide their decision-making process.

6.9. DECLARATIONS

Ethics approval and consent to participate

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of SJUHC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon request.

Funding

Institute of Cancer Research (ICR) of the Canadian Institutes of Health Research (CIHR) ***grant number: 118694***, Cancer Research Society (CRS), Canadian Cancer Society Research Institute (CCSRI), Ontario Institute for Cancer Research (OICR), Pediatric Oncology Group of Ontario (POGO), Garron Family Cancer Centre at SickKids Hospital (Ontario), The Terry Fox foundation, FRQS Applied Medical Genetics Network and Sainte-Justine Hospital Foundation supported the PETALE study.

Kateryna Petrykey is a scholar of the Cole Foundation.

Competing interests

The authors declare no competing financial interests.

Acknowledgements

The authors would like to thank all childhood ALL survivors and their parents for the participation in the PETALE study.

6.10. TABLES AND FIGURES

Table 1. Studies investigating genotypic associations with anthracycline-induced cardiotoxicity in childhood cancer patients.

SNP	Gene	Study approach	Study cohort	ACT definition	Anthracycline type	Follow-up, years, mean or median (range)	Author	Replication
Protective variant	*CAT*	**Candidate genes:** genes that encode ROS metabolizing enzymes *SOD2* **and** *CAT; glutathione S-transferases* **GSTT1** *and* **GSTM1**	Long-term survivals of childhood ALL treated with BFM protocols	**LVEDD, LVESD**, septal and posterior wall thickness in diastole, **LVEF, FS**.	Not specified	Every 6 months for the first 3 years and yearly afterwards.	Rajic et *al.*, 2009 (1)	**no**
rs10836235								
Risk variant	*GSTP1*	**Candidate genes:** 36 candidate SNPs in triple-drug **MAP** (**methotrexate, doxorubicin, cisplatin**) pathway genes	Osteosarcoma pediatric patients	**LVEF**	**Doxorubicin**	**Months** 41.5 (12-93)	Windsor et *al.*, 2012 (2)	**no**
rs1695								
Risk variant	*CBR3*	**Candidate genes:** functional genetic polymorphisms in ***CBR1*** and ***CBR3*** genes	Childhood cancer survivors (treated for soft tissue sarcoma, leukemia, lymphoma, bone tumors, other)	**LVEF** and/or **FS**	**Doxorubicin**, **daunomycin**, **epirubicin** and **idarubicin**	**Cases**= 7.0 (0.1-35.1), **controls** = 11.2 (0.4-40.3)	Blanco et *al.*, 2012 (3)	**no**
rs1056892 (V244M)								
Risk variants	*ABCC1*	**Candidate gene:** SNPs of the ***ABCC1*** gene	Paediatric patients with ALL treated with anthracycline chemotherapy.	**LV** function alteration measured by **LVEDD** and **LVESD**, **FS**.	**Doxorubicin, Daunorubicin**	End of the treatment 2.0 (0.7–3.4); Latest ECHO 6.3 (2.4–13.7).	Semsei et *al.*, 2012 (4)	**no**
rs246221								
rs3743527								

Risk variants		**Candidate genes: 2977** SNPs in **220** key drug biotransformation genes	Patients treated for **childhood cancer** including AML, Other leukemia, HL/NHL, Osteosarcoma, Rhabdomyosarcoma, Ewing's sarcoma, Other sarcoma, Nephroblastoma, Hepatoblastoma, Neuroblastoma, Carcinoma.	Early- or late-onset **LV dysfunction** measured by **FS** and/or symptoms requiring intervention based on **CTCAEv3**	**Doxorubicin, Daunorubicin, other**	**Cases** = 6.8 (0.1-21.2); **controls** = 8.6 (5.0-18.6)	Visscher et al., 2012 (5) Visscher et *al.*, 2013 (6)	**yes**
rs6759892	*UGT1A6*							
rs17863783								
rs4261716								
rs1149222	*ABCB4*							
rs4148808								
rs4148350	*ABCC1*							
rs17583889	*HNMT*							
rs729147	*ADH7*							
rs2305364	*SLC28A1*							
rs2235047	*ABCB1*							
Protective variants								
rs7853758	*SLC28A3*							
rs4877847								
rs885004								
rs1736557	*FMO3*							
rs2019604	*SPG7*							
rs9514091	*SLC10A2*							
rs7319981								
rs17645700	*HNMT*							
rs2290271	*SLC28A1*							
rs316019	*SLC22A2*							
rs2108623	*CYP4F11*							
rs10426377	*SULT2B1*							
Risk variant	*HFE*	**Candidate gene:** two common ***HFE*** allelic variants: C282Y and H63D	Paediatric patients with high-risk ALL treated with the DFCI protocols	LV structure and function Z-scores.	**Doxorubicin**	Time from registration on original protocol to post-baseline echocardiogram, years, **median**	Lipshultz et *al.*, 2013 (7)	**no**
rs1800562								

						(range) 6.1 (1.0–16.1)		
Risk variant rs2232228	*HAS3*	**Candidate genes:** ITMAT/Broad CARe SNP array to profile common SNPs in **2100** genes relevant to de novo cardiovascular disease.	Childhood cancer survivors with and without cardiomyopathy treated for: ALL/AML, soft tissue sarcoma, HL/NHL, bone tumors, other.	Signs and symptoms and/or **LVEF** and **FS**.	Not specified	**Cases**= 10.0 (0.1-35.1), **controls** = 11.3 (0.9-41.0)	Wang et *al.*, 2014 (8)	**yes**
Protective variant rs4982753	*SLC22A17*	**Candidate gene: 4578** SNPs in more than **300** genes preselected for relevance in drug ADME, including genes relevant for ACT.	The discovery cohort comprised patients recruited from pediatric oncology units and long-term follow-up clinics across Canada.	Early- or late-onset **LV** dysfunction measured by **FS** and/or symptoms requiring intervention based on **CTCAEv3**	**Doxorubicin, Daunorubicin**	**Cases** = 6.8 (0.1-21.2); **controls** = 8.6 (5.0-18.6).	Visscher et *al.*, 2015 (9)	**yes**
Protective variant rs4149178	*SLC22A7*							
Risk variants rs10426628	*SULT2B1*							

Risk variant rs2229774	*RARG*	**GWAS discovery analysis:** 740K SNP assay	Patients treated for childhood cancer including AML, ALL, Other leukemia, HL/NHL, Osteosarcoma, Rhabdomyosarcoma, Ewing's sarcoma, Other sarcoma, Wilms tumor, Hepatoblastoma, Neuroblastoma.	**FS** or signs and symptoms of cardiac compromise requiring intervention based on **CTCAEv3**	**Doxorubicin**	**Cases**= 7.5 (2.5-15.5), **controls** = 9 (7-12)	Aminkeng et *al.*, 2015 (10)	**yes**
Risk variant rs1786814	*CELF4*	**GWAS**	Childhood cancer survivors with and without cardiomyopathy treated for: ALL/AML, sarcoma, HL/NHL, other.	Signs and symptoms and/or **LVEF** and **FS**.	Not specified	**Cases**= 9.4 (0.1-35.1), **controls** = 12.9 (1.4-41)	Wang et *al.*, 2016 (11)	**yes**
Risk variant rs7627754	*ABCC5*	**Candidate genes:** polymorphisms in the DOX metabolic and functional pathway, including genes coding for efflux transporters	Paediatric patients with ALL treated with the DFCI protocols	**LVEDD**, **LVED** posterior wall thickness, **FS**, **LVEF** and **LV** mass.	**Doxorubicin**	**Discovery** cohort, 8.4 (1.0 - 18.0); **Validation** cohort, 5.25 (3.1-9.3)	Krajinovic et *al.*, 2016 (12)	**yes**
Protective variant rs1799983	*NOS3*							

Protective variants		**Candidate genes:** 12 GWAS-identified hypertension-susceptibility loci	Long-term childhood cancer survivors (treated for sarcoma, leukemia, lymphoma, other)	**LVEF**	Standardized to **doxorubicin** equivalents	**Cases** = 21.2(11.2); **controls** = 15.7(7.6)	Hildebrandt et *al.*, 2017 (13)	**no**
rs9327264	*PLCE1*							
rs17249754	*ATP2B1*							
Risk variant	*GPR35*	**Candidate genes: SNP array** included 247870 variants (majority protein-altering), representing **multiple ethnicities** and **complex traits**	Anthracycline-treated pediatric cancer patients (Leukemia, osteosarcoma, Ewing sarcoma)	**LV** dysfunction assessed by **FS** or symptoms/signs.	**Doxorubicin, daunorubicin**, or **epirubicin**	**Cases**= 8.3 (1-24.1), **controls** = 10.5 (1-27.5)	Ruiz-Pinto et *al.*, 2017 (14)	**no**
rs12468485								
Risk variants		**Candidate genes: 70** single nucleotide polymorphisms (SNPs) in 26 genes coding for xenobiotic transporters and metabolizing enzymes	Pediatric ALL and OSC patients	**LVEF** and **FS**	Standardized to **doxorubicin** equivalents	Follow-up categories: 1) **at the diagnosis** (control); 2) **acute phase;** 3) **maintenance chemotherapy**; 4) **end of the treatment;** 5) up to **5 y.** after dx; 6) **5–10 y**. after dx; 7) **10–15 y.** after dx; 8) more than **15 y.** after dx.	Sági et *al.,* 2018 (15)	**no**
rs3740066	*ABCC2*							
rs4646450	*CYP3A5*							
rs1043470	*NQO1*							
rs6591722	*SLC22A6*							

rs7853758	*SLC28A3*							

LVEDD: left ventricle end diastolic diameter; LVESD: left ventricle end systolic diameter; LVEF: left ventricle ejection fraction; FS: fractional shortening; LV: left ventricle; SNP: single nucleotide polymorphism; BFM: Berlin-Frankfurt-Munster protocol; ALL: acute lymphoblastic leukemia; ECHO: Echocardiogram; CTCAEv3: Common Terminology Criteria for Adverse Events version 3; AML: acute myelogenous leukemia; HL: Hodgkin's Lymphoma; NHL: Non-Hodgkin's Lymphoma; DFCI: Dana-Farber Cancer Institute; GWAS: genome-wide association study.

Table 2. Summary of the common variants' analysis in the PETALE cohort (N=236), risk associations*.

Gene	Fractional Shortening	LVEF-Mmode	LVEF-Mmode binary	LVEF 2D	LVEF 2D binary	LVEDD z score
ABCB1	**rs2229109****, p=0.036			rs2235035, p=0.026	rs2235035, p=0.009	rs2235035, p=0.043
					rs2235035, p=0.019	
ABCC1				rs111263400, p=0.028; rs4238623, p=0.038		rs246232, p=0.0005; rs903880, p=0.0052; rs3765129, p=0.024
						rs246232, p=0.008; rs3765129, p=0.006
						rs45511401**, p=0.025;
						rs246232, p=0.017
ABCC5			rs6443924, p=0.049			
ADH7	rs17537595, p=0.046	rs17537595, p=0.014				
CYP4F11						rs1064796, p=0.013
FMO2	**rs2020870****, p=0.021	**rs2020870****, p=0.027	**rs2020870****, p=0.006			
	rs2020870**, p=0.00095	**rs2020870****, p=0.002				
FMO3	rs2066528, p=0.009; **rs2266780****, p=0.013; rs1920149, p=0.042	**rs1736557****, p=0.046	rs1920149, p=0.045; **rs2266782****, p=0.045	**rs2266780****, p=0.020		
	rs1736557**, p=0.020	**rs2266782****, p=0.029; rs1920149, p=0.032	**rs1736557****, p=0.012	**rs2266780****, p=0.020		

	rs2066528, p=0.045; **rs2266780****, p=0.010; rs1920149, p=0.024; **rs2266782****, p=0.033		rs1920149, p=0.013; **rs2266782****, p=0.012			
	rs2266782**, p=0.026;		**rs2266780****, p=0.0490			
HNMT			rs3842548, p=0.040			
			rs3842548, p=0.016			
NOS3			rs891512, p=0.0498	rs891512, p=0.009	rs891512, p=0.018	
				rs891512, p=0.028	rs891512, p=0.019	
RARG			**rs2229774****, p=0.021			
SLC10A2	rs8000956, p=0.041; rs279940, p=0.050					
SLC22A2	rs316019, p=0.024		rs2774230, p=0.018			
			rs2774230, p=0.032			
SLC22A7	rs70953680, p=0.011	rs70953680, p=0.015	rs70953680, p=0.0002			
SLC28A1	rs4980349, p=0.036	rs8187763, p=0.049	**rs2242046****, p=0.037; rs8187763, p=0.037		**rs2242046****, p=0.025	rs4980349, p=0.027; **rs2290272****, p=0.037
					rs2242046**, p=0.010	rs3825875, p=0.029; **rs2290272****, p=0.0099
SPG7	**rs66845605**, p=0.021; rs461405, p=0.038	**rs66845605**, p=0.036	rs461405, p=0.026; **rs66845605**, p=0.040			
			rs66845605, p=0.001; **rs12960****, p=0.020			

SULT2B1		rs2302948, p=0.032				
UGT1A6		**rs6759892****, p=0.033	**rs6759892****, p=0.037	**rs6759892****, p=0.017	**rs6759892****, p=0.015	
		rs6759892**, p=0.019				

LVEF: left ventricular ejection fraction; LVEDD: Left Ventricular End Diastolic Diameter.
* results reflect association for minor/derived alleles as obtained through allelic ratio analyzes implemented in PLINK
**functional variants
Variants that were reported in the original studies or in LD with reported variants are in bold and highlighted in red

Color code:

All cohort
Standard risk group
High risk group
High risk patients who received dexrazoxane
High risk patients who did not receive dexrazoxane

Table 3. Summary of the common variants' analysis in the PETALE cohort (N=236), protective associations*.

Gene	Fractional Shortening	LVEF M-mode	LVEF M-mode binary	LVEF 2D	LVEF 2D binary	LVEDD z score
ABCB1				rs2235013, p=0.045		
				rs2235013, p=0.010; **rs2032582****, p=0.042	**rs2032582****, p=0.022	
ABCB4						rs31675, p=0.020
ABCC1	rs2230671, p=0.011			rs2230671, p=0.025	rs2230671, p=0.0009	
	rs35588, p=0.045	rs35588, p=0.034	rs35588, p=0.0054			
ABCC5	rs7636910, p=0.003; rs56348509, p=0.02	rs7636910, p=0.012				rs56373140, p=0.046
						rs56373140, p=0.032
ADH7	rs75076500, p=0.047					
	rs1154460, p=0.037	rs1154460, p=0.038				
	rs1154470, p=0.023					
CYP4F11			rs2305804, p=0.034	rs2305801, p=0.021; rs1064796, p=0.047	rs2305801, p=0.034	rs2305801, p=0.03
				rs2305801, p=0.017		rs2305804, p=0.048
				rs2305801, p=0.048		

FMO2	rs2020861, p=0.013; **rs2020863****, p=0.016	rs2020861, p=0.038; **rs2020862****, p=0.040		rs28969551, p=0.013		**rs2020863****, p=0.045
	rs2020861, p=0.005; **rs2020862****, p=0.012; **rs2020863****, p=0.017					
FMO3					rs143661234, p=0.044	
NOS3	**rs1799983****, p=0.038	**rs1799983****, p=0.031	**rs1799983****, p=0.01		**rs1799983****, p=0.033	**rs1799983****, p=0.026
SLC10A2						rs2301157, p=0.015
SLC22A2				rs2774230, p=0.0125	rs2774230, p=0.0125	
SLC28A1			rs2305367, p=0.0056	rs8025045, p=0.024	rs1562885, p=0.018	rs2290269, p=0.025
					rs1562885, p=0.012; rs2305366, p=0.013; rs3825875, p=0.044	rs2290269, p=0.044
SLC28A3	rs7867504, p=0.028					

LVEF: left ventricular ejection fraction; LVEDD: Left Ventricular End Diastolic Diameter.
* results reflect association for minor/derived alleles as obtained through allelic ratio analyzes implemented in PLINK
** functional variants
Variants that were reported in the original studies are in bold and highlighted in red

Color code:

All cohort
Standard risk group
High risk group
High risk patients who received dexrazoxane
High risk patients who did not receive dexrazoxane

Table 4. SKAT-O analysis of the rare variants in the PETALE cohort using WES data (N=236).

Gene	Fractional Shortening	LVEF M-mode	LVEF M-mode binary	LVEF 2D	LVEF 2D binary	LVEDD z score
ABCC5	p=0.0146	p=0.0105				
	p=0.0090	p=0.0302				
	p=0.0111					
ADH7		p=0.0440		p=0.0392	**p=0.0013**	
					p=0.0083	
					p=0.0184	
CELF4			p=0.0077			
HAS3	p=0.0445	p=0.0502			p=0.0289	
HNMT			p=0.0320	p=0.0101		p=0.0431
						p=0.0394
NOS3						p=0.0034
SLC10A2	p=0.0085	**p=0.0022**	p=0.0272			
	p=0.0062	p=0.0043	p=0.0050			
SLC22A2		p=0.0392				
		p=0.0299				
SLC28A1					p=0.0182	
					p=0.0115	

LVEF: left ventricle ejection fraction; LVEDD: left ventricle end diastolic diameter.
Top-ranking associations, those that were detected with the false discovery rate (FDR) ≤ 5%, are in bold.

Color code:

All cohort
Standard risk group
High risk group
High risk patients who received dexrazoxane
High risk patients who did not receive dexrazoxane

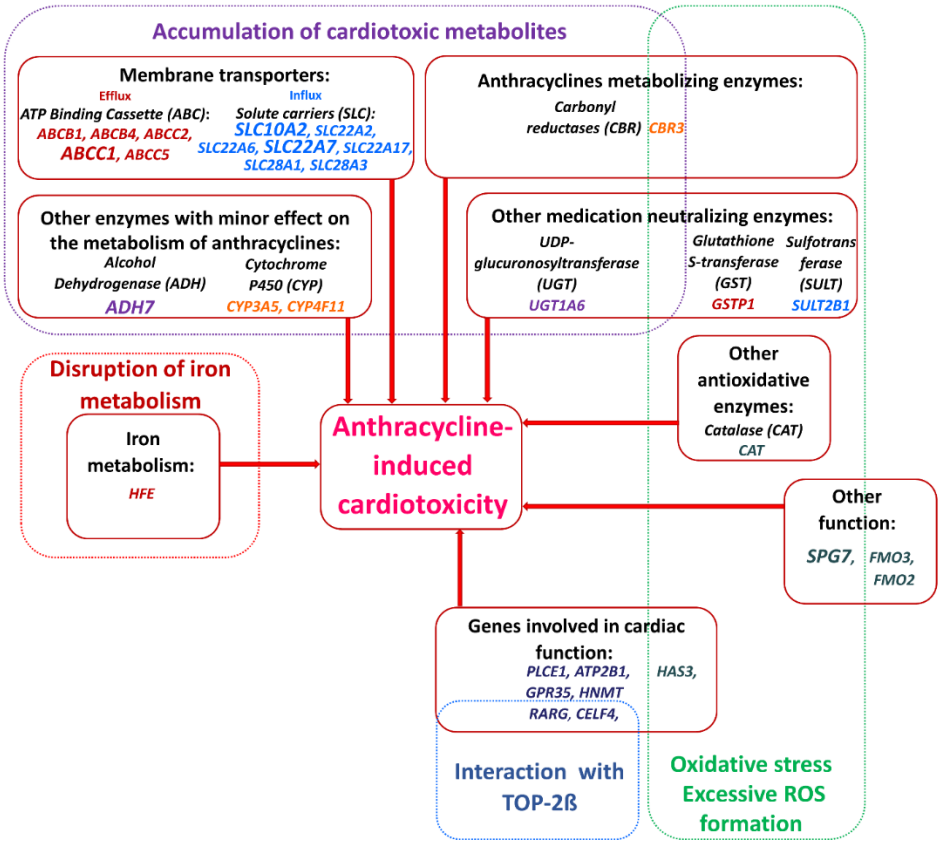


Figure 1. Summary of suggested pathophysiological mechanisms that mediate anthracycline-induced cardiotoxicity (ACT) and identified genetic variations associated with ACT in patients treated for childhood cancer.

The pharmacogenomic markers presented in this systematic review support 4 main pathophysiological mechanisms related to ACT:

a. Oxidative stress;

b. DNA damage through interference with topoisomerase II;

c. Accumulation of cardiotoxic metabolites;

d. Disruption of iron metabolism.

6.11. REFERENCES

1. Hudson MM, Ness KK, Gurney JG, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. *Jama.* 2013;309(22):2371-2381. • **article of interest, describes treatment-related late adverse effects in childhood cancer patients.**
2. Lipshultz SE, Alvarez JA, Scully RE. Anthracycline associated cardiotoxicity in survivors of childhood cancer. *Heart.* 2008;94(4):525-533.
3. Lipshultz SE, Karnik R, Sambatakos P, Franco VI, Ross SW, Miller TL. Anthracycline-related cardiotoxicity in childhood cancer survivors. *Curr Opin Cardiol.* 2014;29(1):103-112.
4. Franco VI, Lipshultz SE. Cardiac complications in childhood cancer survivors treated with anthracyclines. *Cardiol Young.* 2015;25 Suppl 2:107-116.
5. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians.* 2019;69(5):363-385.
6. Howlader N NA, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2016, National Cancer Institute. Bethesda, MD. April 2019.
7. Bhakta N, Liu Q, Ness KK, et al. The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *Lancet.* 2017;390(10112):2569-2582.
8. Lipshultz SE, Diamond MB, Franco VI, et al. Managing chemotherapy-related cardiotoxicity in survivors of childhood cancers. *Paediatr Drugs.* 2014;16(5):373-389. • **article of interest, describes mechanisms of anthracycline-induced cardiotoxicity**
9. Marcoux S, Drouin S, Laverdiere C, et al. The PETALE study: Late adverse effects and biomarkers in childhood acute lymphoblastic leukemia survivors. *Pediatr Blood Cancer.* 2017;64(6). •• **article of considerable interest, explains in details the whole study design of the PETALE cohort**
10. Chow EJ, Leger KJ, Bhatt NS, et al. Paediatric cardio-oncology: epidemiology, screening, prevention, and treatment. *Cardiovasc Res.* 2019;115(5):922-934.
11. Fulbright JM. Review of cardiotoxicity in pediatric cancer patients: during and after therapy. *Cardiol Res Pract.* 2011;2011:942090.
12. Yeh ET, Bickford CL. Cardiovascular complications of cancer therapy: incidence, pathogenesis, diagnosis, and management. *J Am Coll Cardiol.* 2009;53(24):2231-2247.
13. Hudson MM, Leisenring W, Stratton KK, et al. Increasing Cardiomyopathy Screening in At-Risk Adult Survivors of Pediatric Malignancies: A Randomized Controlled Trial. *Journal of Clinical Oncology.* 2014;32(35):3974-3981.
14. Lipshultz SE, Colan SD, Gelber RD, Perez-Atayde AR, Sallan SE, Sanders SP. Late cardiac effects of doxorubicin therapy for acute lymphoblastic leukemia in childhood. *N Engl J Med.* 1991;324(12):808-815.
15. Kremer LC, van Dalen EC, Offringa M, Voute PA. Frequency and risk factors of anthracycline-induced clinical heart failure in children: a systematic review. *Ann Oncol.* 2002;13(4):503-512.

16. Octavia Y, Tocchetti CG, Gabrielson KL, Janssens S, Crijns HJ, Moens AL. Doxorubicin-induced cardiomyopathy: from molecular mechanisms to therapeutic strategies. *J Mol Cell Cardiol.* 2012;52(6):1213-1225.
17. Olson RD, Mushlin PS. Doxorubicin cardiotoxicity: analysis of prevailing hypotheses. *FASEB J.* 1990;4(13):3076-3086.
18. Chow EJ, Antal Z, Constine LS, et al. New Agents, Emerging Late Effects, and the Development of Precision Survivorship. *J Clin Oncol.* 2018;36(21):2231-2240.
19. Barry E, Alvarez JA, Scully RE, Miller TL, Lipshultz SE. Anthracycline-induced cardiotoxicity: course, pathophysiology, prevention and management. *Expert Opinion on Pharmacotherapy.* 2007;8(8):1039-1058. **• article of interest, describes mechanisms of anthracycline-induced cardiotoxicity**
20. Farmakis D, Parissis J, Filippatos G. Insights into onco-cardiology: atrial fibrillation in cancer. *J Am Coll Cardiol.* 2014;63(10):945-953.
21. Volkova M, Russell R, 3rd. Anthracycline cardiotoxicity: prevalence, pathogenesis and treatment. *Curr Cardiol Rev.* 2011;7(4):214-220.
22. Lipshultz SE. Exposure to anthracyclines during childhood causes cardiac injury. *Semin Oncol.* 2006;33(3 Suppl 8):S8-14.
23. Hutchins KK, Siddeek H, Franco VI, Lipshultz SE. Prevention of cardiotoxicity among survivors of childhood cancer. *Br J Clin Pharmacol.* 2017;83(3):455-465.
24. Fornaro A, Olivotto I, Rigacci L, et al. Comparison of long-term outcome in anthracycline-related versus idiopathic dilated cardiomyopathy: a single centre experience. *European Journal of Heart Failure.* 2018;20(5):898-906.
25. Ewer MS, Ewer SM. Cardiotoxicity of anticancer treatments. *Nat Rev Cardiol.* 2015;12(9):547-558.
26. Swain SM, Whaley FS, Ewer MS. Congestive heart failure in patients treated with doxorubicin: a retrospective analysis of three trials. *Cancer.* 2003;97(11):2869-2879.
27. van Dalen EC, van der Pal HJ, Kok WE, Caron HN, Kremer LC. Clinical heart failure in a cohort of children treated with anthracyclines: a long-term follow-up study. *Eur J Cancer.* 2006;42(18):3191-3198.
28. Mulrooney DA, Yeazel MW, Kawashima T, et al. Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort. *Bmj.* 2009;339:b4606.
29. Dewilde S, Carroll K, Nivelle E, Sawyer J. Evaluation of the cost-effectiveness of dexrazoxane for the prevention of anthracycline-related cardiotoxicity in children with sarcoma and haematologic malignancies: a European perspective. *Cost Eff Resour Alloc.* 2020;18:7-7.
30. Brown TR, Vijarnsorn C, Potts J, Milner R, Sandor GG, Fryer C. Anthracycline induced cardiac toxicity in pediatric Ewing sarcoma: a longitudinal study. *Pediatr Blood Cancer.* 2013;60(5):842-848.
31. Janeway KA, Grier HE. Sequelae of osteosarcoma medical therapy: a review of rare acute toxicities and late effects. *Lancet Oncol.* 2010;11(7):670-678.
32. Blanco JG, Sun C-L, Landier W, et al. Anthracycline-related cardiomyopathy after childhood cancer: role of polymorphisms in carbonyl reductase genes--a report from the Children's Oncology Group. *J Clin Oncol.* 2012;30(13):1415-1421.

33. Linschoten M, Teske AJ, Cramer MJ, Wall Evd, Asselbergs FW. Chemotherapy-Related Cardiac Dysfunction. *Circulation: Genomic and Precision Medicine.* 2018;11(1):e001753.
34. Sawyer DB, Peng X, Chen B, Pentassuglia L, Lim CC. Mechanisms of anthracycline cardiac injury: can we identify strategies for cardioprotection? *Prog Cardiovasc Dis.* 2010;53(2):105-113.
35. Corremans R, Adão R, De Keulenaer GW, Leite-Moreira AF, Brás-Silva C. Update on pathophysiology and preventive strategies of anthracycline-induced cardiotoxicity. *Clinical and Experimental Pharmacology and Physiology.* 2019;46(3):204-215. • **article of interest, describes mechanisms of anthracycline-induced cardiotoxicity**
36. Megias-Vericat JE, Martinez-Cuadron D, Herrero MJ, et al. Pharmacogenetics of Metabolic Genes of Anthracyclines in Acute Myeloid Leukemia. *Curr Drug Metab.* 2018;19(1):55-74. • **article of interest, describes mechanisms of anthracycline-induced cardiotoxicity**
37. Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev.* 2004;56(2):185-229. • **article of interest, describes mechanisms of anthracycline-induced cardiotoxicity**
38. Neudorf U, Schönecker A, Reinhardt D. Cardio-toxicity in childhood cancer survivors "Cure is not enough". *J Thorac Dis.* 2018;10(Suppl 35):S4344-S4350.
39. Tripaydonis A, Conyers R, Elliott DA. Pediatric Anthracycline-Induced Cardiotoxicity: Mechanisms, Pharmacogenomics, and Pluripotent Stem-Cell Modeling. *Clinical pharmacology and therapeutics.* 2019;105(3):614-624.
40. Gewirtz DA. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem Pharmacol.* 1999;57(7):727-741.
41. Zhang S, Liu X, Bawa-Khalfe T, et al. Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nature Medicine.* 2012;18(11):1639-1642.
42. Sinha BK. Free radicals in anticancer drug pharmacology. *Chemico-Biological Interactions.* 1989;69(4):293-317.
43. Doroshow JH. Anthracycline antibiotic-stimulated superoxide, hydrogen peroxide, and hydroxyl radical production by NADH dehydrogenase. *Cancer research.* 1983;43(10):4543-4551.
44. Goodman J, Hochstein P. Generation of free radicals and lipid peroxidation by redox cycling of adriamycin and daunomycin. *Biochemical and Biophysical Research Communications.* 1977;77(2):797-803.
45. Lipshultz SE, Sambatakos P, Maguire M, et al. Cardiotoxicity and Cardioprotection in Childhood Cancer. *Acta Haematologica.* 2014;132(3-4):391-399.
46. Kliment CR, Suliman HB, Tobolewski JM, et al. Extracellular superoxide dismutase regulates cardiac function and fibrosis. *Journal of molecular and cellular cardiology.* 2009;47(5):730-742.
47. Fogli S, Nieri P, Breschi MC. The role of nitric oxide in anthracycline toxicity and prospects for pharmacologic prevention of cardiac damage. *FASEB J.* 2004;18(6):664-675.
48. Sinha BK, Mason RP. IS METABOLIC ACTIVATION OF TOPOISOMERASE II POISONS IMPORTANT IN THE MECHANISM OF CYTOTOXICITY? *J Drug Metab Toxicol.* 2015;6(3).

49. Vavrova A, Jansova H, Mackova E, et al. Catalytic inhibitors of topoisomerase II differently modulate the toxicity of anthracyclines in cardiac and cancer cells. *PLoS One.* 2013;8(10):e76676.
50. Farmakis D, Mantzourani M, Filippatos G. Anthracycline-induced cardiomyopathy: secrets and lies. *European Journal of Heart Failure.* 2018;20(5):907-909.
51. Vejpongsa P, Yeh ETH. Prevention of Anthracycline-Induced Cardiotoxicity: Challenges and Opportunities. *Journal of the American College of Cardiology.* 2014;64(9):938-945.
52. Stewart DJ, Grewaal D, Green RM, et al. Concentrations of doxorubicin and its metabolites in human autopsy heart and other tissues. *Anticancer Res.* 1993;13(6a):1945-1952.
53. Blanco JG, Leisenring WM, Gonzalez-Covarrubias VM, et al. Genetic polymorphisms in the carbonyl reductase 3 gene CBR3 and the NAD(P)H:quinone oxidoreductase 1 gene NQO1 in patients who developed anthracycline-related congestive heart failure after childhood cancer. *Cancer.* 2008;112(12):2789-2795.
54. Wojnowski L, Kulle B, Schirmer M, et al. NAD(P)H oxidase and multidrug resistance protein genetic polymorphisms are associated with doxorubicin-induced cardiotoxicity. *Circulation.* 2005;112(24):3754-3762.
55. Huang K, Hu S, Sparreboom A. Drug transporters and anthracycline-induced cardiotoxicity. *Pharmacogenomics.* 2018;19.
56. Zhai X, Wang H, Zhu X, et al. Gene polymorphisms of ABC transporters are associated with clinical outcomes in children with acute lymphoblastic leukemia. *Arch Med Sci.* 2012;8(4):659-671.
57. Milosevic G, Kotur N, Krstovski N, et al. Variants in TPMT, ITPA, ABCC4 and ABCB1 Genes As Predictors of 6-mercaptopurine Induced Toxicity in Children with Acute Lymphoblastic Leukemia. *J Med Biochem.* 2018;37(3):320-327.
58. Hattinger CM, Biason P, Iacoboni E, et al. Candidate germline polymorphisms of genes belonging to the pathways of four drugs used in osteosarcoma standard chemotherapy associated with risk, survival and toxicity in non-metastatic high-grade osteosarcoma. *Oncotarget.* 2016;7(38):61970-61987.
59. Ansari M, Sauty G, Labuda M, et al. Polymorphism in multidrug resistance-associated protein gene 3 is associated with outcomes in childhood acute lymphoblastic leukemia. *Pharmacogenomics J.* 2012;12(5):386-394.
60. Umlauf J, Horký M. Molecular biology of doxorubicin-induced cardiomyopathy. *Exp Clin Cardiol.* 2002;7(1):35-39.
61. Kotamraju S, Chitambar CR, Kalivendi SV, Joseph J, Kalyanaraman B. Transferrin receptor-dependent iron uptake is responsible for doxorubicin-mediated apoptosis in endothelial cells: role of oxidant-induced iron signaling in apoptosis. *J Biol Chem.* 2002;277(19):17179-17187.
62. Bansal N, Adams MJ, Ganatra S, et al. Strategies to prevent anthracycline-induced cardiotoxicity in cancer survivors. *Cardiooncology.* 2019;5:18-18.
63. Hasinoff BB, Herman EH. Dexrazoxane: how it works in cardiac and tumor cells. Is it a prodrug or is it a drug? *Cardiovasc Toxicol.* 2007;7(2):140-144.
64. Hildebrandt MAT, Reyes M, Wu X, et al. Hypertension Susceptibility Loci are Associated with Anthracycline-related Cardiotoxicity in Long-term Childhood Cancer Survivors.

Scientific Reports. 2017;7(1):9698. •• **article of considerable interest, describes genetic variants included in the review**
65. Krajinovic M, Elbared J, Drouin S, et al. Polymorphisms of ABCC5 and NOS3 genes influence doxorubicin cardiotoxicity in survivors of childhood acute lymphoblastic leukemia. *Pharmacogenomics J.* 2016;16(6):530-535. •• **article of considerable interest, describes genetic variants included in the review**
66. Lipshultz SE, Lipsitz SR, Kutok JL, et al. Impact of hemochromatosis gene mutations on cardiac status in doxorubicin-treated survivors of childhood high-risk leukemia. *Cancer.* 2013;119(19):3555-3562. •• **article of considerable interest, describes genetic variants included in the review**
67. Rajić V, Aplenc R, Debeljak M, et al. Influence of the polymorphism in candidate genes on late cardiac damage in patients treated due to acute leukemia in childhood. *Leukemia & Lymphoma.* 2009;50(10):1693-1698. •• **article of considerable interest, describes genetic variants included in the review**
68. Ruiz-Pinto S, Pita G, Patino-Garcia A, et al. Exome array analysis identifies GPR35 as a novel susceptibility gene for anthracycline-induced cardiotoxicity in childhood cancer. *Pharmacogenet Genomics.* 2017;27(12):445-453. •• **article of considerable interest, describes genetic variants included in the review**
69. Sági JC, Egyed B, Kelemen A, et al. Possible roles of genetic variations in chemotherapy related cardiotoxicity in pediatric acute lymphoblastic leukemia and osteosarcoma. *BMC Cancer.* 2018;18(1):704. •• **article of considerable interest, describes genetic variants included in the review**
70. Semsei AF, Erdelyi DJ, Ungvari I, et al. ABCC1 polymorphisms in anthracycline-induced cardiotoxicity in childhood acute lymphoblastic leukaemia. *Cell Biol Int.* 2012;36(1):79-86. •• **article of considerable interest, describes genetic variants included in the review**
71. Visscher H, Rassekh SR, Sandor GS, et al. Genetic variants in SLC22A17 and SLC22A7 are associated with anthracycline-induced cardiotoxicity in children. *Pharmacogenomics.* 2015;16(10):1065-1076. •• **article of considerable interest, describes genetic variants included in the review**
72. Visscher H, Ross CJD, Rassekh SR, et al. Pharmacogenomic Prediction of Anthracycline-Induced Cardiotoxicity in Children. *Journal of Clinical Oncology.* 2012;30(13):1422-1428. •• **article of considerable interest, describes genetic variants included in the review**
73. Wang X, Liu W, Sun CL, et al. Hyaluronan synthase 3 variant and anthracycline-related cardiomyopathy: a report from the children's oncology group. *J Clin Oncol.* 2014;32(7):647-653.
74. Wang X, Sun CL, Quinones-Lombrana A, et al. CELF4 Variant and Anthracycline-Related Cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study. *J Clin Oncol.* 2016;34(8):863-870. •• **article of considerable interest, describes genetic variants included in the review**
75. Aminkeng F, Bhavsar AP, Visscher H, et al. A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. *Nat Genet.*

2015;47(9):1079-1084. •• **article of considerable interest, describes genetic variants included in the review**
76. Schumacher T, Benndorf RA. ABC Transport Proteins in Cardiovascular Disease-A Brief Summary. *Molecules.* 2017;22(4):589.
77. Visscher H, Ross CJD, Rassekh SR, et al. Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. *Pediatric Blood & Cancer.* 2013;60(8):1375-1381. •• **article of considerable interest, describes genetic variants included in the review**
78. Flens MJ, Zaman GJ, van der Valk P, et al. Tissue distribution of the multidrug resistance protein. *Am J Pathol.* 1996;148(4):1237-1247.
79. de Grouw EPLM, Raaijmakers MHGP, Boezeman JB, et al. Preferential expression of a high number of ATP binding cassette transporters in both normal and leukemic CD34+CD38− cells. *Leukemia.* 2006;20(4):750-754.
80. Girardi E, Cesar-Razquin A, Lindinger S, et al. A widespread role for SLC transmembrane transporters in resistance to cytotoxic drugs. *Nat Chem Biol.* 2020.
81. César-Razquin A, Snijder B, Frappier-Brinton T, et al. A Call for Systematic Research on Solute Carriers. *Cell.* 2015;162(3):478-487.
82. Aminkeng F, Ross CJ, Rassekh SR, et al. Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. *Br J Clin Pharmacol.* 2016;82(3):683-695.
83. Vulsteke C, Pfeil AM, Maggen C, et al. Clinical and genetic risk factors for epirubicin-induced cardiac toxicity in early breast cancer patients. *Breast cancer research and treatment.* 2015;152(1):67-76.
84. Reichwagen A, Ziepert M, Kreuz M, et al. Association of NADPH oxidase polymorphisms with anthracycline-induced cardiotoxicity in the RICOVER-60 trial of patients with aggressive CD20(+) B-cell lymphoma. *Pharmacogenomics.* 2015;16(4):361-372.
85. Cano-Soldado P, Pastor-Anglada M. Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition. *Medicinal Research Reviews.* 2012;32(2):428-457.
86. Mata JF, García-Manteiga JM, Lostao MP, et al. Role of the human concentrative nucleoside transporter (hCNT1) in the cytotoxic action of 5[Prime]-deoxy-5-fluorouridine, an active intermediate metabolite of capecitabine, a novel oral anticancer drug. *Mol Pharmacol.* 2001;59(6):1542-1548.
87. Cano-Soldado P, Pastor-Anglada M. Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition. *Med Res Rev.* 2012;32(2):428-457.
88. Nagasawa K, Nagai K, Ohnishi N, Yokoyama T, Fujimoto S. Contribution of specific transport systems to anthracycline transport in tumor and normal cells. *Curr Drug Metab.* 2001;2(4):355-366.
89. Shen H, Lai Y, Rodrigues AD. Organic Anion Transporter 2: An Enigmatic Human Solute Carrier. *Drug Metab Dispos.* 2017;45(2):228-236.
90. Cheng Y, Vapurcuyan A, Shahidullah M, Aleksunes LM, Pelis RM. Expression of organic anion transporter 2 in the human kidney and its potential role in the tubular secretion of guanine-containing antiviral drugs. *Drug Metab Dispos.* 2012;40(3):617-624.

91. Kobayashi Y, Ohshiro N, Sakai R, Ohbayashi M, Kohyama N, Yamamoto T. Transport mechanism and substrate specificity of human organic anion transporter 2 (hOat2 [SLC22A7]). *J Pharm Pharmacol.* 2005;57(5):573-578.
92. Shibayama Y, Ushinohama K, Ikeda R, et al. Effect of methotrexate treatment on expression levels of multidrug resistance protein 2, breast cancer resistance protein and organic anion transporters Oat1, Oat2 and Oat3 in rats. *Cancer Sci.* 2006;97(11):1260-1266.
93. Lekawanvijit S, Adrahtas A, Kelly DJ, Kompa AR, Wang BH, Krum H. Does indoxyl sulfate, a uraemic toxin, have direct effects on cardiac fibroblasts and myocytes? *Eur Heart J.* 2010;31(14):1771-1779.
94. Liu S, Wang BH, Kompa AR, Lekawanvijit S, Krum H. Antagonists of organic anion transporters 1 and 3 ameliorate adverse cardiac remodelling induced by uremic toxin indoxyl sulfate. *Int J Cardiol.* 2012;158(3):457-458.
95. Bennett KM, Liu J, Hoelting C, Stoll J. Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23. *Mol Cell Biochem.* 2011;352(1-2):143-154.
96. Jung D, Fried M, Kullak-Ublick GA. Human apical sodium-dependent bile salt transporter gene (SLC10A2) is regulated by the peroxisome proliferator-activated receptor alpha. *J Biol Chem.* 2002;277(34):30559-30566.
97. Jonker JW, Schinkel AH. Pharmacological and physiological functions of the polyspecific organic cation transporters: OCT1, 2, and 3 (SLC22A1-3). *J Pharmacol Exp Ther.* 2004;308(1):2-9.
98. Mordente A, Meucci E, Silvestrini A, Martorana GE, Giardina B. New developments in anthracycline-induced cardiotoxicity. *Curr Med Chem.* 2009;16(13):1656-1672.
99. Tesauro M, Thompson WC, Rogliani P, Qi L, Chaudhary PP, Moss J. Intracellular processing of endothelial nitric oxide synthase isoforms associated with differences in severity of cardiopulmonary diseases: cleavage of proteins with aspartate vs. glutamate at position 298. *Proc Natl Acad Sci U S A.* 2000;97(6):2832-2835.
100. Siegel D, Yan C, Ross D. NAD(P)H:quinone oxidoreductase 1 (NQO1) in the sensitivity and resistance to antitumor quinones. *Biochemical pharmacology.* 2012;83(8):1033-1040.
101. Armenian SH, Ding Y, Mills G, et al. Genetic susceptibility to anthracycline-related congestive heart failure in survivors of haematopoietic cell transplantation. *Br J Haematol.* 2013;163(2):205-213.
102. Iacobucci I, Lonetti A, Candoni A, et al. Profiling of drug-metabolizing enzymes/transporters in CD33+ acute myeloid leukemia patients treated with Gemtuzumab-Ozogamicin and Fludarabine, Cytarabine and Idarubicin. *The Pharmacogenomics Journal.* 2013;13(4):335-341.
103. Birley AJ, James MR, Dickson PA, et al. ADH single nucleotide polymorphism associations with alcohol metabolism in vivo. *Hum Mol Genet.* 2009;18(8):1533-1542.
104. Clerk A, Cullingford TE, Fuller SJ, et al. Signaling pathways mediating cardiac myocyte gene expression in physiological and stress responses. *Journal of Cellular Physiology.* 2007;212(2):311-322.
105. Preedy VR, Patel VB, Reilly ME, Richardson PJ, Falkous G, Mantle D. Oxidants, antioxidants and alcohol: implications for skeletal and cardiac muscle. *Front Biosci.* 1999;4:e58-66.

106. Li S-Y, Ren J. Cardiac overexpression of alcohol dehydrogenase exacerbates chronic ethanol ingestion-induced myocardial dysfunction and hypertrophy: role of insulin signaling and ER stress. *Journal of molecular and cellular cardiology.* 2008;44(6):992-1001.
107. Ge W, Yuan M, Ceylan AF, Wang X, Ren J. Mitochondrial aldehyde dehydrogenase protects against doxorubicin cardiotoxicity through a transient receptor potential channel vanilloid 1-mediated mechanism. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease.* 2016;1862(4):622-634.
108. Chaudhary KR, Batchu SN, Seubert JM. Cytochrome P450 enzymes and the heart. *IUBMB Life.* 2009;61(10):954-960.
109. Powell WS, Rokach J. Biosynthesis, biological effects, and receptors of hydroxyeicosatetraenoic acids (HETEs) and oxoeicosatetraenoic acids (oxo-ETEs) derived from arachidonic acid. *Biochimica et biophysica acta.* 2015;1851(4):340-355.
110. Zordoky BNM, Anwar-Mohamed A, Aboutabl ME, El-Kadi AOS. Acute doxorubicin cardiotoxicity alters cardiac cytochrome P450 expression and arachidonic acid metabolism in rats. *Toxicology and Applied Pharmacology.* 2010;242(1):38-46.
111. Seubert JM, Zeldin DC, Nithipatikom K, Gross GJ. Role of epoxyeicosatrienoic acids in protecting the myocardium following ischemia/reperfusion injury. *Prostaglandins Other Lipid Mediat.* 2007;82(1-4):50-59.
112. Zhang Y, El-Sikhry H, Chaudhary KR, et al. Overexpression of CYP2J2 provides protection against doxorubicin-induced cardiotoxicity. *American journal of physiology Heart and circulatory physiology.* 2009;297(1):H37-H46.
113. L'Ecuyer T, Allebban Z, Thomas R, Vander Heide R. Glutathione S-transferase overexpression protects against anthracycline-induced H9C2 cell death. *Am J Physiol Heart Circ Physiol.* 2004;286(6):H2057-2064.
114. Emadi A, Karp JE. The clinically relevant pharmacogenomic changes in acute myelogenous leukemia. *Pharmacogenomics.* 2012;13(11):1257-1269.
115. Windsor RE, Strauss SJ, Kallis C, Wood NE, Whelan JS. Germline genetic polymorphisms may influence chemotherapy response and disease outcome in osteosarcoma: a pilot study. *Cancer.* 2012;118(7):1856-1867. **•• article of considerable interest, describes genetic variants included in the review**
116. Waldenström A, Martinussen HJ, Gerdin B, Hällgren R. Accumulation of hyaluronan and tissue edema in experimental myocardial infarction. *J Clin Invest.* 1991;88(5):1622-1628.
117. Gordon AM, Homsher E, Regnier M. Regulation of contraction in striated muscle. *Physiol Rev.* 2000;80(2):853-924.
118. Kim J, Ogai A, Nakatani S, et al. Impact of blockade of histamine H2 receptors on chronic heart failure revealed by retrospective and prospective randomized studies. *J Am Coll Cardiol.* 2006;48(7):1378-1384.
119. Zeng Z, Shen L, Li X, et al. Disruption of histamine H2 receptor slows heart failure progression through reducing myocardial apoptosis and fibrosis. *Clin Sci (Lond).* 2014;127(7):435-448.
120. He G-H, Cai W-K, Zhang J-B, et al. Associations of Polymorphisms in HRH2, HRH3, DAO, and HNMT Genes with Risk of Chronic Heart Failure. *Biomed Res Int.* 2016;2016:1208476-1208476.

121. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol.* 2002;3(9):639-650.
122. Armenian SH, Lacchetti C, Lenihan D. Prevention and Monitoring of Cardiac Dysfunction in Survivors of Adult Cancers: American Society of Clinical Oncology Clinical Practice Guideline Summary. *Journal of Oncology Practice.* 2017;13(4):270-275.
123. Lemasters JJ, Theruvath TP, Zhong Z, Nieminen AL. Mitochondrial calcium and the permeability transition in cell death. *Biochim Biophys Acta.* 2009;1787(11):1395-1401.
124. Shanmughapriya S, Rajan S, Hoffman NE, et al. SPG7 Is an Essential and Conserved Component of the Mitochondrial Permeability Transition Pore. *Mol Cell.* 2015;60(1):47-62.
125. Baines CP. The Cardiac Mitochondrion: Nexus of Stress. *Annual Review of Physiology.* 2010;72(1):61-80.
126. Hurst S, Baggett A, Csordas G, Sheu SS. SPG7 targets the m-AAA protease complex to process MCU for uniporter assembly, Ca(2+) influx, and regulation of mitochondrial permeability transition pore opening. *J Biol Chem.* 2019;294(28):10807-10818.
127. Dolphin CT, Beckett DJ, Janmohamed A, et al. The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. *J Biol Chem.* 1998;273(46):30599-30607.
128. Schugar RC, Brown JM. Emerging roles of flavin monooxygenase 3 in cholesterol metabolism and atherosclerosis. *Curr Opin Lipidol.* 2015;26(5):426-431.
129. Robinson-Cohen C, Newitt R, Shen DD, et al. Association of FMO3 Variants and Trimethylamine N-Oxide Concentration, Disease Progression, and Mortality in CKD Patients. *PloS one.* 2016;11(8):e0161074-e0161074.
130. Phillips IR, Shephard EA. Flavin-containing monooxygenase 3 (FMO3): genetic variants and their consequences for drug metabolism and disease. *Xenobiotica.* 2020;50(1):19-33.
131. England J, Drouin S, Beaulieu P, et al. Genomic determinants of long-term cardiometabolic complications in childhood acute lymphoblastic leukemia survivors. *BMC Cancer.* 2017;17(1):751.
132. Petrykey K, Lippé S, Robaey P, et al. Influence of genetic factors on long-term treatment related neurocognitive complications, and on anxiety and depression in survivors of childhood acute lymphoblastic leukemia: The Petale study. *PLoS One.* 2019;14(6):e0217314.
133. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1-2):279-284.
134. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. *Genetics.* 2005;171(2):783-790.
135. Cropp CD, Komori T, Shima JE, et al. Organic anion transporter 2 (SLC22A7) is a facilitative transporter of cGMP. *Molecular pharmacology.* 2008;73(4):1151-1158.
136. Hediger MA, Clemencon B, Burrier RE, Bruford EA. The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med.* 2013;34(2-3):95-107.
137. Dawson PA. Role of the intestinal bile acid transporters in bile acid and drug disposition. *Handb Exp Pharmacol.* 2011(201):169-203.
138. Balakrishnan A, Polli JE. Apical sodium dependent bile acid transporter (ASBT, SLC10A2): a potential prodrug target. *Mol Pharm.* 2006;3(3):223-230.

139. Klutho PJ, Dashek RJ, Song L, Baines CP. Genetic manipulation of SPG7 or NipSnap2 does not affect mitochondrial permeability transition. *Cell Death Discov.* 2020;6:5-5.

6.12. SUPPLEMENTAL MATERIAL

Supplemental Table 1. PETALE cohort: patient demographics and clinical characteristics.

		N	%
Sex	Males	115	48.7
	Females	121	51.3
Prognostic risk group	Standard risk	108	45.8
	High risk*	128	54.2
DFCI protocol	(87-01)	18	7.6
	(91-01)	48	20.3
	(95-01)	68	28.8
	(00-01)	75	31.8
	(05-01)	27	11.5
Age at diagnosis - median (range)	9 (0-18)		
Time since the end of treatment	More than 10 years	168	71.2
	Less than 10 years	68	28.8
	Median	**Min**	**Max**
	12	3	24
Dexrazoxane treatment in High-risk patients	Yes	68	28.8
	No	60	25.4
Dexrazoxane cumulative dose	**Median**	**Max**	**Min**

(mg/m2)	2961.8		3706.2	1433.8
Doxorubicin cumulative dose (mg/m2)	**Median**		**Max**	**Min**
	207.2		479.2	41.3
	Landmark	**Threshold**	**Max**	**Min**
FS (%)	less then	28	46.33	24
LVEF M-Mode (%)	less then	55	77.3	47
LVEF 2D (%)	less then	55	66.8	39
LVEDD (mm) z-score	more than	2	4.28	-1.91
binary outcomes				
	Affected	**Unaffected**	**Missing**	
LVEF M-Mode (%)	50 (21.2%)	186 (78.8%)	0 (0.0%)	
LVEF 2D (%)	85 (36.0%)	143 (60.6%)	8 (3.4%)	

DFCI: Dana-Farber Cancer Institute; FS: fractional shortening; LVEF: left ventricular ejection fraction; LVEDD: Left Ventricular End Dia stolic Diameter.
*Criteria for High-risk stratification were mainly attributed based on age, white blood cell count, immunophenotype (presence of T-cell markers) and combination of these factors; as well as central nervous system (CNS) status and Minimal residual disease at diagnosis.

Supplemental Table 2. Top-ranking associations from common variants analysis, of the PETALE cohort using WES data.

LVEDD z score, All cohort, N=236														
Gene	**chr**	**position**	**Ref allele**	**Var allele**	**function**	**rs number**	**MAF**	**P value**	**FDR**	**Genotype 22/12/11**	**Mean 22/12/11**			**SD**
ABCC1	16	16130524	C	G	intronic	rs246232	0.33	0.0005	0.05	23/106/103	0.95/0.83/0.39			1.1/0.9/0.98
LVEF Mmode (%) binary, SR group, N=108														
Gene	**chr**	**position**	**Ref allele**	**Var allele**	**function**	**rs number**	**MAF**	**P value allelic test**	**FDR**	**Additive model**	**0***	**1**	**2**	**P value**
SLC22A7	6	43267831	G	A	intronic	rs70953680	0.1	0.0002	0.01	Unaffected	63	11	1	**0.003**
										Affected	10	6	3	
Gene	**chr**	**position**	**Ref allele**	**Var allele**	**function**	**rs number**	**MAF**	**P value allelic test**	**FDR**	**Dominant model**	**12+22****		**11**	**P value**
SPG7	16	89616790	G	A	intronic	rs66845605	0.1	0.001	0.03	Unaffected	13		60	**0.007**
										Affected	9		10	
LVEF 2D (%) binary, SR group, N=108														
Gene	**chr**	**position**	**Ref allele**	**Var allele**	**function**	**rs number**	**MAF**	**P value allelic test**	**FDR**	**Dominant model**	**12+22****		**11**	**P value**
ABCC1	16	16228242	G	A	intronic	rs2230671	0.2	0.0009	0.047	Unaffected	32		31	**0.002**
										Affected	8		31	

LVEDD: left ventricular end diastolic diameter; LVEF: left ventricle ejection fraction; chr: chromosome; Ref: reference; Var: variant; MAF: minor allele frequency;
SD: standard deviation; FDR: false discovery rate; *ABCC1: ATP Binding Cassette Subfamily C Member 1; SLC22A7: Solute Carrier Family 22 Member 7, ADH7: Alcohol Dehydrogenase 7.*
*Genotypes were recoded as follows: 0-homozygote wild type; 1-heterozygote variant; 2-homozygote variant.
** Genotypes were recoded as follows: 11-homozygote wild type; 12-heterozygote variant; 22-homozygote variant.

Supplemental Table 3. SKAT-O analysis of the rare variants in the PETALE cohort using WES data.

LVEF Mmode (%), All cohort, N=236																
Gene	SKAT-O		SNPs tested				Linear regression model for all variants combined				Collapsed variants**					
	P value	FDR	position	function	rs number	MAF	11/ 12+22	Mean	SD	P value	11/ 12+22*	Mean	SD	P value	Beta [95% CI]	
SLC10A2	0.002	0.036	103698506[a]	exonic	.	0.002	131/60	60.1/ 58.7	5.7/ 6.1	0.121	**179/28**	**60.5/ 56.3**	**5.9/ 4.9**	**0.001**	**(-0.24) [(-0.38)-(-0.11)]**	
			103701648[a]	exonic	rs61966074	0.002										
			103701672[a]	exonic	rs71640248	0.002										
			103701689[a]	exonic	rs139024168	0.002										
			103701690[a]	exonic	rs56398830	0.015										
			103703609[a]	exonic	rs150229163	0.002										
			103704995[a]	exonic	rs201929731	0.002										
			103705001[a]	exonic	rs199920394	0.002										
			103705050	exonic	rs41281678	0.034										
			103705098	**intronic**	**rs66842575***	0.015										
			103705132	**intronic**	**rs67736127***	0.016										
			103710571	intronic	rs142412008	0.004										
			103710635[a]	**exonic**	**rs60380298***	0.015										
			103710685[a]	**exonic**	**rs117447044***	0.002										
			103710721[a]	exonic	rs201341384	0.002										
			103718211	**intronic**	**rs56182000***	0.015										
			103718236[a]	exonic	.	0.004										
			103718308[a]	**exonic**	**rs55971546***	0.045										
			103718370[a]	exonic	rs143297386	0.002										
			103718421[a]	exonic	rs201392859	0.002										
			103718471	**exonic**	**rs41281680***	0.015										
			103718616	**UTR5**	**rs41281682***	0.009										

LVEF 2D (%), All cohort, N=236												
Gene	SKAT-O		SNPs tested				All variants combined					
	P value	FDR	position	function	rs number	MAF	11/12+22	ADH7	11	12+22	P value ***	OR[95% CI]
			100334324[a]	exonic	.	0.002						
			100336649	**exonic**	*****	**0.004**			63	10		

ADH7	0.0013	0.028	100341705	exonic	rs139516819	0.006	166/16	Affected	(86.3%)	(13.7%)	0.056	2.7 [0.94-7.9]	
			100341927	exonic	.	0.002		Unaffected	103 (94.5%)	6 (5.5%)			
			100349136[a]	**exonic**	*	**0.004**							
			100349160	intronic	.	0.002	**Significant combination***						
			100349343	intronic	.	0.002							
			100349765[a]	exonic	.	0.002	**11/12+22***	**ADH7**	**11**	**12+22**	**P value *****	**OR[95% CI]**	
			100349793	**intronic**	**rs72681953***	**0.025**							
			100349797	intronic	.	0.002	218/11	Affected	81 (89.0%)	10 (11.0%)	**0.001**	**16.9 [2.13-134.6]***	
			100349828	intronic	.	0.002							
			100350648	intronic	.	0.002		Unaffected	137 (99.3%)	1 (0.7%)			
			100356528	UTR5	.	0.003							

LVEF: left ventricle ejection fraction; FDR: false discovery rate; MAF : minor allele frequency; *SLC10A2: Solute Carrier Family 10 Member 2; ADH7: Alcohol Dehydrogenase 7*.

*SNPs that are identified as the most important contributors to the association signal obtained through collapsing approach of several SNPs in each gene.

**Collapsed variants (carriers of at least one of rare variants were included into the model, variants with missing values were excluded).

***Chi-square or Fisher test.

[a] Considered functional if identified with at least one of the prediction tools, PolyPhen2 and/or SIFT.

Section C

Chapter 7

Genetic Susceptibility to Hepatic Sinusoidal Obstruction Syndrome in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplantation.

Biol Blood Marrow Transplant. 2020 May;26(5):920-927. doi: 10.1016/j.bbmt.2019.11.026. Epub 2019 Nov 29.

PMID: 31790828 DOI: 10.1016/j.bbmt.2019.11.026

Genetic susceptibility to hepatic sinusoidal obstruction syndrome in pediatric patients undergoing HSCT.

On Behalf of the Pediatric Disease Working Party of the European Society for Blood and Marrow Transplantation.

Running title: Genetics of hepatic SOS in pediatric patients.

Marc Ansari[1,2], Kateryna Petrykey[3,4], Mohamed Aziz Rezgui[3], Veronica Del Vecchio[4], Jacques Cortyl[3], Reginald-Olivier Ralph[4], Tiago Nava[1,2,3], Patrick Beaulieu[3], Pascal St-Onge[3,], Simona Jurkovic Mlakar[1,2], Patricia Huezo-Diaz Curtis[1,2], Chakradhara Rao S. Uppugunduri[1,2], Laurence Lesne[1,2], Yves Théoret[3-5], Yves Chalandon[6], Imke H. Bartelink[7,12], Jaap-Jan Boelens[7,14], Robbert G.M. Bredius[8], Jean-Hugues Dalle[9], Victor Lewis[10], Bill S. Kangarloo[10], Christina Peters[11], Daniel Sinnett[3,13], Henrique Bittencourt[3-5,13] and Maja Krajinovic[3-5,13].

[1]CANSEARCH Research Laboratory, Department of Pediatrics, Faculty of Medicine, University of Geneva, Geneva, Switzerland; [2]Department of Pediatrics, Oncology-Hematology Unit, Geneva University Hospital, Geneva, Switzerland; [3]Charles-Bruneau Cancer Center, Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada; [4]Department of Pharmacology and Physiology, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada; [5]Clinical Pharmacology Unit, Sainte-Justine University Health Center (SJUHC), Montreal, Quebec, Canada; [6]Division of Hematology, Department Oncology, Geneva University Hospital and Faculty of Medicine, University of Geneva, Geneva, Switzerland; [7]Pediatric Blood and Marrow Transplantation Program, University Medical Center Utrecht, The Netherlands; [8]Department of Pediatrics, Division of Immunology, Infectious Diseases and SCT, Leiden University Medical Center, Leiden, The Netherlands; [9]Pediatric Hematology Department, Robert Debré Hospital, Assistance Publique, Hôpitaux de Paris and Paris-Diderot University, Paris, France; [10]Department of Oncology, Alberta Children's Hospital, Calgary, Alberta, Canada; [11]Department of Pediatrics, Stem Cell Transplantation Unit, St Anna Children's Hospital, Vienna, Austria; [12]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Clinical Pharmacology and Pharmacy, Cancer Center Amsterdam, Amsterdam; [13]Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada; [14]Stem cell transplantation and cellular therapy program, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, USA.

Corresponding author:

Maja Krajinovic
Professor, Departments of Pediatrics and Pharmacology and Physiology, University of Montreal

Charles-Bruneau Cancer Center, Sainte-Justine University Health Center (SJUHC),
3175 Chemin de la Côte-Sainte-Catherine, Montreal (Quebec), Canada, H3T 1C5.
E-mail: maja.krajinovic@umontreal.ca
Fax: 1 514-345-4731, Phone: 1 514-345-4931 #6259

7.1. ABSTRACT

Sinusoidal Obstruction Syndrome (SOS) is a well-recognized and potentially life-threatening complication of hematopoietic stem cell transplantation (HSCT). SOS arises from endothelial cell damage and hepatocellular injury mostly due to the transplantation conditioning regimens but also to other patient, disease, and treatment-related factors. Understanding risk factors associated with the development of SOS is critical for early initiation of treatment or prophylaxis. The knowledge about genetic contribution is limited; few studies investigated so far selected set of genes. To get more comprehensive insight in the genetic component, we performed exome-wide association study using genetic variants derived from whole-exome sequencing. The analyses were performed in a discovery cohort composed of 87 pediatric patients undergoing HSCT following busulfan-containing conditioning regimen. Eight lead SNPs were identified after correction for multiple testing and subsequently analyzed in a validation cohort (n=182). Three SNPs were successfully replicated including *rs17146905* (p=0.001), *rs16931326* (p=0.04) and *rs2289971* (p=0.03), located respectively in *UGT2B10, BHLHE22* and *KIAA1715* genes. *UGT2B10* and *KIAA1715* were retained in multivariable model while controlling for non-genetic covariates and previously identified risk variants in *GSTA1* promoter. The modulation of associations by conditioning regimens was noted, *KIAA1715* was dependent on the intensity of conditioning regimen, whereas the effect of *UGT2B10* was equally applicable to all of them. Combined effect of associated loci was also observed (p=0.00006) with genotype-related SOS risk of 9.8. This is the first study addressing the genetic component of SOS at

an exome-wide level and identifying novel genetic variations conferring higher risk of SOS, which might be useful for personalized prevention and treatment strategies.

Keywords

Hepatic sinusoidal obstruction syndrome, hematopoietic stem cell transplantation, busulfan, genetic factors, association study, whole-exome sequencing.

7.2. BACKGROUND

Hepatic sinusoidal obstruction syndrome (SOS), also known as veno-occlusive disease (VOD), is a life-threatening complication that occurs after hematopoietic stem cell transplantation (HSCT)[1]. It is mostly related to the intensity of the conditioning regimen and employed drugs, such as busulfan (Bu), cyclophosphamide (Cy) or melphalan (Mel) that cause sinusoidal endothelial cell damage and hepatocellular injury[2,3]. Other patient, disease and treatment-related factors can also modify the risk of SOS, such as for example patient age, liver dysfunction, concomitant medication, alloreactivity, cytokine-release due to inflammation and engraftment[4-10]. Experimental models showed that endothelial damage triggers formation of gaps between sinusoidal endothelial cells and passage of erythrocyte to the perisinusoidal space[11]. Additionally, locally released cytokines induce activation of cell adhesion molecules, coagulation, and fibrinolytic pathways[12]. Fibrin deposition, clot formation and erythrocyte extravasation have all been reported to contribute to the narrowing of the sinusoids and reduction of hepatic venous outflow, leading to central venular-occlusion, hepatocellular necrosis, hepatic enlargement and ultimately to SOS[2]. Patients with severe forms can have significant complications, including multi-organ failure, and high mortality rate. SOS is reported to occur in up to 18% of pediatric patients after HSCT[7]. Well-established risk factors mentioned above can influence the risk of SOS. Nevertheless, patients with similar treatments, disease and demographics are not equally vulnerable to SOS development suggesting a genetic contribution. Indeed, candidate gene studies have led to the identification of genetic factors contributing to SOS risk including glutathione S transferase (GST) polymorphisms, which might affect Bu metabolism, such as *GSTM1-null* genotype and *GSTA1 *B* haplotype, as well as polymorphisms that might affect glutathione levels and oxidative liver injury[13-18]. However, these studies focused on a selected set of candidate genes. To address the role of genetic susceptibility in a more comprehensive manner, we have used a hypothesis-free

approach and assessed the relationship of SOS development in children undergoing HSCT with genetic variants obtained from whole-exome sequencing (WES). Top-ranking association signals were verified in a replication cohort, identifying novel genetic loci contributing to SOS.

7.3. PATIENTS AND METHODS

7.3.1. Patients groups

Participants were recruited from the Institutional HSCT biobank at Saint-Justine University Health Center (SJUHC), Montreal (Quebec), Canada, and in the context of a multicentric study of European Society for Blood and Marrow Transplantation (EBMT) (ClinicalTrials.gov Identifier: NCT01257854)[19]. The discovery cohort included 87 patients who underwent allogeneic HSCT between 2000 and 2013 at SJUHC, Montreal, (Quebec), Canada, whose DNA was of sufficient quality and quantity to perform WES. The replication cohort was an independent cohort composed of 182 unselected patients, including 61 patients from SJUHC who underwent allogeneic or autologous HSCT, and who were either not included in sequencing due to insufficient DNA quantity or were recruited after the sequencing has been performed (2013-2015). The replication cohort also included 121 pediatric patients who underwent allogeneic HSCT from 2001-2015 in four different centers in Europe and Canada (Geneva University Hospital, University Medical Center Utrecht, Leiden University Medical Center, Robert Debré Hospital, Paris and Alberta Children's Hospital, Calgary) and were in part included in our previous study on *GSTA1* polymorphisms[19]. The characteristics of the discovery and replication cohorts are given in **Tables 1** and **2**, respectively, and are also provided in Supplemental material **(Item S1).** The information on SOS was collected from patients' medical charts. SOS was diagnosed according to the Modified Seattle Criteria[18], which is, the occurrence of two of the following events: unexplained weight gain of more than 2% from baseline because of fluid accumulation, hyperbilirubinemia ≥ 2mg/dL, hepatomegaly, or upper right quadrant pain of liver origin.

7.3.2. Whole exome sequencing

WES in the discovery cohort was performed on germline DNA, extracted from peripheral blood or saliva samples before transplant as described previously[20]. Briefly, exomes were captured in solution

with Agilent's SureSelect Human All Exon V5 + UTRs kit, and sequenced on the Illumina HiSeq2500 platform (mean coverage of 40X*)* at SJUHC integrated clinical genomic centre in pediatrics. Reads were aligned to the hg19 reference genome using BWA-MEM[21]. PICARD[22] was used to mark PCR duplicates and collect sequencing quality control metrics. Variant calling was performed using the Haplotype Caller and quality score recalibration was performed using Variant Recalibrator, both implemented in the Genome Analysis Tool Kit (GATK)[23]. Variants were selected based on the variant quality score (VQSR = PASS) and minimum depth of coverage (DP ≥ 10). The annotation of the identified germline variants was performed using ANNOVAR[24]. Common SNPs located in exons and UTRs with minor allele frequency (MAF) higher than 5% were selected for the analyses. They were filtered to exclude variants exceeding missingness rate of 20%, not in Hardy-Weinberg equilibrium (P<0.001)[25], or with pairwise linkage disequilibrium (LD, r2>0.8). To further reduce the complexity of the analysis and focus on the coding variants with potential causal effects, only non-synonymous variants with predicted functional effect, nonsense variants and variants in splicing sites were conserved. The predicted effect of non-synonymous variants on the protein function was assessed *in silico* using SIFT and PolyPhen-2 [26,27]. All above filtering resulted in 4,946 common exonic variants and 28,540 common SNPs located in UTRs that were retained for the analyses.

7.3.3. Association study

The analysis between genetic variants obtained from WES data and SOS was performed by allelic ratio implemented in PLINK v.1.07[28] using chi-square or Fisher's exact test, as applicable. Analyses were corrected for multiple testing using the Benjamini-Hochberg procedure[29] for the false discovery rate (FDR) with a cut-off value of < 5%. Variants significantly associated with SOS were subsequently analyzed by SPSS (version 24, SPSS Inc, NY) using cumulative incidence of SOS and one minus survival curves in Kaplan–Meier (KM) framework according to the best genetic model presented

relative to the minor allele. The difference between genotypes was assessed by log-rank test. Univariate Cox regression analysis was used to estimate a hazard ratio (HR) with 95% confidence interval (CI). Variants that were significantly associated with SOS were genotyped in the replication cohort using the Sequenom platform at the McGill University, Montreal (Quebec), Canada and Genome Quebec Innovation Centre, or by PCR-coupled Allele Specific Oligonucleotide (ASO) hybridization assay[30]. The pre-transplant DNA was available for 175 patients. The amplification was not equally efficient for all loci and the maximum number of individuals with successful genotyping data ranged from 167 to 175, depending on the locus analyzed. The analyses of cumulative incidence of SOS in relation to the genotypes were performed in KM framework, as explained above. Multivariable Cox regression (backwards stepwise selection method) was used to estimate the impact of associated genotypes in the presence of other covariates in the replication cohort. Other covariates included age as a continuous variable, sex, diagnosis (non-malignant disease, hematological malignancies, and neuroblastoma), frequency of Bu administration (once or four times daily), type of conditioning regimen categorized according to the number of alkylating agents and whether patient received or not total body irradiation (TBI)[31]. *GSTA1* haplotype groups defined previously to be a risk factor for SOS in the same cohorts [16,19] were also included in the multivariate analyses as well as dose adjusted cumulative area under the curve (cumAUC, mg×h/L) estimated from the first dose AUC and each individual dose received. *GSTA1* haplotypes groups are defined by promoter polymorphisms, which are not available in WES data set, therefore previously obtained genotypes were either used or were obtained by PCR-ASO and resulting haplotypes/diplotypes were recoded based on their metabolic capacity (slow metabolizers vs. remaining groups). Stratified analyses according to conditioning regimen were also performed.

7.4. RESULTS

The association analysis using WES data in the discovery cohort revealed eight loci significantly associated with SOS after multiple testing adjustment (**Table 3**, p ranged from $1.5x10^{-5}$ to $8x10^{-7}$). All loci (*HADH, rs17511319; FAT3 rs11823754, UGT2B10 rs17146905; ZNF608 rs75323508; AMPH rs2810; BHLHE22 rs16931326; AGPAT3 rs11537798;* and *KIAA1715 rs2289971*) were in 3'UTR of respective genes. The associated genes are implicated in different cellular functions such as transcriptional regulation, lipid homeostasis, or glucuronidation **(Table S1).** Cumulative SOS incidence in relation to the genotypes is shown in **Fig 1**. In most cases, the dominant model was the most appropriate given the low number of homozygotes for the minor alleles, except for SNPs in *UGT2B10* and *FAT3* gene where the risk increased in an additive manner with each copy of the minor allele. The cumulative risk of SOS ranged from 4.8 to 12.8; HR for the carriers of the minor allele at *UGT2B10 rs17146905* was 4.8 (95% CI =2.3-10.4, $p=4x10^{-6}$) and for *rs17511319* in *HADH* gene was 12.8, 95% CI, 4.1-40 ($p=1.5x10^{-8}$). There was only one case of SOS in Bu/fludarabine (Flu) - based conditioning regimens, thus results mostly reflected association in patients that received Bu-Cy combinations. All associations were equally present in patients with malignant (n=45) and non-malignant disease (n=42) and among patients of European ancestry (n=62) (**Table S2**).

The variants in these genes were further analyzed by genotyping in the replication cohort; the results are summarized in **Table 4.** The association was confirmed for *UGT2B10*, *BHLHE22* and *KIAA1715* gene variants ($p<0.05$). The frequency of risk genotype ranged from 23.1-34.6% in patients with SOS as compared to 10.3-11.4% seen in patients without this complication (Table 4). Although borderline significant association was also noted for *HADH* (p=0.05), the direction of the effect was opposite as compared to the effect seen in discovery cohort. This locus was not considered replicated. The

cumulative incidence of SOS in relation to *KIAA1715, UGT2B10* and *BHLHE22* (p=0.03, 0.001 and 0.04, respectively) is also depicted in **Fig S1**.

Multivariable analyses (**Table 5**) were subsequently performed and included *KIAA1715, UGT2B10* and *BHLHE22* genotypes and non-genetic factors (age, sex, diagnosis, conditioning regimen, cumAUC and TBI). Additionally, *GSTA1* haplotype groups (classified as the slow vs. fast and normal metabolizers) were included in the model based on our previous studies showing in discovery and replication cohort higher risk of SOS for *GSTA1* diplotypes defining slow metabolizing capacity[16,19]. Please note that *GSTA1* promoter polymorphisms defining metabolic capacity were not available in WES dataset and could not be revealed through the WES data analyses in discovery cohort. In the multivariable analysis, *UGT2B10* and *KIAA1715* remained associated with higher risk of SOS (HR =4.7, 95% CI=2.0-11.2, p=0.0004 and HR= 2.7, 95% CI=1.0-7.5, p=0.05, **Table 5**), while controlling for other explanatory covariates. Other factors that remained in the final model included *GSTA1* haplotype groups, disease diagnosis (categorized as non-malignant disease, hematological malignancies, or neuroblastoma) and conditioning regimen in which two or more alkylating agents were classified against Busulfan only, with the higher risk noted for neuroblastoma or more intensive conditioning regimen. The maximal number of cases with available genotypes for non-malignant diseases, hematological malignancies, and neuroblastoma was 73, 92 and 10, and respective number of SOS cases was 8, 13 and 5. 57 patients received condition regimens with one alkylating agent (Bu in Bu-Flu based regimens), 4 cases had SOS; 118 patients received more than one alkylating agent mostly represented by Bu/Cy, Bu/Mel and Bu/Cy/Mel combination, of which 22 had SOS.

We previously reported that association of the *GSTA1* genotypes with SOS was present in double or triple alkylator setting[16,19], and **Table 6** presents stratified analyses in replication cohort by conditioning

regimen relative to the genotypes retained in multivariable model. Beside *GSTA1*, the effect of *KIAA1715* was dependent on the number of alkylating agents (p=0.01), whereas the effect of *UGT2B10* was seen irrespectively of the type of conditioning regimen and was also present in Bu-Flu based protocols (p=0.001). A model combining risk alleles of *KIAA1715*, *UGT2B10* and *GSTA1* was then built and tested in a more intensive conditioning regimen (**Fig 2**). The risk of SOS increased in additive manner from 3.9 (95% CI=1.4-10.4, p=0.007) in patients with one risk genotype at any locus (group 1), to 9.8 (95% CI=2.8-33.8, p=0.0003) in patients with two risk genotypes (group 2) when compared to those with no risk genotypes (group 0). The significance for overall difference across genotypes was 0.00006. There were no patients with three risk genotypes.

7.5. DISCUSSION

This study reports identification of the genetic contribution to SOS in pediatric patients undergoing HSCT. To our knowledge there is no other study that has performed genome or exome wide data analyses in either pediatric or adult patients undergoing HSCT. SOS has been reported as one of the most serious life-threatening complications in the post-transplantation period[1,7]. Identification of patient- and transplantation-specific risk factors for the development of SOS can help guide prophylaxis and treatment of this complication[12]. Our analyses identified three replicated loci, notably *BHLHE22*, *KIAA1715,* and *UGT2B10*, of which the last two were retained in multivariable model, while controlling for other risk factors such as intensity of conditioning regimen, disease indication, Bu exposure, TBI and *GSTA1* haplotypes. None of these genes was previously related to SOS and the exact mechanism of their potential effect is not yet clear. They may contribute to SOS development through mechanisms underlying SOS pathophysiology, conditioning regimen drug pathways or even pathology of the disease for which HSCT is performed.

KIAA1715 is coding for endoplasmic reticulum (ER) junction formation factor also known as a lunapark (LNPK), ubiquitously expressed in number of tissues[32]. The rs2289971 in *KIAA1715* is an expression quantitative trait locus (eQTL) with higher expression noted for minor allele (**Fig S2**), as shown by the analyses of data available through Genotype-Tissue Expression (GTEx) Project[33,34]. Loss-of-function mutations in LNPK lead to aberrant ER structures and increased luminal mass density[32]. When cells are subjected to changes in their extracellular environment, unfolded proteins accumulate in the ER, causing ER stress[35]. This initiates the unfolded protein response (UPR), a signal transduction cascade aiming at restoring cellular homeostasis, which is involved in the pathogenesis of many diseases including chronic liver diseases [32,35,36]. Among other pathways, the UPR plays a significant role in vascular endothelial growth factor A (VEGFA) regulation, which is up regulated in

the presence of endothelial damage[37]. Association of VEGFA levels after conditioning regimen was noted with non-relapse mortality and importantly with SOS; thus, confirming VEGFA significance as an endothelial damage marker in the setting of HSCT[38]. Whether LNPK change in expression can be linked to VEGFA level modulation, or to other mechanisms resulting from ER stress, remains to be determined. *KIAA1715 rs2289971* variant was associated in our study with SOS in patients that received a more intensive conditioning regimen.

The UDP-glucuronosyltransferase (UGT) 2B enzymes are important in the detoxification of a variety of endogenous and exogenous compounds, including many hormones, drugs and carcinogens. Recent observations revealed that human *UGT2B10*, mostly expressed in liver, catalyzes N-glucuronidation of amine-containing compounds[39]. This eventually can affect the levels of biologically active metabolites of Cy and Mel[40]. There is also evidence that UGT expression in the context of chronic lymphocytic leukemia treatment can be affected by fludarabine-containing regimens[41]. It is worth noting that *UGT2B10 rs17146905* was the only variant significantly associated in our study with SOS in Flu-based conditioning regimen. The *UGT2B10* is extensively regulated through alternative splicing [42], but also by microRNA (miR)[43]. Interestingly, *rs17146905* SNP is predicted to lead to target-gain for hsa-miR-454-5p[44].

The *BHLHE22* gene encodes a protein that belongs to the basic helix-loop-helix (bHLH) family of transcription factors that regulate cell fate determination, proliferation, and differentiation[45]. This gene is thought to play a role mostly in neural circuit assembly[46]. However, its downregulation by microRNA was recently reported in the formation of insulin-producing cells[47]. It was also one of three genes whose methylation was most predictive for endometrial cancer[48], suggesting wider regulatory role of this gene in different tissues. How this protein can be connected to SOS is not clear. One of the possibilities is a crosstalk with hypoxia induced factors (HIF), as shown for the other bHLH members[49].

HIF pathway in turn is well known for its regulation of VEGF[35]. The other possibility might lie in the proximity of *BHLHE22* to *CYP7B1* whose hepatic activity is implicated in the inactivation of oxysterols[50] and could suggest a LD between their variants. Loss of *CYP7B1* activity is associated with liver failure in children[51-53]. Nevertheless, *BHLHE22 rs16931326* variant was not retained in a multivariable model, which might suggest its relatively minor role.

Among non-genetic factors included in multivariable model, highest risk of SOS was noted for neuroblastoma and more intensive conditioning regimen. Both factors were previously reported to be associated with higher incidence of SOS[4,5].

Our study suffers certain drawbacks such as limited sample size, heterogeneity in terms of diagnosis and conditioning regimen within and between cohorts and study design that did not include the promoter variants. In spite of uniform SOS diagnostic criteria across study centers, it is worth noting a retrospective nature of the study with the wide period for patients' enrolment, which ended in 2015; therefore, the study did include last up to date diagnostic criteria[6], what might have biased the estimate of SOS incidence. Likewise, the information on SOS prophylaxis was missing in the majority of patients of the replication cohort, for which reason this co-variate was not included in the multivariable model.

7.6. CONCLUSIONS

We used WES data to perform whole exome/adjacent UTR analysis of the genetic component of SOS in pediatric patients undergoing HSCT. Despite certain study drawbacks, three loci conferring higher risk of SOS were successfully replicated in independent patient group. Most of the associations were found in more intensive conditioning regimens (double or triple alkylator setting) and could increase the risk of SOS through combined gene effect, including previously identified *GSTA1* diplotypes underlying low metabolic capacity. These loci have not been previously identified as potential SOS predictors. Although we acknowledge that they still need to be investigated through functional assays, additional replication studies (e.g., with larger sample size, different diseases, and adult population) and prospective evaluation with up to date SOS diagnostic criteria, the current findings could help further understand the role of genetic component on pathophysiology of the SOS. Combining these genetic markers with known risk factors may lead to prediction models to identify children who might be highly susceptible to SOS in a HSCT setting and could possibly benefit from early prophylactic intervention.

7.7. LIST OF ABBREVIATIONS

SOS: Sinusoidal Obstruction Syndrome
HSCT: Hematopoietic Stem Cell Transplantation
WES: Whole exome sequencing
SNP: Single nucleotide polymorphism
VOD: Veno-Occlusive Disease
Bu: Busulfan
Cy: Cyclophosphamide
Mel: Melphalan
GST: glutathione S transferase
SJUHC: Sainte-Justine University Health Center
EBMT: European Society for Blood and Marrow Transplantation
MAF: Minor allele frequency
LD: linkage disequilibrium
FDR: False discovery rate
KM: Kaplan–Meier
HR: hazard ratio
CI: confidence interval
cumAUC: cumulative area under the curve
ER: endoplasmic reticulum
eQTL: expression quantitative trait locus
GTEx: Genotype-Tissue Expression
***LNPK*:** lunapark (endoplasmic reticulum junction formation factor)
UPR: unfolded protein response
VEGFA: vascular endothelial growth factor A
UGT: UDP-glucuronosyltransferase
bHLH: basic helix-loop-helix
HIF: hypoxia induced factors

7.8. ACKNOWLEDGEMENTS

The authors would like to thank all patients and their parents for the participation in the study, as well as all study collaborators for their valuable contribution.

Funding

This investigation was supported by grants from the Swiss National Science Foundation (**grant number 153389**), CANSEARCH Foundation, AOK Foundation and Charles-Bruneau Foundation.

7.9. DECLARATIONS

Ethics approval and consent to participate

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the relevant institutional review boards or ethics committees.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available upon request.

The request should be made to the data access committee composed of senior authors of this study: Dr. M. Krajinovic, maja.krajinovic@umontreal.ca Dr. H. Bittencourt, henrique.bittencourt.hsj@ssss.gouv.qc.ca; Dr. M. Ansari, marc.ansari@hcuge.ch, and President of Ethics committee at CHU Sainte Justine, G. Cardinal, genevieve.cardinal@recherche-ste-justine.qc.ca

Authorship Contributions

MA, HB, and MK designed the study; SD supervised WES analyses; PS-O and PB contributed to bioinformatics analyses; LL, YT, YC, IHB, JJB, RGB, JHD, VL, BK, CP, MA, HB, MAR contributed to patients' sample and data processing; KP, JC, VDV and MK executed computational and statistical analysis; TN, SJM, PH-DC, CRSU contributed to interpretation of results; MAR and ROR performed the replication analysis; MK drafted the manuscript and all authors revised it critically.

Financial disclosure statement

The authors declare no competing financial interests.

7.10. TABLES AND FIGURES

7.10.1. Table1. Demographic and treatment characteristics of discovery cohort (n=87)

Demographic Characteristics		Patients	
		N	%
Gender	Male	40	46
	Female	47	54
Ethnicity	Caucasians	62	71.3
	Other	25	28.7
Diagnosis[1]	Malignancies	45	51.7
	Non-Malignancies	42	48.3
HLA compatibility	Unrelated donor	50	57.5
	Related donor	1	1.1
	HLA identical sibling	36	41.4
Stem Cell Source[1]	BM	43	49.4
	PBSC	2	2.3
	Cord blood	42	48.3
Conditioning	Bu/Cy	62	71.3
	Bu/Cy/VP16	5	5.7
	Bu/Flu	19	21.8
	Bu/Flu/Thiotepa	1	1.1
Busulfan Protocol	One dose per day	67	77
	Four doses per day	20	23
Chemotherapy Regimen[1]	Myeloablative	68	78.2
	Myeloablative with reduced toxicity	16	18.4
	Non-Myeloablative	3	3.4
SOS	Yes	12	13.8
	No	75	86.2
Prophylaxis of SOS	Ursodeoxycholic acid	87	100
Age in years Median (Range)	7.4 (0.1-23.5)		
cumAUC (mg×h/L) Median (Range)	59.6 (25.5-79.0)		
GSTA1[2]	13(14.9%)/74(85.1%)		

Abbreviations: BM: Bone Marrow; PBSC: Peripheral Blood Stem Cells; Bu: Busulfan; CY: Cyclophosphamide; VP16: Etoposide; Flu: Fludarabine; SOS: Sinusoidal Obstruction Syndrome; cumAUC, cumulative area under the

curve; GSTA1, glutathione S transferase A1 [1] Further details of discovery cohort, particularly regarding diagnosis, stem cell source and chemotherapy regimens are provided in supplemental material (Item S1) [2]Number and frequency of diplotypes, as derived from genotype data, with and without reduced metabolic capacity.

7.10.2. Table 2. Demographic and treatment characteristics of replication cohort (n=182).

Characteristics		Patients	
		N	%
Centers	CHU St-Justine, Montreal (Canada)	61	33.5
	Geneva University Hospital, Geneva (Switzerland)	4	2.2
	Robert Debré, University Hospital, Paris (France)	13	7.1
	University Medical Center, Utrecht (Netherlands)	66	36.3
	Alberta Children's Hospital, Calgary (Canada)	38	20.9
Gender	Male	109	59.9
	Female	73	40.1
Ethnicity	Caucasian	150	82.4
	Other	27	14.8
	Not available data	5	2.8
Diagnosis[1]	Hematologic Malignancies	100	54.9
	Neuroblastoma	10	5.5
	Non-Malignancies	72	39.6
HLA compatibility	Unrelated donor	121	66.5
	Related donor	3	1.6
	Autologous	14	7.7
	HLA identical sibling	44	24.2
Stem Cell Source[1]	BM	76	41.8
	PBSC	42	23.1
	Cord blood	63	34.6
	BM+PBSC	1	0.5
Conditioning	Bu/Cy	71	39.0
	Bu/Cy/Mel	31	17.0
	Bu/Cy/VP16	7	3.8
	Bu/Mel	12	6.6
	Bu/Mel/Ara-C	1	0.5
	Bu/Mel/Gem	3	1.7
	Bu/Flu	54	29.7
	Bu/Flu/Thiotepa	3	1.7
Busulfan Protocol	One dose per day	122	67
	Four doses per day	60	33
Chemotherapy Regimen	Myeloablative	125	68.7
	Myeloablative with reduced toxicity[4]	50	27.5
	Non-Myeloablative	7	3.8
Total Body Irradiation[1]	Yes[5]	28	15.4
	No	154	84.6

SOS	Yes	27	14.8
	No	155	85.2
Prophylaxis of SOS	Ursodeoxycholic acid	60	33.0
	Defibrotide	5	2.7
	Heparin	8	4.4
	Defibrotide & Ursodeoxycholic acid	4	2.2
	Not available data	105	57.7
Age in years Median (Range)	4.71 (0.0-21)		
cumAUC (mg×h/L) Median (Range)	61.9 (31.6-118.7)		
GSTA1[2]	24(15.3%)/133(84.7%)		

Abbreviations: BM: Bone Marrow; PBSC: Peripheral Blood Stem Cells; SOS: Sinusoidal Obstruction Syndrome; Bu: Busulfan; CY: Cyclophosphamide; Flu: Fludarabine; Mel: Melphalan; VP16: Etoposide; Gem: Gemcitabine. [1] Further details of replication cohort, particularly regarding diagnosis, stem cell source, chemotherapy regimen and total body irradiation are provided in supplemental material (Item S1) [2]Number and frequency of diplotypes with and without reduced metabolic capacity.

7.10.3. Table 3. Top-ranking loci of SOS identified in discovery cohort through exome-wide association study.

Locus	SNP	Gene	MAF	Allelic P-value	Allelic* Ratio SOS+	Allelic Ratio SOS-	Allelic OR (95% CI)
chr2_176788639 (3'UTR)	rs2289971 (T>C)	*KIAA1715*	0.09	$3.4x10^{-6}$	8/16	7/143	10.2 (3.3 –31.9)
chr4_108956331 (3'UTR)	rs17511319 (A>G)	*HADH*	0.059	$1.2x10^{-5}$	7/17	2/148	30.5 (5.9 – 158.6)
chr4_69696638 (3'UTR)	rs17146905 (A>G)	*UGT2B10*	0.11	$6.9x10^{-6}$	9/15	2/140	8.4 (3.0 – 23.9)
chr5_123973164 (3'UTR)	rs75323508 (C>T)	*ZNF608*	0.07	$1.3x10^{-5}$	7/17	6/144	9.9 (3.0 – 32.8)
chr7_38424328 (3'UTR)	rs2810 (T>C)	*AMPH*	0.09	$1.1x10^{-5}$	8/16	8/142	8.9 (2.9 – 26.9)
chr8_65495333 (3'UTR)	rs16931326 (G>A)	*BHLHE22*	0.09	$1.1x10^{-5}$	8/16	8/142	8.9 (2.9 – 26.9)
chr11_92625944 (3'UTR)	rs11823754 (G>T)	*FAT3*	0.10	$8.3x10^{-7}$	9/15	8/142	10.7 (3.6 – 31.7)
chr21_45403546 (3'UTR)	rs11537798 (A>G)	*AGPAT3*	0.07	$1.3x10^{-5}$	7/17	6/144	9.9 (3.0 – 32.8)

The results are obtained by allelic ratio implemented in PLINK. *KIAA1715*, lunapark (LNPK); *HADH*, hydroxyacyl-CoA dehydrogenase; *UGT2B10*, UDP, glucuronosyltransferase family 2 member B10; *ZNF608*, zinc finger protein 608; *AMPH*, amphiphysin; *BHLHE22*, basic helix-loop-helix family member e22; *FAT3*, FAT atypical cadherin 3; *AGPAT3*, 1-acylglycerol-3-phosphate O-acyltransferase 3; MAF, minor allele frequency in entire cohort. *Allelic ratio is presented as ratio of minor vs. major allele in patients with (+) and without (-) Sinusoidal Obstruction Syndrome (SOS). OR, odds ratio; CI, confidence interval.

7.10.4. Table 4. Summary of the analysis of top-ranking association signals in replication cohort.

Gene/SNP	N of patients (p*)	SOS cases in genotype groups (%) **	Risk genotypes in SOS groups (%)***
KIAA1715	N=175	7/24 (29.2)	7/26 (26.9)
rs2289971	(0.03)	19/151 (12.6)	17/149 (11.4)
HADH	N=172	0/21 (0)	
rs17511319	(0.05)	26/151 (17.2)	
UGT2B10	N=172	9/24 (37.5)	9/26 (34.6)
rs17146905	(0.001)	17/148 (11.5)	15/146 (10.3)
ZNF608	N=174	1/14 (7.1)	
rs75323508	(0.4)	25/161 (15.5)	
AMPH	N=173	6/39 (15.4)	
rs2810	(0.9)	20/134 (14.9)	
BHLHE22	N=169	6/21 (28.6)	6/26 (23.1)
rs16931326	(0.04)	20/148 (13.5)	15/143 (10.5)
FAT3	N=171	4/26 (15.4)	
rs11823754	(1.0)	22/145 (15.2)	
AGPAT3	N=167	2/31 (6.5)	
rs11537798	(0.1)	25/136 (18.4)	

*P is obtained by log-rank test

**Analyses are done according to dominant model. Number of SOS cases in minor allele carriers (heterozygous and homozygous individuals combined) and non-carriers (homozygotes for major allele), with the frequency given in brackets. ***Number of risk genotypes (presence of at least one copy of minor allele, with the frequency in brackets), in patients with SOS (upper line) and without SOS (lower line) for significantly associated loci.

7.10.5. Table 5. Variables retained in stepwise selection in multivariate Cox regression model in replication cohort.

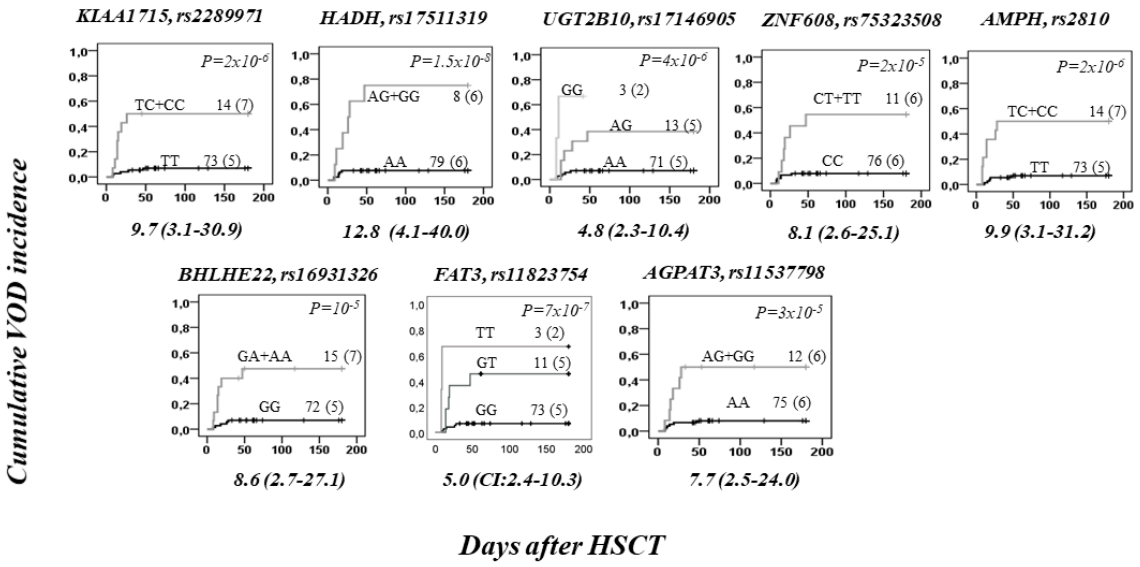
Variable	P-value	HR (95%-CI)
UGT2B10 rs17146905	0.0004	4.7 (2.0-11.5)
KIAA1715 rs2289971	0.05	2.7 (1.0-7.5)
GSTA1	0.02	3.1 (1.2-8.0)
Disease (hematological malignancies)	0.5	1.4 (0.5-3.7)
Disease (neuroblastoma)	0.003	6.1 (1.9-20.1)
Conditioning regimen	0.05	3.2 (1.0-9.9)

HR, hazard ratio, CI, confidence interval.
UGT2B10 and *KIAA1715* are analyzed according to dominant model in which carriers of minor alleles are compared to major allele homozygotes. *GSTA1* diplotypes associated with reduced metabolic capacity are compared to remaining groups. Conditioning regimens are categorised into more than one vs. one alkylating agent (latter corresponding to Bu only or Flu containing regimen). Disease indication is categorised into non-malignant disease, hematological malignancies, and neuroblastoma. Presented HRs reflect risk of hematological malignancies and neuroblastoma; p=0.009 if the risk is compared across disease categories. Other co-variables that were not retained in the final model included age at HSCT and dose-adjusted cumAUC as continues variables, sex, frequency of Bu administration (once vs. 4 times a day), use or not of total body irradiation and *BHLHE22 rs16931326* genotype.

7.10.6. Table 6. Analysis of SOS associated genotypes in replication cohort in relation to conditioning regimen.

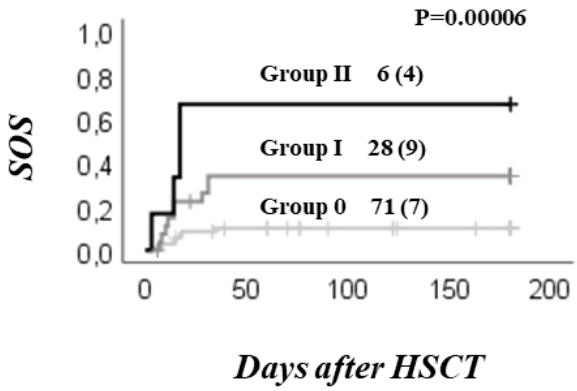
	One alkylating agent (Bu)		Two and more alkylating agents	
Gene/SNP	**N of patients (p*)**	**SOS cases in genotype groups (%) ****	**N of patients (p*)**	**SOS cases in genotype groups (%) ****
KIAA1715	N= 57	1/9 (11.1)	N= 118	6/15 (40.0)
rs2289971	(p = 0.6)	3/48 (6.3)	(p = 0.01)	16/103 (15.5)
UGT2B10	N= 56	3/9 (33.3)	N= 116	6/15 (40.0)
rs17146905	(p = 0.001)	1/47(2.1)	(p = 0.04)	16/101 (15.8)
*GSTA1****	N= 50	0/10 (0)	N= 107	6/14 (42.9)
	(p = 0.3)	4/40 (10.0)	(p = 0.003)	14/93 (15.1)

*P is obtained by log-rank test
**Number of SOS cases in patients with and without the minor allele with the frequencies given in brackets.
*** *GSTA1,* diplotypes with and without reduced metabolic capacity.



7.10.7. Figure 1. Cumulative incidence of SOS in relation to top ranking loci identified through exome-wide association study of discovery cohort.

Cumulative incidence of sinusoidal obstruction syndrome (SOS), plotted for indicated genotype groups according to dominant or additive model. The gene name and SNP rs number are given at the top of the panels. The total number of patients in each group with number of patients with SOS in brackets, and p value derived by log-rank, is depicted on all plots. Hazard ratio with 95% confidence interval in brackets is indicated below panels.



7.10.8. Figure 2. Model combining risk alleles of KIAA1715, UGT2B10 and GSTA1 in intensive conditioning regimen.

Cumulative incidence of SOS in patients that received two or more alkylating agents and have 0, 1 or 2 risk genotypes at any locus. Risk loci are minor alleles for *UGT2B10 rs17146905* and *KIAA1715 rs2289971,* and diplotypes associated with reduced *GSTA1* metabolic capacity. P value derived by log-rank test for the difference across genotypes groups is indicated on the plot. HR (95% CI) for group 1 vs. group 0 is 3.9 (1.4-10.4, p=0.007) and for group 2 vs. group 0 is 9.8 (2.8-33.8, p=0.0003). Group 1 are patients with one risk genotype at any locus, group 2 are patients with any two risk genotypes and group 0 are patients with no risk genotypes; there were no patients with three risk genotypes.

7.11. REFERENCES

1. Coppell JA, Richardson PG, Soiffer R, et al. Hepatic veno-occlusive disease following stem cell transplantation: incidence, clinical course, and outcome. *Biol Blood Marrow Transplant.* 2010;16(2):157-168.
2. Qiao J, Fu J, Fang T, et al. Evaluation of the effects of preconditioning regimens on hepatic veno-occlusive disease in mice after hematopoietic stem cell transplantation. *Exp Mol Pathol.* 2015;98(1):73-78.
3. Qiao J, Huang Y, Xia Y, et al. Busulfan and cyclosphamide induce liver inflammation through NLRP3 activation in mice after hematopoietic stem cell transplantation. *Sci Rep.* 2015;5:17828.
4. Strouse C, Zhang Y, Zhang MJ, et al. Risk Score for the Development of Veno-Occlusive Disease after Allogeneic Hematopoietic Cell Transplant. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* 2018;24(10):2072-2080.
5. Schechter T, Perez-Albuerne E, Lin TF, et al. Veno-occlusive disease after high-dose busulfan-melphalan in neuroblastoma. *Bone marrow transplantation.* 2018.
6. Corbacioglu S, Carreras E, Ansari M, et al. Diagnosis and severity criteria for sinusoidal obstruction syndrome/veno-occlusive disease in pediatric patients: a new classification from the European society for blood and marrow transplantation. *Bone marrow transplantation.* 2018;53(2):138-145.
7. Barker CC, Butzner JD, Anderson RA, Brant R, Sauve RS. Incidence, survival and risk factors for the development of veno-occlusive disease in pediatric hematopoietic stem cell transplant recipients. *Bone marrow transplantation.* 2003;32(1):79-87.
8. Copelan EA. Hematopoietic stem-cell transplantation. *N Engl J Med.* 2006;354(17):1813-1826.
9. Lucarelli G, Galimberti M, Polchi P, et al. Bone marrow transplantation in patients with thalassemia. *N Engl J Med.* 1990;322(7):417-421.
10. Lucarelli G, Clift RA, Galimberti M, et al. Marrow transplantation for patients with thalassemia: results in class 3 patients. *Blood.* 1996;87(5):2082-2088.
11. Braet F, Riches J, Geerts W, Jahn KA, Wisse E, Frederik P. Three-dimensional organization of fenestrae labyrinths in liver sinusoidal endothelial cells. *Liver Int.* 2009;29(4):603-613.
12. Dalle JH, Giralt SA. Hepatic Veno-Occlusive Disease after Hematopoietic Stem Cell Transplantation: Risk Factors and Stratification, Prophylaxis, and Treatment. *Biol Blood Marrow Transplant.* 2016;22(3):400-409.
13. Kallianpur AR, Hall LD, Yadav M, et al. The hemochromatosis C282Y allele: a risk factor for hepatic veno-occlusive disease after hematopoietic stem cell transplantation. *Bone marrow transplantation.* 2005;35(12):1155-1164.
14. Efrati E, Zuckerman T, Ben-Ami E, Krivoy N. MTHFR C677T/A1298C genotype: a possible risk factor for liver sinusoidal obstruction syndrome. *Bone marrow transplantation.* 2014;49(5):726-727.
15. Srivastava A, Poonkuzhali B, Shaji RV, et al. Glutathione S-transferase M1 polymorphism: a risk factor for hepatic venoocclusive disease in bone marrow transplantation. *Blood.* 2004;104(5):1574-1577.

16. Ansari M, Rezgui MA, Theoret Y, et al. Glutathione S-transferase gene variations influence BU pharmacokinetics and outcome of hematopoietic SCT in pediatric patients. *Bone marrow transplantation.* 2013;48(7):939-946.
17. Huezo-Diaz Curtis P, Uppugunduri CRS, Muthukumaran J, et al. Association of CTH variant with sinusoidal obstruction syndrome in children receiving intravenous busulfan and cyclophosphamide before hematopoietic stem cell transplantation. *Pharmacogenomics J.* 2018;18(1):64-69.
18. McDonald GB, Hinds MS, Fisher LD, et al. Veno-occlusive disease of the liver and multiorgan failure after bone marrow transplantation: a cohort study of 355 patients. *Annals of internal medicine.* 1993;118(4):255-267.
19. Ansari M, Curtis PH, Uppugunduri CRS, et al. GSTA1 diplotypes affect busulfan clearance and toxicity in children undergoing allogeneic hematopoietic stem cell transplantation: a multicenter study. *Oncotarget.* 2017;8(53):90852-90867.
20. Spinella JF, Healy J, Saillour V, et al. Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. *BMC Cancer.* 2015;15:539.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
22. Picard Tools - By Broad Institute. https://broadinstitute.github.io/picard/. Accessed.
23. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
25. Rohlfs RV, Weir BS. Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics.* 2008;180(3):1609-1616.
26. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7:Unit7 20.
27. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.
28. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
29. Benjamini YH, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological* 1995(57):289–300.
30. Labuda D, Krajinovic M, Richer C, et al. Rapid detection of CYP1A1, CYP2D6, and NAT variants by multiplex polymerase chain reaction and allele-specific oligonucleotide assay. *Anal Biochem.* 1999;275(1):84-92.
31. Mohty M, Malard F, Abecassis M, et al. Sinusoidal obstruction syndrome/veno-occlusive disease: current situation and perspectives-a position statement from the European Society for Blood and Marrow Transplantation (EBMT). *Bone marrow transplantation.* 2015;50(6):781-789.
32. Zhao Y, Zhang T, Huo H, Ye Y, Liu Y. Lunapark Is a Component of a Ubiquitin Ligase Complex Localized to the Endoplasmic Reticulum Three-way Junctions. *J Biol Chem.* 2016;291(35):18252-18262.

33. Genotype-Tissue Expression (GTEx). https://gtexportal.org/. Accessed.
34. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213.
35. Paridaens A, Laukens D, Vandewynckel YP, et al. Endoplasmic reticulum stress and angiogenesis: is there an interaction between them? *Liver Int.* 2014;34(6):e10-18.
36. Samali A, Fitzgerald U, Deegan S, Gupta S. Methods for monitoring endoplasmic reticulum stress and the unfolded protein response. *Int J Cell Biol.* 2010;2010:830307.
37. Ghosh R, Lipson KL, Sargent KE, et al. Transcriptional regulation of VEGF-A by the unfolded protein response pathway. *PLoS One.* 2010;5(3):e9575.
38. Moiseev IS, Lapin SV, Surkova EA, Lerner MY, Vavilov VN, Afanasyev BV. Level of vascular endothelial growth factor predicts both relapse and nonrelapse mortality after allogeneic hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant.* 2013;19(12):1677-1682.
39. Kato Y, Izukawa T, Oda S, et al. Human UDP-glucuronosyltransferase (UGT) 2B10 in drug N-glucuronidation: substrate screening and comparison with UGT1A3 and UGT1A4. *Drug Metab Dispos.* 2013;41(7):1389-1397.
40. Groehler At, Villalta PW, Campbell C, Tretyakova N. Covalent DNA-Protein Cross-Linking by Phosphoramide Mustard and Nornitrogen Mustard in Human Cells. *Chem Res Toxicol.* 2016;29(2):190-202.
41. Gruber M, Bellemare J, Hoermann G, et al. Overexpression of uridine diphospho glucuronosyltransferase 2B17 in high-risk chronic lymphocytic leukemia. *Blood.* 2013;121(7):1175-1183.
42. Labriet A, Allain EP, Rouleau M, Audet-Delage Y, Villeneuve L, Guillemette C. Post-transcriptional Regulation of UGT2B10 Hepatic Expression and Activity by Alternative Splicing. *Drug Metab Dispos.* 2018;46(5):514-524.
43. Dluzen DF, Sutliff AK, Chen G, Watson CJ, Ishmael FT, Lazarus P. Regulation of UGT2B Expression and Activity by miR-216b-5p in Liver Cancer Cell Lines. *J Pharmacol Exp Ther.* 2016;359(1):182-193.
44. Gong J, Tong Y, Zhang HM, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum Mutat.* 2012;33(1):254-263.
45. Stevens JD, Roalson EH, Skinner MK. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation. *Differentiation.* 2008;76(9):1006-1022.
46. Ross SE, McCord AE, Jung C, et al. Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron.* 2012;73(2):292-303.
47. Bai C, Gao Y, Li X, et al. MicroRNAs can effectively induce formation of insulin-producing cells from mesenchymal stem cells. *J Tissue Eng Regen Med.* 2017;11(12):3457-3468.
48. Huang RL, Su PH, Liao YP, et al. Integrated Epigenomics Analysis Reveals a DNA Methylation Panel for Endometrial Cancer Detection Using Cervical Scrapings. *Clin Cancer Res.* 2017;23(1):263-272.
49. Button EL, Bersten DC, Whitelaw ML. HIF has Biff - Crosstalk between HIF1a and the family of bHLH/PAS proteins. *Exp Cell Res.* 2017;356(2):141-145.

50. Lorbek G, Lewinska M, Rozman D. Cytochrome P450s in the synthesis of cholesterol and bile acids--from mouse models to human diseases. *FEBS J.* 2012;279(9):1516-1533.
51. Stiles AR, McDonald JG, Bauman DR, Russell DW. CYP7B1: one cytochrome P450, two human genetic diseases, and multiple physiological functions. *J Biol Chem.* 2009;284(42):28485-28489.
52. van Mil SW, Houwen RH, Klomp LW. Genetics of familial intrahepatic cholestasis syndromes. *J Med Genet.* 2005;42(6):449-463.
53. Ueki I, Kimura A, Nishiyori A, et al. Neonatal cholestatic liver disease in an Asian patient with a homozygous mutation in the oxysterol 7alpha-hydroxylase gene. *J Pediatr Gastroenterol Nutr.* 2008;46(4):465-469.

7.12. SUPPLEMENTAL MATERIAL

7.12.1. Table S1. Function of the genes associated with a SOS in discovery cohort.

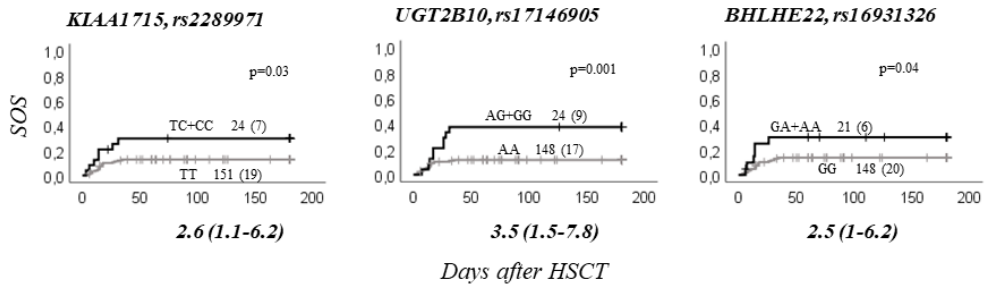
Gene	Role
KIAA1715	Endoplasmic reticulum Junction Formation Factor
HADH	Fatty Acid Oxidation
UGT2B10	Phase 2 Metabolism
ZNF608	Transcription Regulation
AMPH	Cytoplasmic surface of synaptic vesicles
BHLHE22	Transcription factors that regulate cell fate determination, proliferation, and differentiation
FAT3	Cell adhesion
AGPAT3	De novo phospholipid biosynthetic pathway

KIAA1715, lunapark (LNPK); HADH, hydroxyacyl-CoA dehydrogenase; UGT2B10UDP, glucuronosyltransferase family 2 member B10; ZNF608, zinc finger protein 608; AMPH, amphiphysin; BHLHE22, basic helix-loop-helix family member e22; FAT3, FAT atypical cadherin 3; AGPAT3, 1-acylglycerol-3-phosphate O-acyltransferase 3.

7.12.2. Table S2. Relationship between genotypes and SOS in patients of discovery cohort relative to diagnosis and population.

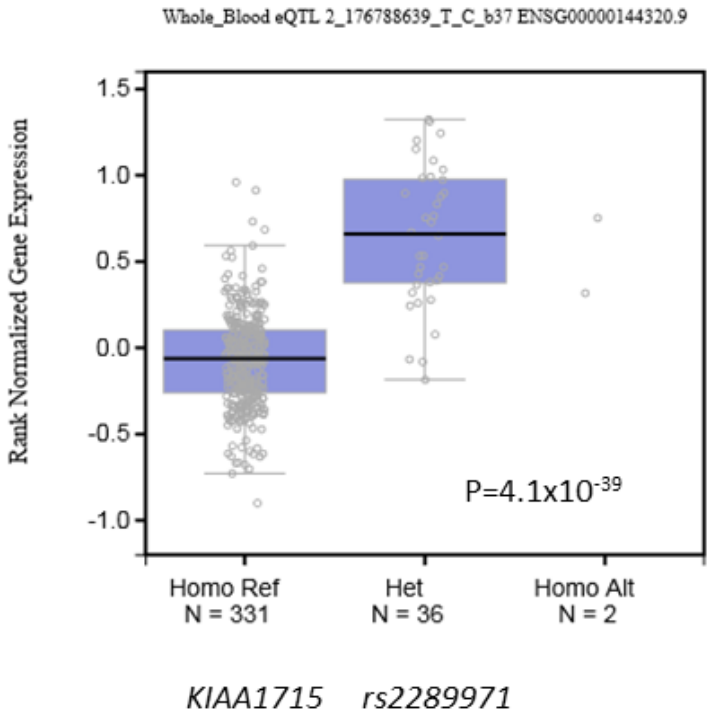
Gene/SNP	MODEL	Malignant (n=45) *	Non-Malignant (n=42) *	European ancestry (n=62) *
KIAA1715 rs2289971	DOM	p = 0.001 4/7 (57.1) 3/38 (7.9)	p = 0.001 3/7 (42.9) 2/35 (5.7)	$p = 2x10^{-6}$ 6/12 (50) 2/50 (4)
HADH rs17511319	DOM	p = 0.001 3/5 (60) 4/40 (10)	$p < 10^{-7}$ 3/3 (100) 2/39 (5.1)	$p < 10^{-7}$ 4/5 (80) 4/57 (7)
UGT2B10 rs17146905	ADD	p = 0.001 2/3 (66.7) 2/6 (33.3) 3/36 (8.3)	p = 0.006 3/7 (42.9) ** 2/35 (5.7)	p = 0.001 1/2 (50) 4/10 (40) 3/50 (6)
ZNF608 rs75323508	DOM	p = 0.01 4/9 (44.4) 3/36 (8.3)	$P = 4x10^{-5}$ 2/2 (100) 3/40 (7.5)	$p < 10^{-7}$ 6/10 (60) 2/52 (3.8)
AMPH rs2810	DOM	p = 0.004 4/9 (44.4) 3/36 (8.3)	$p = 3x10^{-5}$ 3/5 (60) 2/37 (5.4)	p = 0.02 3/8 (37.5) 5/54 (9.3)
BHLHE22 rs16931326	DOM	p = 0.006 3/6 (50) 4/39 (10.3)	$p = 3x10^{-4}$ 4/9 (44.4) 1/33 (3.0)	$p < 10^{-7}$ 5/7 (7.1) 3/55 (5.5)
FAT3 rs11823754	DOM	p = 0.02 3/7 (42.9) 4/38 (10.5)	$p = 8x10^{-6}$ 4/7 (57.1) 1/35	$p = 7x10^{-5}$ 5/10 (50) 3/52 (5.8)
AGPAT3 rs11537798	DOM	p = 0.01 4/9 (44.4) 3/36 (8.3)	$p = 3x10^{-4}$ 2/3 (2.9) 3/39 (7.7)	p = 0.003 4/10 (40) 4/52 (7.7)

*, P is obtained by log-rank test. Number of SOS cases in all patients with the minor allele (heterozygous and homozygous individuals combined for dominant (DOM) model or each of them depicted separately for additive (ADD) model) followed by number of SOS in all patients without minor allele (homozygotes for major allele). SOS frequency for genotype groups is given in brackets. **, no homozygotes for minor allele for these loci was observed.



7.12.3. Figure S1. Cumulative incidence of SOS in replication cohort plotted for indicated genotype groups.

The gene name and SNP rs number are given at the top of the panels. Total number of patients in each group with number of patients with SOS in brackets, and p value derived by log-rank, is depicted on all plots. Hazard ratio with 95% confidence interval in brackets is indicated below panels.



7.12.4. Figure S2. Relationship of KIAA1715 rs2289971 with the mRNA expression.

Representative example depicting significant change in expression according to the KIAA1715 rs2289971 genotype, as obtained from public database: The Genotype-Tissue Expression (GTEx) Project.

7.12.5. Item S1. Demographic and treatment characteristics of replication cohort.

Demographic and treatment characteristics of replication cohort related to Table 1.

Hematologic Malignancies include Acute Lymphoblastic Leukemia (n=3); Acute Myeloid Leukemia (n=27); Myelodysplastic Syndrome (n=13); Chronic Myeloid Leukemia (n=2). Non-Malignancies include Hemoglobinopathy (n=13), Immunodeficiency (n=12), Metabolic Disease (n=5), Chronic Granulomatous Disease (n=7) and Hemophagocytic Syndrome (n=5).

Among 42 patients who received a cord blood transplant, there were 3 patients who received double cord blood.

Patients who received myeloablative regimen with reduced toxicity (n=16), received a Bu/Flu combination with a cumulative dose of Bu >8mg/kg.

None of patients received inotuzumab or gemtuzumab or total body irradiation (TBI).

Demographic and treatment characteristics of replication cohort related to Table 2.

Hematologic Malignancies include Acute Lymphoblastic Leukemia (n=26); Acute Myeloid Leukemia (n=37); Biphenotypic Acute Leukemia (n=2); Myelodysplastic Syndrome (n=30); Chronic Myeloid Leukemia (n=2); Lymphoma (3).

Non-Malignancies include Bone Marrow Failure Syndrome (n=5), Hemoglobinopathy (n=12), Immunodeficiency (n=19), Metabolic Disease (n=22), Chronic Granulomatous Disease (n=1), Hemophagocytic Syndrome (n=10) and Osteoporosis (n=3).

Among 63 patients who received a cord blood transplant, there were 3 patients who received the double cord blood.

Patients who received myeloablative regimen with reduced toxicity (n=50) received a Bu/Flu combination with a cumulative dose of Bu >8mg/kg.

None of patients received inotuzumab or gemtuzumab.

Among patients that received autologous transplantation, 10 patients were with neuroblastoma and 4 patients had Hodgkin (n=2) and non-Hodgkin lymphoma (n=2). None of patients had a tandem HSCT and neuroblastoma patients did not receive therapeutic Metaiodobenzylguanidine (MIBG).

Patients from Alberta Children's Hospital, Calgary (n=28) received a TBI, 4 Gy given in 2 fractions of 2 Gy each. Of these, 24 patients received myeloablative regimen with reduced toxicity and 4 patients received non-myeloablative regimen.

Section C

Chapter 8

Genetic susceptibility to acute graft versus host disease in pediatric patients undergoing HSCT.

Bone Marrow Transplant. 2021 Nov;56(11):2697-2704. doi: 10.1038/s41409-021-01386-8. Epub 2021 Jul 2.

PMID: 34215854 DOI: 10.1038/s41409-021-01386-8

Genetic susceptibility to acute graft versus host disease in pediatric patients undergoing HSCT.

Running title: Genetics of aGVHD in pediatric patients.

Marc Ansari[1,2], Kateryna Petrykey[3,4], Mohamed Aziz Rezgui[3], Veronica Del Vecchio[3,4], Jacques Cortyl[3], Milad Ameur[3,4], Tiago Nava[1,2], Patrick Beaulieu[3], Pascal St-Onge[3], Simona Jurkovic Mlakar[1,2], Chakradhara Rao S. Uppugunduri[1,2], Yves Théoret[3-5], Imke H. Bartelink[6,11], Jaap-Jan Boelens[6,12], Robbert G.M. Bredius[7], Jean-Hugues Dalle[8], Victor Lewis[9], Bill S. Kangarloo[9], Selim Corbacioglu[10], Daniel Sinnett[3,13], Henrique Bittencourt[3,13] and Maja Krajinovic[3-5,13].

[1]Cansearch research platform for paediatric oncology and haematology, Department of Paediatrics, Gynaecology and Obstetrics, Faculty of Medicine, University of Geneva, Geneva, Switzerland; [2]Department of Women, Child and Adolescent, Onco-Hematology Unit, Geneva University Hospital, Geneva, Switzerland; [3]Charles-Bruneau Cancer Center, Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada; [4]Department of Pharmacology and Physiology, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada; [5]Clinical Pharmacology Unit, CHU Sainte-Justine, Montreal, Quebec, Canada; [6]Pediatric Blood and Marrow Transplantation Program, University Medical Center Utrecht, The Netherlands; [7]Department of Pediatrics, Division of Immunology, Infectious Diseases and SCT, Leiden University Medical Center, Leiden, The Netherlands; [8]Pediatric Hematology Department, Robert Debré Hospital, Assistance Publique, Hôpitaux de Paris and Paris-Diderot University, Paris, France; [9]Department of Oncology, Alberta Children's Hospital, Calgary, Alberta, Canada; [10]Department of Hematology, Oncology, and Stem Cell Transplantation University Children's Hospital Regensburg, Germany, [11]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Clinical Pharmacology and Pharmacy, Cancer Center Amsterdam, Amsterdam, The Netherlands; [12]Stem cell transplantation and cellular therapy program, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, USA, [13]Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.

Corresponding author:
Maja Krajinovic
Charles-Bruneau Cancer Center, Sainte-Justine University Health Center (SJUHC),
3175 Chemin de la Côte-Sainte-Catherine, Montreal (Quebec), Canada, H3T 1C5
E-mail: maja.krajinovic@umontreal.ca
Phone: 514-345-4931 #6259

Keywords: hematopoietic stem cell transplantation, acute graft versus host disease, whole-exome sequencing, exome-wide association study, genetic factors.

8.1. ABSTRACT

The most frequent complication of allogeneic hematopoietic stem cell transplantation is acute graft versus host disease (aGVHD). Proliferation and differentiation of donor T cells initiate inflammatory response affecting the skin, liver, and gastrointestinal tract. Besides recipient–donor HLA disparities, disease type, and the conditioning regimen, variability in the non-HLA genotype has an impact on aGVHD onset, and genetic variability of key cytokines and chemokines was associated with increased risk of aGVHD. To get further insight into the recipient genetic component of aGVHD grades2-4 in pediatric patients, we performed an exome-wide association study in a discovery cohort (n= 87). Nine loci sustained correction for multiple testing and were analyzed in a validation group (n=168). Significant associations were replicated for *ERC1* rs1046473*, PLEK* rs3816281, *NOP9* rs2332320 and *SPRED1* rs11634702 variants through the interaction with non-genetic factors. The *ERC1* variant was significant among patients that received the transplant from HLA-matched related individuals (p=0.03), bone marrow stem cells recipients (p=0.007), and serotherapy-negative patients (p=0.004). *NOP9, PLEK,* and *SPRED1* effects were modulated by stem cell source, and serotherapy ($p<0.05$). Furthermore, *ERC1* and *PLEK* SNPs correlated with aGVHD3-4 independently of non-genetic covariates (p=0.02 and p=0.003). This study provides additional insight into the genetic component of moderate to severe aGVHD.

8.2. INTRODUCTION

Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is used in the treatment of hematologic malignant and non-malignant diseases incurable by conventional treatment[1]. However, the success of allo-HSCT is limited by transplant-associated toxicities. The most frequent complication is acute Graft versus Host Disease (aGVHD), occurring in up to 30-60% of children undergoing HSCT[2-4]. Acute GVHD is characterized by an inflammatory response and a combination of signs and symptoms that target the skin, liver, and gastrointestinal tract[1,5]. The major pathophysiological mechanism includes damage of the host tissues by conditioning regimen, which raises the level of proinflammatory cytokines activating host antigen-presenting cells, resulting in donor T cells stimulation. Proliferation and differentiation of donor T cells, in turn, lead to rapid intracellular biochemical cascades activating cellular and inflammatory mediators such as tumor necrosis factors, interleukins, interferons, and nitric oxide, resulting in tissue injury[5-7]. Major clinical factors associated with aGVHD include the degree of HLA compatibility between donor and recipient, patient age, sex, stem-cell source, GVHD prophylaxis, underlying disease, and conditioning regimen[1,5]. Donor T-cell recognition that induces aGVHD can be directed against host major histocompatibility complex (MHC) and/or minor histocompatibility antigen (miH) disparities[5,8]. The role of miH is supported by genome-wide analysis of single nucleotide polymorphisms (SNPs) resulting in amino acid coding differences between recipients and donors[9]. Additionally, SNPs for chemokines, cytokines, and their receptors, costimulatory molecules, micro-RNAs, and drug-metabolizing enzymes have been reported as associated with aGVHD risk. Genetic associations reported so far indicate the limited utility of a single SNP as a predictive biomarker. Indeed, predictive models combining several genetic variants have been developed[6,10].

To further address the role of recipient genetic susceptibility in aGVHD, we have assessed the relationship of aGVHD development in children undergoing HSCT with genetic variants obtained from whole-exome sequencing (WES). Top-ranked association signals were further verified in an independent cohort, identifying novel genetic loci associated with higher aGVHD incidence.

8.3. PATIENTS AND METHODS

8.3.1. Patients' groups

Participants were recruited from the Institutional HSCT biobank at Sainte-Justine University Health Center (SJUHC), Montreal (Quebec), Canada, and in the context of a multicentric study of the European Society for Blood and Marrow Transplantation (EBMT) (ClinicalTrials.gov Identifier: NCT01257854)[11]. Discovery and replication cohorts were presented in a recently reported study on the association between WES and post-HSCT sinusoidal obstructive syndrome (SOS) in pediatric patients[12]. The discovery group included 87 children who underwent allo-HSCT between 2000 and 2013 at SJUHC. The replication cohort was an independent cohort composed of all remaining (unselected) 168 patients who underwent allo-HSCT. These included 47 patients from SJUHC, who were either not included in sequencing due to insufficient DNA quantity or were recruited after the sequencing has been performed (2013-2015). The remaining patients are 121 pediatric patients who underwent allo-HSCT from 2001-2015 in four different centers in Europe and Canada (Geneva University Hospital, University Medical Center Utrecht, Leiden University Medical Center, Robert Debré Hospital, Paris, and Alberta Children's Hospital, Calgary). The characteristics of the discovery and replication cohorts are as previously described[12] and are also given in **Table 1** and Supplemental material (**Item S1**). The information on aGVHD was collected from patients' medical charts and was based on skin rash, serum bilirubin levels, diarrhea, and upper gastrointestinal symptoms according to the 1994 Consensus Conference[13]. Clinical presentation was used to differentiate acute from chronic form. Acute GvHD was diagnosed and considered up to day 180 post-HSCT to consider late cases of aGvHD[14]. Only moderate and severe cases diagnosed as grades 2 to 4 were considered for the analyses. Prophylaxis of GvHD was given for all patients, four patients of the replication cohort did not have available

data. Prophylaxis includes calcineurin inhibitors associated or not with methotrexate, mycophenolate mofetil (MMF), and/or steroids. Details are provided in Supplemental material (**Item S1**).

8.3.2. Whole exome sequencing

WES was performed on the patient's DNA in the discovery cohort as described previously[12]. Exomes were captured in solution with Agilent's SureSelect Human All Exon V5 + untranslated (UTRs) kit and sequenced on the Illumina HiSeq2500 platform (mean coverage of 40X) at SJUHC integrated clinical genomic center in pediatrics. Common SNPs that were included in the association analyses were defined by minor allele frequency (MAF) higher than 5%, according to the frequency reported in public datasets[15]. They were filtered to exclude variants exceeding a missingness rate of 5%, not in Hardy-Weinberg equilibrium (P<0.001), or with pairwise linkage disequilibrium (LD, r2>0.8). To further reduce the complexity of the analysis and focus on the coding variants with potential causal effects or potentially acting as miH, for exonic variants, only non-synonymous variants with predicted functional effect, nonsense variants, and variants in splicing sites were conserved[16,17]. All the above filtering resulted in 4,911 common exonic variants and 27,086 common SNPs located in UTRs that were retained for the analyses.

8.3.3. Association study

For all association analyses, patients with and without aGVHD were considered as cases and controls, respectively. The analysis between genetic variants and aGVHD grade 2-4 was performed by allelic ratio implemented in PLINK v.1.07[18] using chi-square or Fisher's exact test, as applicable. Analyses were corrected for multiple testing using false discovery rate (FDR)[19] with a cut-off value of < 1%, which was close to Bonferroni corrected p=1 x10^{-5} for exonic variants and p=2 x10^{-6} for variants located in UTRs. Variants significantly associated with aGVHD2-4 were

subsequently analyzed by SPSS (version 24, SPSS Inc, NY) using one minus survival curves in Kaplan-Meier (KM) framework. The difference between genotypes was assessed by a log-rank test. Univariate and multivariate Cox regression analysis was used to estimate a hazard ratio (HR) with a 95% confidence interval (CI). Cumulative incidence of aGVHD in relation to the genotype was additionally estimated using competing events (death or death and relapse) and Gray test analyses using EZR software[20]. Nine SNPs sustained correction for multiple testing and were genotyped in the replication cohort using the Sequenom platform at the Genome Quebec Innovation Centre, or by PCR-coupled Allele-Specific Oligonucleotide (ASO) hybridization assay[21]. The amplification was not equally efficient for all loci and the maximum number of individuals with successful genotyping data ranged from 147 to 164, depending on the locus analyzed. The replication cohort has sufficient power (≥90%) to detect similar effect sizes as the ones obtained for these loci in the discovery cohort. The analyses of cumulative incidence of aGVHD2-4 and aGVHD3-4 in relation to the genotypes were performed in the KM framework and accounting for a competing event, as explained above. Univariable and multivariable Cox regression (stepwise selection method) was used to estimate the impact of non-genetic covariates. The covariates included age as a continuous variable, sex, diagnosis (non-malignant disease and hematological malignancies), type of conditioning regimen categorized according to the number of alkylating agents, to which total body irradiation (TBI) was added as the "equivalent" of one alkylating agent, busulfan (Bu) exposure (cumulative area under the curve (cumAUC, mg×h/L)[11,22], type of donor (HLA identical related individuals vs other donors); stem cell source (peripheral blood stem cell vs. bone marrow or cord blood) and whether patients received (or not) serotherapy. Covariates with $p \leq 0.1$ were retained for further stratified analyses with genetic

factors. Genetic factors were considered significantly replicated if $p<0.05$ and if more than one individual harbored risk genotype among aGVHD cases.

8.4. RESULTS

The association analysis using allelic ratio and WES data of the discovery cohort revealed 11 loci significantly associated with aGVHD2-4 after multiple testing adjustments (**Table 2**), of which 5 were in the coding region and 6 in UTRs. The difference in aGVHD2-4 incidence according to the genotypic model is shown in **Table 3** and **Supplemental Figure 1.** Two variants did not sustain correction for multiple testing in the genotyping model and 9 variants were carried further to the replication phase. Among these 9 loci, the hazard ratio (HR) of aGVHD2-4 in the discovery cohort ranged from 4.0 to 13.5 and all of them remained significant in the multivariable model with the inclusion of non-genetic covariates (**Supplemental Table 1**). The cumulative aGVHD2-4 incidence in the discovery cohort (obtained in 1-KM framework) did not change if competing event analyses were performed **(Supplemental Figure 2** and **Supplemental Table 2).** Details of association analyses in the replication group are presented in **Supplemental Table 3;** none of the tested loci was significant in the entire cohort or in patients of European ancestry.

We next investigated whether the impact of genetic variants is modulated by clinical risk factors. For that, we analyzed non-genetic covariates through a univariable (**Supplemental Table 4)** and multivariable Cox regression model, the latter showed associations of patient age, type of donor, serotherapy, and stem cell source with aGVHD2-4 (**Table 4**). The association of genetic variants with aGVHD2-4 in relationship with these covariates was then assessed through stratified analyses. Significant associations were found for the *ERC1, PLEK, NOP9* and *SPRED1* SNPs (rs1046473, rs3816281 rs2332320, and rs11634702, respectively, $p<0.05$, **Figure 1**). The association followed either the additive (*ERC1*) or recessive genetic model for the remaining loci, with the risk value ranging from 2.2 to 5.8. The *ERC1* variant was significant among patients that received the transplant from HLA-identical siblings ($p=0.03$). It was also significant among

recipients of bone marrow grafts (p=0.007), likely due to the correlation with HLA compatibility **(Supplemental Table 5)** and among patients that did not receive serotherapy (p=0.004). *PLEK* was significant among serotherapy-negative patients (p=0.04). *SPRED1* was significant among serotherapy positive patients (p=0.02) and *NOP9* among patients who received cord blood stem cells (p=0.01). We next verified whether any of the 9 variants that were carried forward for replication analyses were predictive of aGVHD3-4. Significant associations were found for the same variants in the *ERC1* (rs1046473, p=0.02) and *PLEK* (rs381628, p=0.003) genes in the entire replication cohort independently of non-genetic covariates (**Figure 2**). Serotherapy and HLA compatibility between donors and recipients were clinical factors associated with aGVHD3-4 (p=0.0002 and 0.01, respectively). None of the results of the replication cohort were affected by competing event analyses **(Supplemental Table 6).**

8.5. DISCUSSION

This study reports the identification of the recipient genetic contribution to higher aGVHD grades in pediatric patients undergoing HSCT using WES data. Although an exome-wide association study was performed on overall survival, transplantation- and disease-related mortality[23], to our knowledge, there is no other study that has performed such analyses with aGVHD in either pediatric or adult settings. We identified four novel loci that can contribute to aGVHD development, notably *ERC1*, *PLEK, NOP9*, and *SPRED1,* which were validated in the replication cohort. It is worth mentioning that they were associated with aGVHD2-4 development through interaction with known clinical predisposing factors, such as serotherapy, HLA compatibility, and stem cell source, whereas *ERK1* and *PLEK* gene variants were also associated with aGVHD3-4 development irrespective of non-genetic covariates. None of these genes was previously related to aGVHD and the exact mechanism of their potential effect is not yet clear. Their possible connection to aGVHD development is summarized below.

ERC1 (ELKS/RAB6-interacting/CAST family member 1) participates in the canonical DNA damage response (DDR) signaling pathway, which is interesting in the context of its possible connection to an aGVHD onset. The treatment administered during conditioning regimens, like the one administered to the patients in this study, induces DNA damage and activates DDR signaling resulting in cell cycle arrest - cellular senescence, which has been described to be associated with pro-inflammatory secretion controlled by the NF-kB system involved in aGVHD pathogenesis[24-28]. Indeed, knock-down of the *ERC1* gene resulted in reduced activity of NF-kB[29]. Two SNPs from the *NFKB1* gene were recently identified associated with aGVHD in recipients allografted from HLA-identical donors[28]. The rs1046473 in *ERC1* gene associated with aGVHD in our study has been predicted to result in gain and loss of binding sites for several miRNAs [30,31]

(**Supplemental Table 3**), which might modulate the *ERC1* expression, potentially leading to increased NFkB activity.

There are several studies showing that the *PLEK* (pleckstrin) gene provides a common genetic background for several autoimmune diseases and inflammatory conditions[32,33]. The gene expression profile of several such conditions, including periodontitis, cardiovascular disease, rheumatoid arthritis, and ulcerative colitis, identified *PLEK* as the only gene significantly overexpressed in all of them implicating its participation in the common inflammatory pathways underlying these diseases[34]. This is in accordance with a known role of PLEK as a protein kinase C substrate, inducibly (including bacterial stimuli) expressed by macrophages[34]. Importantly, PLEK was suggested to be an important intermediate in the secretion and activation pathways of the pro-inflammatory cytokines, TNF-a and IL-1ß[35,36], secreted by lymphocytes, macrophages, or tissue cells providing a possible connection to the aGVHD pathophysiology. The rs3816281 in *PLEK* is predicted to lead to amino-acid substitution that might affect protein function (**Supplemental Table 3**). The same variant is identified as an expression quantitative trait locus, most likely due to LD with the promoter variants in the same gene, as shown by the analyses of data available through the Genotype-Tissue Expression (GTEx) Project (**Supplemental Figure 3**)[37]. Interestingly, preliminary analyses of our in-house RNAseq data in relation to half-maximal inhibitory concentration (IC50) of Busulfan (Bu) in lymphoblastoid cell lines (LCLs) showed a 1.5 increase in *PLEK* expression levels after Bu treatment among Bu-sensitive LCLs (**Supplemental Figure 4**), further advocating potential implication of this gene.

SPRED1 (sprouty-related, EVH1 domain-containing protein 1), among other functions, acts as a tumor suppressor and is a negative regulator of the RAS MAPK pathway[38]. It is highly expressed in hematopoietic cells and negatively regulates hematopoiesis by suppressing stem cell factor and

interleukin-3 (IL-3)-induced ERK activation[38]. *SPRED1* plays a role in leukemia predisposition[38,39]. It was also shown to play a role in the repair of the endothelium after vascular injury[40], possibly linking it to the damage of vascular endothelium and endothelial cells associated with aGVHD[41-43]. Additionally, several studies implicated mast cell involvement in GVHD pathogenesis[44,45]; better survival was identified in the murine T cell-mediated model of GVHD in which recipients lack mast cells[45]. *SPRED1* in turn negatively regulates mast cell activation, as shown by *SPRED1*- knockout mice[46]. SPRED1 was also shown to play an important role in the proliferation, apoptosis, and cytokine secretion of innate lymphoid cells[47], which are associated with reduced susceptibility to GVHD[48]. Further work is needed to demonstrate whether any of these hypotheses are true and to establish a direct functional link of rs11634702 with aGVHD.

The knowledge about nuclear protein Nop9 is limited. *NOP9* is similar to the PUF family of RNA-binding proteins. Both Nop9 and Puf-A participate in sequence-specific RNA recognition and preribosomal RNA processing[49,50]. Alterations in expression level or sequence of *NOP9* and Puf-A are associated with breast cancer, autoimmunity, and learning/language impairment[50-53]. One of the genes that belong to PUF RNA binding family members is identified as a minor histocompatibility antigen[52,53]. Whether a similar role can be attributed to *NOP9* remains to be defined. Both rs11634702 *SPRED1* and rs2332320 *NOP9* are located in 3'UTR and have been predicted to result in gain and loss of binding sites for several miRNAs[30,31] (**Supplemental Table 3**) which might modulate their expression. In that regard, miRNA was indeed found to regulate the translation of *SPRED1*[54].

We have to emphasize certain drawbacks of the study, such as a limited number of patients, which could have affected the study power and the accuracy of analyses within the subgroups and overestimated the effect sizes. Likewise, there is a lack of genetic analyses in donors as well as

cohort's heterogeneity, including treatment characteristics, that could have resulted in different frequencies of aGVHD between cohorts. Given the heterogeneity in GVHD prophylaxis, we could not investigate the relationship between a particular prophylaxis type and the genetic component. In conclusion, we used WES data to perform whole exome/adjacent UTR analysis of the genetic component of aGVHD in pediatric patients undergoing HSCT. Four loci conferring the higher risk of aGVHD were replicated in the independent patient groups; and are dependent on the interaction with non-genetic factors such as serotherapy, stem cell source, and donor type. The loci identified here have not been previously linked to aGVHD but have been linked to inflammatory and autoimmune diseases and, despite certain study drawbacks, could help to further understand the role of the genetic component on the pathophysiology of the aGVHD. Given that prior work assessing the genetics of aGVHD indicates that the SNPs located in the several cytokines combined with clinical risk factors improve aGVHD prediction[6], it would be interesting to assess whether the polygenic risk score can be further improved with the loci discovered in this study. Combining genetic markers with the best predictive ability with other known risk factors may lead to prediction models to identify children who might be highly susceptible to aGVHD in an HSCT setting and could possibly benefit from early prophylactic and therapeutic interventions.

8.6. DECLARATIONS

Ethics approval and consent

Written informed consent was obtained from every patient or parent/legal guardian. The study was conducted in accordance with the Declaration of Helsinki and was approved by Research Ethics Board of SJUHC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available upon request.
The request should be made to the data access committee composed of senior authors of this study:
Dr. M. Krajinovic, maja.krajinovic@umontreal.ca;
Dr. H. Bittencourt, henrique.bittencourt.hsj@ssss.gouv.qc.ca;
Dr. M. Ansari, marc.ansari@hcuge.ch,
and President of Ethics committee at SJUHC, G. Cardinal,
genevieve.cardinal@recherche-ste-justine.qc.ca.

Competing interests

The authors declare no competing financial interests.

Funding

This investigation was supported by grants from the Swiss National Science Foundation (grant number 153389), CANSEARCH Foundation, AOK Foundation, and Foundation of Charles-Bruneau.

Acknowledgements

Reported On Behalf of the Pediatric Disease Working Party of the European Society for Blood and Marrow Transplantation.

The authors would like to thank all patients and their parents for the participation in the study, as well as all study collaborators for their valuable contribution.

Author Contributions

MA, HB, and MK designed the study; DS supervised WES analyses; PS-O and PB contributed to bioinformatics analyses; YT, IHB, JJB, RGMB, JHD, VL, BSK, SC, MA, HB, MAR contributed to patients' sample and data processing; KP, JC, VDV and MK executed computational and statistical analysis; TN, SJM, CRSU contributed to interpretation of results; MAR and MA performed the replication analysis; MK drafted the manuscript and all authors revised it critically.

8.7. TABLES AND FIGURES

8.7.1. Table 1. Demographic and treatment characteristics of discovery (n=87) and replication cohort (n=168).

Demographic Characteristics		Cohorts	
		Discovery	Replication
		Patients	
		N (%)	N (%)
Centers	CHU St-Justine, Montreal (Canada)	87 (100)	47 (28)
	Geneva University Hospital, Geneva (Switzerland)	–	4 (2.4)
	Robert Debré, University Hospital, Paris (France)	–	13 (7.7)
	University Medical Center, Utrecht (Netherlands)	–	66 (39.3)
	Alberta Children's Hospital, Calgary (Canada)	–	38 (22.6)
Gender[2]	Male	40 (46)	101 (60.1)
	Female	47 (54)	67 (39.9)
Ethnicity	European ancestry	69 (79.3)	147 (87.5)
	Other	18 (20.7)	16 (9.5)
	Not available data	–	5 (3)
Diagnosis[1]	Malignancies	45 (51.7)	94 (56)
	Non-Malignancies	42 (48.3)	74 (44)
HLA compatibility[1]	Unrelated donor	51 (58.6)	121 (72)
	Identical related donor	30 (34.5)	44 (26.2)
	Non-identical related donor	6 (6.9)	3 (1.8)
Stem Cell Source[1,2]	BM	43 (49.4)	70 (41.7)
	PBSC	2 (2.3)	31 (18.4)
	Cord blood	42 (48.3)	63 (37.5)
	BM+PBSC	–	4 (2.4)
Conditioning	Bu/Cy	62 (71.3)	71 (42.3)
	Bu/Cy/VP16	5 (5.7)	7 (4.2)

	Bu/Flu	19 (21.8)	54 (32.1)
	Bu/Flu/Thiotepa	1 (1.1)	3 (1.8)
	Bu/Mel	–	2 (1.2)
	Bu/Mel/Cy	–	31 (18.5)
Chemotherapy Regimen[1,]	Myeloablative	68 (78.2)	113 (67.3)
	Myeloablative with reduced toxicity	16 (18.4)	48 (28.6)
	Non-Myeloablative	3 (3.4)	7 (4.2)
Total Body Irradiation[1,2]	Yes	–	28 (16.7)
	No	–	140 (83.3)
aGvHD grade 2-4[2]	Yes	11 (12.6)	45 (26.8)
	No	76 (87.4)	123 (73.2)
aGvHD grade 3-4	Yes	5 (5.7)	17 (10.1)
	No	82 (94.3)	151 (89.9)
Serotherapy[2]	No serotherapy	23 (26.4)	23 (13.7)
	Antithymocyte Globulin	56 (64.4)	121 (72)
	Alemtuzumab (Campath)	8 (9.2)	2 (1.2)
	NA	–	22 (13.1)
Age in years Median (Range)[2]		7.4 (0.1-23.5)	4.71 (0.0-21)
cumAUC (mg×h/L) Median (Range)[2]		59.6 (25.5-79.0)	61.14 (31.6-118.7)

Abbreviations: BM: Bone Marrow; PBSC: Peripheral Blood Stem Cells; Bu: Busulfan; CY: Cyclophosphamide; Mel: Melphalan; Flu: Fludarabine; VP16: Etoposide; aGvHD: Acute Graft-versus-Hot Disease; cumAUC, cumulative area under the curve; [1]Further details of discovery and replication cohort, particularly regarding diagnosis, stem cell source, chemotherapy regimens, HLA compatibility, total body irradiation and aGvHD prophylaxis are provided in supplemental material (Item S1). [2]Characteristics that are significantly different between discovery and replication cohort at $p<0.05$, as obtained by chi-square for categorical variables and Mann-Whitney for continuous variables, for HLA compatibility, identical related individuals are compared to the rest of the donors and for the conditioning regimens, Flu-based regimens are compared to the remaining ones; NA, not available.

8.7.2. Table 2. Top-ranking loci of aGvHD grade 2-4 identified in discovery cohort through exome-wide association study using allelic ratio test.

Locus	SNP	Gene	MAF (%)	P-value	Allelic* Ratio GvHD +	Allelic Ratio GvHD -	OR (95% CI)
chr3:122003769 (exonic)	rs1042636 (A>G)	*CASR*	6.9	$8x10^{-7}$	7/15	5/147	13.7 (3.9 –48.6)
chr15:89195245 (exonic)	rs59188950 (C>T)	*ISG20*	13.2	$1.8x10^{-6}$	10/12	13/139	8.9 (3.2 – 24.6)
chr16:48258198 (exonic)	rs17822931 (A>G)	*ABCC11*	13.2	$1.8x10^{-6}$	10/12	13/139	8.9 (3.2 – 24.6)
chr17:32647831 (exonic)	rs1133763 (A>C)	*CCL8*	11.5	$3.7x10^{-6}$	9/13	11/141	8.9 (3.1 – 25.3)
chr2:68607947 (exonic)	rs3816281 (G>T)	*PLEK*	25.9	$1.5x10^{-5}$	14/8	31/121	6.8 (2.6 – 17.7)
chr15:38647386 (3'UTR)	rs11634702 (A>G)	*SPRED1*	10.0	$2.2x10^{-7}$	9/13	8/140	12.1 (4.0– 36.7)
chr1:109289353 (5'UTR)	rs1124427 (G>C)	*STXBP3*	27.6	$4.0x10^{-7}$	16/6	32/120	10.0 (3.6– 27.6)
chr12:1600978 (3'UTR)	rs1064125 (A>T)	*ERC1*	14.9	$8.0x10^{-7}$	11/11	15/137	9.1 (3.4– 24.6)
chr14:24776219 (3'UTR)	rs2332320 (T>C)	*NOP9*	10.9	$1.4x10^{-6}$	9/13	10/142	9.8 (3.4– 28.5)
chr12:1604490 (3'UTR)	rs1046473 (A>T)	*ERC1*	29.9	$2.6x10^{-6}$	16/6	36/116	8.6 (3.1– 23.6)
chr6:159456212 (3'UTR)	rs4709265 (A>T)	*TAGAP*	7.5	$3.4x10^{-6}$	7/15	6/146	11.4 (3.4– 38.2)

The results are obtained by allelic ratio implemented in PLINK. *CASR,* calcium sensing receptor; *ISG20,* interferon stimulated exonuclease gene 20; *ABCC11,* ATP binding cassette subfamily C member 11; *CCL8,* C-C motif chemokine ligand 8; *PLEK*, pleckstrin; *SPRED1*, sprouty related EVH1 domain containing 1; STXBP3; syntaxin binding protein 3; *ERC1,* ELKS/RAB6-interacting/CAST family member 1; *NOP9*, NOP9 nucleolar protein; *TAGAP,* T cell activation RhoGTPase activating protein. *Allelic ratio is presented as ratio of minor vs. major allele in patients with (+) and without (-) acute graft versus host disease (aGvHD), grade, 2-4, for which OR (odds ratio) with CI (confidence interval) is calculated; p value reflects the difference in allelic ratio between patients with and without aGvHD.

8.7.3. Table 3. Genotypic model of top-ranking loci of aGvHD grade 2-4 in the discovery cohort.

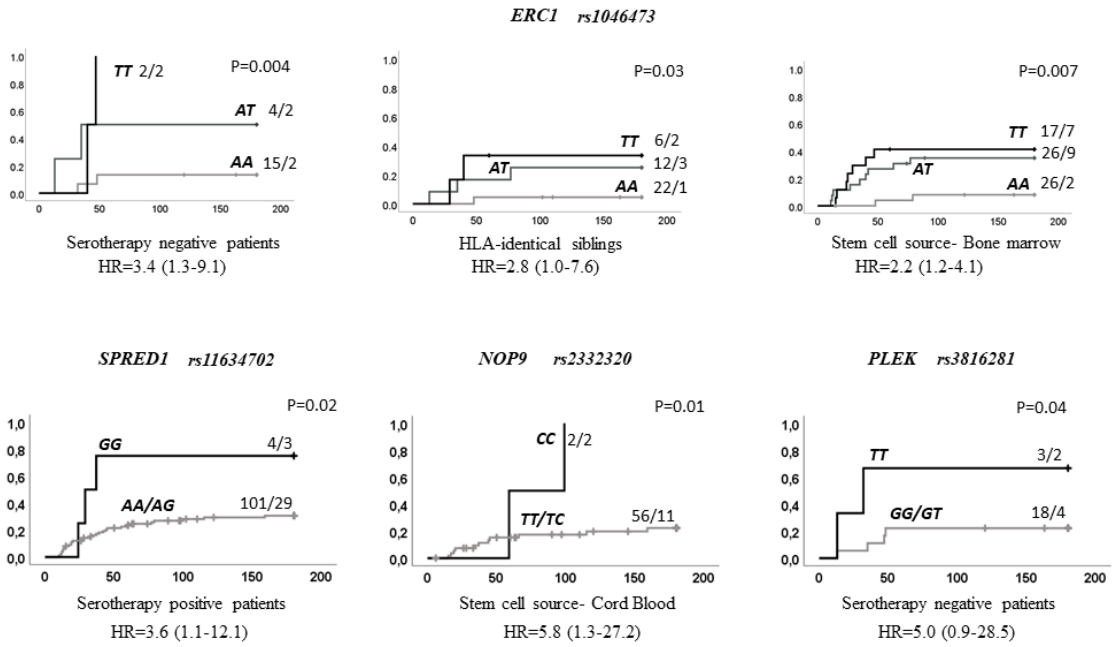
Gene and SNP	P	Model	HR (95%-CI)
CASR rs1042636 (exonic)*	$7.6x10^{-8}$	D	13.5 (3.9 – 46.3)
ISG20 rs59188950 (exonic)	$6x10^{-6}$	A	4.3 (2.1 – 9.1)
ABC11 rs17822931 (exonic)	$1x10^{-5}$	A	4.0 (2.0 – 8.2)
CCL8 rs1133763 (exonic)	$1x10^{-5}$	A	4.8 (2.2 – 10.8)
PLEK rs3816281 (exonic)	$8x10^{-6}$	R	9.2 (2.8 – 30.4)
SPRED1* rs11634702 *(UTR)	$1.6x10^{-7}$	D	12.8(3.7 – 43.8)
STXBP3* rs1124427 *(UTR)**	$4.0x10^{-6}$	A	5.4 (2.3 – 12.8)
ERC1 rs1064125 (UTR)	$7.6x10^{-7}$	A	4.8 (2.3 – 9.9)
NOP9* rs2332320 *(UTR)	$1.0x10^{-6}$	A	4.9 (2.3 – 10.4)
ERC1 rs1046473 (UTR)	$2.4x10^{-7}$	A	9.3 (3.5 – 24.9)
TAGAP* rs4709265 *(UTR)**	$1.0x10^{-5}$	A	4.4 (2.0 – 9.4)

P-KM, p value is obtained by log-rank test in Kaplan -Meier framework for the cumulative difference between genotype groups, HR (hazard ratio) with 95% CI (confidence interval) obtained through univariate Cox regression analyses, Analyses are done according to the model that best fits the data, A, additive, D, dominant or R, recessive *, there was no homozygotes individuals for minor CASR allele; ** variants that did not sustain correction for multiple testing in genotypic model ($2x10^{-6}$ for UTR variants), and not retained for the analyses in replication group.

8.7.4. Table 4. Non-genetic variables retained in stepwise selection in multivariate Cox regression model in the replication cohort.

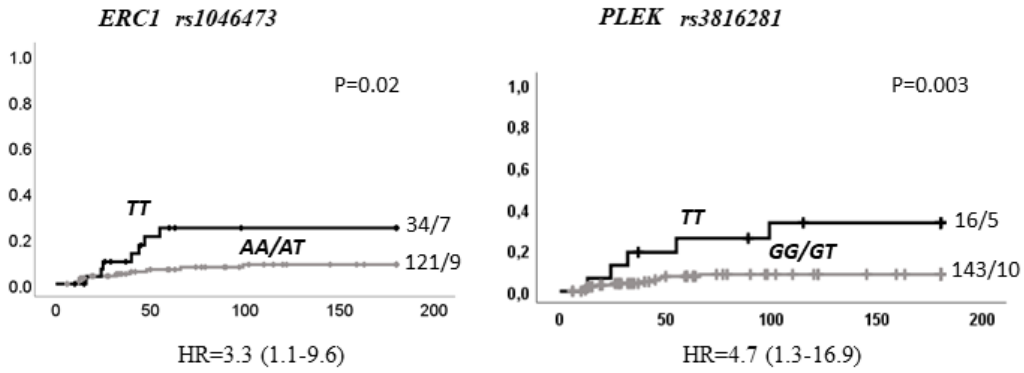
Variable	P-value	HR (95%-CI)
Donor type	0.001	5.6 (2.0-15.7)
Serotherapy	0.01	0.3 (0.1-0.8)
Age	0.02	1.07 (1.0-1.13)
Cell stem source	0.1	1.8 (0.9-3.6)

Donor type, unrelated or unmatched donors vs. HLA-identical siblings; serotherapy (yes vs. no); Stem source, peripheral blood vs others; Age is continuous variable. Other variables that were included but not retained in the final model are sex, baseline disease, conditioning regimens, and cumulative busulfan AUC.



8.7.5. Figure 1. Cumulative incidence of aGVHD 2-4 in relation to top-ranking loci identified through validation analysis of replication cohort.

Cumulative incidence of acute graft versus host disease (aGVHD), grades 2-4, plotted for indicated genotypes in particular patient subgroups. The gene name and SNP rs number are given at the top of the panels. P value is derived by log-rank test. Total number of patients in each genotype group followed by the number of patients with aGVHD 2-4 is indicated next to each curve. Hazard ratio (HR) with 95% confidence interval (in brackets) is calculated through univariate Cox regression according to recessive model for *PLEK*, *SPRED1,* and *NOP9* and according to additive genetic model for *ERC1*.



8.7.6. Figure 2. Cumulative incidence of aGVHD 3-4 in replication cohort relative to the variants in *ERC1* and *PLEK* genes.

Cumulative incidence of acute graft versus host disease (aGVHD), grades 3-4, plotted for indicated genotypes in the entire replication cohort. The gene name and SNP rs number are given at the top of the panels. P value derived by log-rank test for the difference across genotypes is indicated on the plot. Hazard ratio (HR) with 95% confidence interval (CI) in brackets is indicated below panels and is derived through multivariate Cox regression model, which included besides genotype, non-genetic factors significantly associated with aGVHD 3-4 (HLA compatibility and serotherapy, p=0.01 and 0.0002, respectively).

8.8. REFERENCES

1. Adom D, Rowan C, Adeniyan T, Yang J, Paczesny S. Biomarkers for Allogeneic HCT Outcomes. *Frontiers in Immunology*. 2020;11(673).
2. Shaw PJ, Kan F, Woo Ahn K, et al. Outcomes of pediatric bone marrow transplantation for leukemia and myelodysplasia using matched sibling, mismatched related, or matched unrelated donors. *Blood*. 2010;116(19):4007-4015.
3. Eapen M, Horowitz MM, Klein JP, et al. Higher mortality after allogeneic peripheral-blood transplantation compared with bone marrow in children and adolescents: the Histocompatibility and Alternate Stem Cell Source Working Committee of the International Bone Marrow Transplant Registry. *J Clin Oncol*. 2004;22(24):4872-4880.
4. Alsultan A, Giller RH, Gao D, et al. GVHD after unrelated cord blood transplant in children: characteristics, severity, risk factors and influence on outcome. *Bone marrow transplantation*. 2011;46(5):668-675.
5. Zeiser R, Blazar BR. Acute Graft-versus-Host Disease — Biologic Process, Prevention, and Therapy. *New England Journal of Medicine*. 2017;377(22):2167-2179.
6. Martínez-Laperche C, Buces E, Aguilera-Morillo MC, et al. A novel predictive approach for GVHD after allogeneic SCT based on clinical variables and cytokine gene polymorphisms. *Blood Adv*. 2018;2(14):1719-1737.
7. Ferrara JLM, Levine JE, Reddy P, Holler E. Graft-versus-host disease. *The Lancet*. 2009;373(9674):1550-1561.
8. Dzierzak-Mietla M, Markiewicz M, Siekiera U, et al. Occurrence and Impact of Minor Histocompatibility Antigens' Disparities on Outcomes of Hematopoietic Stem Cell Transplantation from HLA-Matched Sibling Donors. *Bone Marrow Research*. 2012;2012:257086.
9. Martin PJ, Levine DM, Storer BE, et al. Genome-wide minor histocompatibility matching as related to the risk of graft-versus-host disease. *Blood*. 2017;129(6):791-798.
10. Kim DDH, Yun J, Won H-H, et al. Multiple single-nucleotide polymorphism-based risk model for clinical outcomes after allogeneic stem-cell transplantation, especially for acute graft-versus-host disease. *Transplantation*. 2012;94(12):1250-1257.
11. Ansari M, Curtis PH, Uppugunduri CRS, et al. GSTA1 diplotypes affect busulfan clearance and toxicity in children undergoing allogeneic hematopoietic stem cell transplantation: a multicenter study. *Oncotarget*. 2017;8(53):90852-90867.
12. Ansari M, Petrykey K, Rezgui MA, et al. Genetic Susceptibility to Hepatic Sinusoidal Obstruction Syndrome in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplantation. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation*. 2020;26(5):920-927.
13. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus Conference on Acute GVHD Grading. *Bone marrow transplantation*. 1995;15(6):825-828.
14. Holtan SG, Khera N, Levine JE, et al. Late acute graft-versus-host disease: a prospective analysis of clinical outcomes and circulating angiogenic factors. *Blood*. 2016;128(19):2350-2358.
15. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
16. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7 20.

17. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.
18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
19. Benjamini YH, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological* 1995(57):289–300.
20. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone marrow transplantation.* 2013;48(3):452-458.
21. Labuda D, Krajinovic M, Richer C, et al. Rapid detection of CYP1A1, CYP2D6, and NAT variants by multiplex polymerase chain reaction and allele-specific oligonucleotide assay. *Anal Biochem.* 1999;275(1):84-92.
22. Huezo-Diaz P, Uppugunduri CR, Tyagi AK, Krajinovic M, Ansari M. Pharmacogenetic aspects of drug metabolizing enzymes in busulfan based conditioning prior to allogenic hematopoietic stem cell transplantation in children. *Curr Drug Metab.* 2014;15(3):251-264.
23. Zhu Q, Yan L, Liu Q, et al. Exome chip analyses identify genes affecting mortality after HLA-matched unrelated-donor blood and marrow transplantation. *Blood.* 2018;131(22):2490-2499.
24. Liu T, Zhang L, Joo D, Sun SC. NF-κB signaling in inflammation. *Signal Transduct Target Ther.* 2017;2:17023-.
25. Magenau J, Runaas L, Reddy P. Advances in understanding the pathogenesis of graft-versus-host disease. *Br J Haematol.* 2016;173(2):190-205.
26. Jeanette LDS, Bottero V, Young DB, Shevchenko A, Mercurio F, Verma IM. Activation of Transcription Factor NF-κB Requires ELKS, an IκB Kinase Regulatory Subunit. *Science.* 2004;304(5679):1963-1967.
27. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature.* 2009;461(7267):1071-1078.
28. Kuba A, Raida L, Mrazek F, et al. NFKB1 gene single-nucleotide polymorphisms: implications for graft-versus-host disease in allogeneic hematopoietic stem cell transplantation. *Ann Hematol.* 2020;99(3):609-618.
29. Alpay M, Backman LR, Cheng X, et al. Oxidative stress shapes breast cancer phenotype through chronic activation of ATM-dependent signaling. *Breast cancer research and treatment.* 2015;151(1):75-87.
30. Liu CJ, Fu X, Xia M, Zhang Q, Gu Z, Guo AY. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Res.* 2020.
31. Gong J, Tong Y, Zhang HM, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum Mutat.* 2012;33(1):254-263.
32. Medrano LM, Pascual V, Bodas A, et al. Expression patterns common and unique to ulcerative colitis and celiac disease. *Ann Hum Genet.* 2019;83(2):86-94.
33. Hoppmann N, Graetz C, Paterka M, et al. New candidates for CD4 T cell pathogenicity in experimental neuroinflammation and multiple sclerosis. *Brain.* 2015;138(Pt 4):902-917.

34. Lundmark A, Davanian H, Båge T, et al. Transcriptome analysis reveals mucin 4 to be highly associated with periodontitis and identifies pleckstrin as a link to systemic diseases. *Sci Rep.* 2015;5:18475.
35. Ding Y, Kantarci A, Badwey JA, Hasturk H, Malabanan A, Van Dyke TE. Phosphorylation of Pleckstrin Increases Proinflammatory Cytokine Secretion by Mononuclear Phagocytes in Diabetes Mellitus. *The Journal of Immunology.* 2007;179(1):647-654.
36. Hasturk H, Kantarci A, Van Dyke TE. Oral inflammatory diseases and systemic inflammation: role of the macrophage. *Front Immunol.* 2012;3:118.
37. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213.
38. Pasmant E, Gilbert-Dussardier B, Petit A, et al. SPRED1, a RAS MAPK pathway inhibitor that causes Legius syndrome, is a tumour suppressor downregulated in paediatric acute myeloblastic leukaemia. *Oncogene.* 2015;34(5):631-638.
39. Olsson L, Castor A, Behrendtz M, et al. Deletions of IKZF1 and SPRED1 are associated with poor prognosis in a population-based series of pediatric B-cell precursor acute lymphoblastic leukemia diagnosed between 1992 and 2011. *Leukemia.* 2014;28(2):302-310.
40. Jansen F, Yang X, Hoelscher M, et al. Endothelial microparticle-mediated transfer of MicroRNA-126 promotes vascular endothelial cell repair via SPRED1 and is abrogated in glucose-damaged endothelial microparticles. *Circulation.* 2013;128(18):2026-2038.
41. Biedermann BC. Vascular endothelium and graft-versus-host disease. *Best Practice & Research Clinical Haematology.* 2008;21(2):129-138.
42. Tichelli A, Gratwohl A. Vascular endothelium as 'novel' target of graft-versus-host disease. *Best Practice & Research Clinical Haematology.* 2008;21(2):139-148.
43. Cao X-N, Kong Y, Song Y, et al. Impairment of bone marrow endothelial progenitor cells in acute graft-versus-host disease patients after allotransplant. *British Journal of Haematology.* 2018;182(6):870-886.
44. Galli SJ, Grimbaldeston M, Tsai M. Immunomodulatory mast cells: negative, as well as positive, regulators of immunity. *Nat Rev Immunol.* 2008;8(6):478-486.
45. Leveson-Gower DB, Sega EI, Kalesnikoff J, et al. Mast cells suppress murine GVHD in a mechanism independent of CD4+CD25+ regulatory T cells. *Blood.* 2013;122(22):3659-3665.
46. Ishizaki T, Tamiya T, Taniguchi K, et al. miR126 positively regulates mast cell proliferation and cytokine production through suppressing Spred1. *Genes to Cells.* 2011;16(7):803-814.
47. Suzuki M, Morita R, Hirata Y, Shichita T, Yoshimura A. Spred1, a Suppressor of the Ras-ERK Pathway, Negatively Regulates Expansion and Function of Group 2 Innate Lymphoid Cells. *J Immunol.* 2015;195(3):1273-1281.
48. Munneke JM, Björklund AT, Mjösberg JM, et al. Activated innate lymphoid cells are associated with a reduced susceptibility to graft-versus-host disease. *Blood.* 2014;124(5):812-821.
49. Thomson E, Rappsilber J, Tollervey D. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. *RNA.* 2007;13(12):2165-2174.

50. Zhang J, Teramoto T, Qiu C, et al. Nop9 recognizes structured and single-stranded RNA elements of preribosomal RNA. *RNA.* 2020;26(8):1049-1059.
51. Nudel R, Simpson NH, Baird G, et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav.* 2014;13(4):418-429.
52. Turpeinen H, Ojala PJ, Ojala K, Miettinen M, Volin L, Partanen J. Minor histocompatibility antigens as determinants for graft-versus-host disease after allogeneic haematopoietic stem cell transplantation. *Int J Immunogenet.* 2013;40(6):495-501.
53. Akatsuka Y, Warren EH, Gooley TA, et al. Disparity for a newly identified minor histocompatibility antigen, HA-8, correlates with acute graft-versus-host disease after haematopoietic stem cell transplantation from an HLA-identical sibling. *Br J Haematol.* 2003;123(4):671-675.
54. Sharma SB, Lin CC, Farrugia MK, et al. MicroRNAs 206 and 21 cooperate to promote RAS-extracellular signal-regulated kinase signaling by suppressing the translation of RASA1 and SPRED1. *Mol Cell Biol.* 2014;34(22):4143-4164.

8.9. SUPPLEMENTAL MATERIAL

8.9.1. Supplemental Table 1. Multivariate Cox regression analyses in the discovery cohort.

Variable	P-value	HR (95%-CI)
ERC1 rs1046473	0.001	13.1 (3.1-55.5)
ERC1 rs1064125	0.001	4.6 (1.8-11.4)
ISG20 rs59188950	0.004	8.4 (2.6-27.4)
CASR rs1042636	0.0002	31.2 (5.1-191.5)
ABC11 rs17822931	0.001	4.7 (1.9-11.5)
CCL8 rs1133763	0.001	7.4 (2.4—22.8
PLEK rs3816281	0.001	11.4 (2.7-47.9)
***SPRED1* rs11634702**	0.00008	22.7 (4.8-107.4)
***NOP9* rs2332320**	0.001	16.7 (3.2-86.2)

Multivariate models analyze the effect of each genotype separately with the presence of non-genetic covariates: donor type, serotherapy, stem source, age, sex, baseline disease, conditioning regimens, and cumulative busulfan AUC. Genetic model for each locus is the same as indicated in **Table 3**. Variants that did sustain correction for multiple testing (*STXBP3* and *TAGAP*) were not included in the multivariate model.

8.9.2. Supplemental Table 2. The difference between genotype groups in the discovery cohort as obtained by competing event analysis and Gray test.

Gene	SNP	p1	p2	p3	p4
ISG20	rs59188950	$7x10^{-5}$	0.8	$7x10^{-5}$	0.6
CCL8	rs1133763	$3x10^{-5}$	0.9	$3x10^{-5}$	0.5
SPRED1	rs11634702	$3x10^{-6}$	1.0	$3x10^{-6}$	0.5
NOP9	rs2332320	$2x10^{-5}$	0.9	$2x10^{-5}$	0.5
ERC1	rs1046473	$5x10^{-7}$	0.7	$5x10^{-7}$	0.6
PLEK	rs3816281	$3x10^{-5}$	0.4	$2x10^{-5}$	0.2
ERC1	rs1064125	$2x10^{-5}$	0.6	$2x10^{-5}$	0.6
ABCC11	rs17822931	$1x10^{-4}$	0.8	$1x10^{-4}$	0.5
CASR	rs1042636	$5x10^{-7}$	0.3	$8x10^{-7}$	0.6

p1 and p3 reflect the comparison between genotype groups for aGvHD 2-4 with inclusion of a death, or death and relapse, respectively, as a competing event. p2 and p4 are the significance of comparison across categories of corresponding competing event. The graphs obtained by competing event analyses are presented in Supplemental Figure 2. The genetic model is the same as presented for 1-KM analyses for each of these SNPs in **Supplemental Figure 1.**

8.9.3. Supplemental Table 3. Analysis of top-ranking loci of aGvHD grade 2-4 in the replication cohort.

Locus	SNP	Gene	Functional impact	P KM	Minor allele carriers* Event+	Minor allele carriers Event-	P KM**	Model
chr3:122003769 (exonic)	rs1042636 (A>G)	*CASR*	Arg990Gly[1]	0.35	2/41	2/115	0.5	R
chr15:89195245 (exonic)	rs59188950 (C>T)	*ISG20*	His34Tyr[2]	0.37	1/43	7/114	0.62	R
chr16:48258198 (exonic)	rs17822931 (A>G)	*ABCC11*	Gly180Arg[1]	0.18	0/44	6/115	0.34	R
chr17:32647831 (exonic)	rs1133763 (A>C)	*CCL8*	Lys69Gln[1]	0.22	10/45	39/119	0.35	D
chr2:68607947 (exonic)	rs3816281 (G>T)	*PLEK*	Lys97Asn[1]	0.17	7/43	9/116	0.3	R
chr15:38647386 (3'UTR)	rs11634702 (A>G)	*SPRED1*	gain or loss of miR target[3]	0.07	3/40	2/107	0.77	R
chr12:1600978 (3'UTR)	rs1064125 (A>T)	*ERC1*	gain of miR target[3]	0.49	3/43	11/112	0.25	R
chr14:24776219 (3'UTR)	rs2332320 (T>C)	*NOP9*	gain or loss of miR target[3]	0.29	10/41	38/116	0.52	D
chr12:1604490 (3'UTR)	rs1046473 (A>T)	*ERC1*	gain or loss of miR target[3]	0.1	28/43	58/112	0.09	D

The chromosome position of the variant, type of nucleotide substitution, dbSNP identification number, gene name and functional impact are provided. [1]damaging impact of amino-acid substitution as identified by SIFT or PolyPhen; [2]only identified as missense variant in particular isoforms; [3]change in miR (microRNA) binding sites, as identified by miRNASNP-v3 database (30, 31). P-KM, p value is obtained by log-rank test in Kaplan-Meier framework in entire cohort; * Minor allele carriers according to indicated genetic model among patients with (+) and without (-) event. Analyses are done according to the model that best fits the data; A, additive, D, dominant or R, recessive. ** p value obtained by log-rank test in Kaplan-Meier framework when the analyses are restricted to Caucasians only.

8.9.4. Supplemental Table 4. Univariate Cox regression analyses for non-genetic covariates of the replication cohort.

Variable	P-value	HR (95%-CI)
Sex	0.5	0.8 (0.4-1.5)
Age	0.08	1.05 (1.0-1.1)
Baseline disease	0.03	2.0 (1.1-3.8)
Donor type	0.02	2.7 (1.2-6.4)
Serotherapy	0.9	0.9 (0.4-2.1)
Conditioning regimen	0.5	0.8 (0.4-1.7)
Cell stem source	0.05	1.9 (1.0-3.7)
Cumulative Bu AUC	0.1	1.02 (1.0-1.04)

These covariates included age as a continuous variable, sex (male vs females), baseline disease (hematological malignancies vs non-malignant disease), type of a donor (HLA identical related individuals vs other donors); whether patients received (or not) serotherapy, type of conditioning regimen categorized according to the number of alkylating agents (more than one vs one), to which total body irradiation (TBI) was added as the "equivalent" of one alkylating agent, stem cell source (peripheral blood stem cell vs. bone marrow or cord blood) and busulfan cumulative area under the curve (cumAUC, mg×h/L), estimated from the first dose AUC and each individual dose received.

8.9.5. Supplemental Table 5. Correlation between stem cell source and donor type.

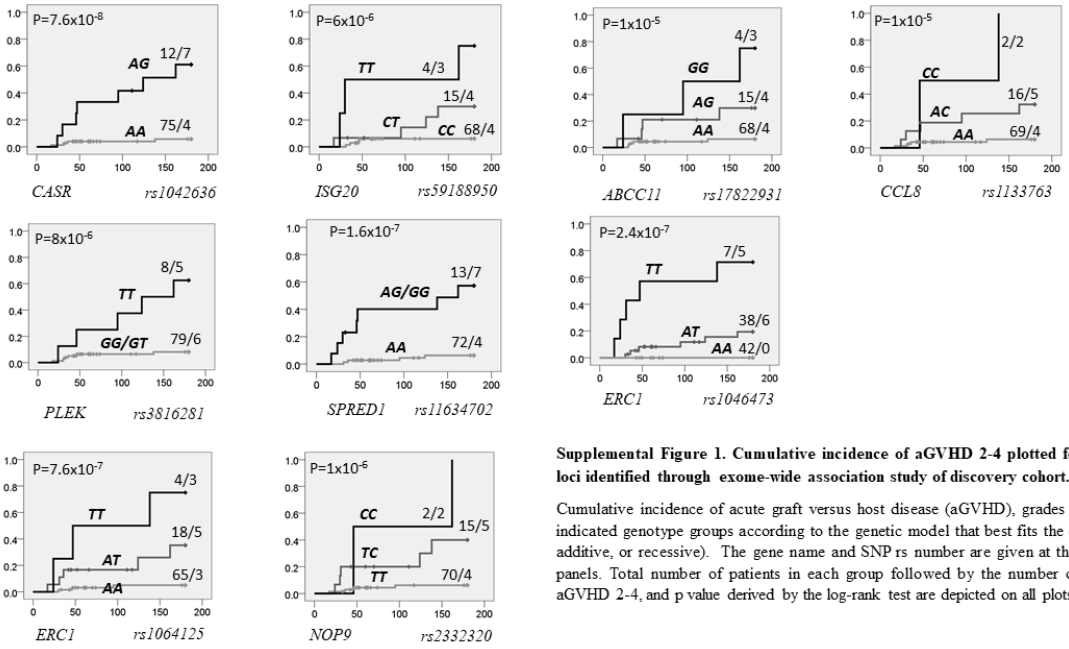
			Stem Cell Source		
			Peripheral or cord blood	Bone Marrow	Total
Donor type	HLA identical siblings	Number	15	29	44
		%	34.1%	65.9%	100.0%
	Remaining donors	Number	79	45	124
		%	63.7%	36.3%	100.0%
Total		Number	94	74	168
		%	56.0%	44.0%	100.0%

High frequency of bone marrow grafts was noted among HLA identical siblings, p=0.001.

8.9.6. Supplemental Table 6. The difference between genotype groups in the replication cohort as obtained by competing event analysis and Gray test.

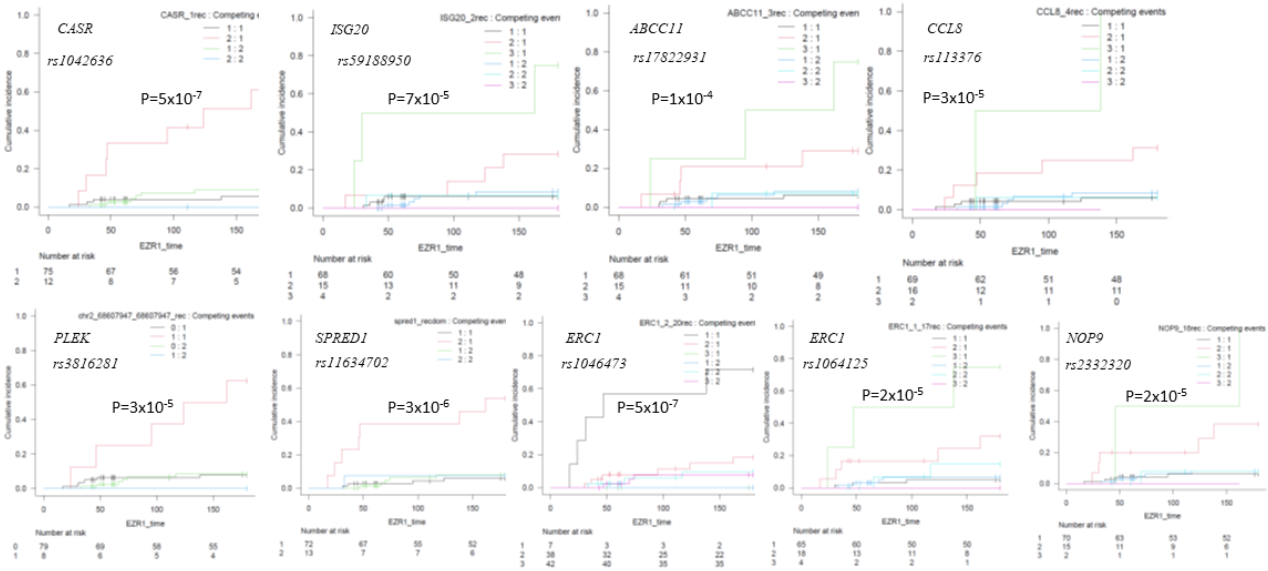
Gene	SNP	Association group	p1	p2
ERC1	*rs1046473*	HLA-identical siblings	0.04	0.7
		Serotherapy negative patients	0.01	-
		Stem cell source- Bone marrow	0.005	0.6
PLEK	*rs3816281*	Serotherapy negative patients	0.06	-
SPRED1	*rs11634702*	Serotherapy positive patients	0.03	0.5
NOP9	*rs2332320*	Stem cell source- Cord blood	0.008	0.7
ERC1	*rs1046473*	entire cohort*	0.02	0.7
PLEK	*rs3816281*	entire cohort*	0.003	0.2

P1 reflect the analysis between genotype groups for aGVHD2-4 with inclusion of a death as a competing event. p2, difference between genotype group for a competing event. Association group is indicated and is same as in Figures 1 and 2 when using corresponding 1-KM analyses for each SNP, - absence of competing events.



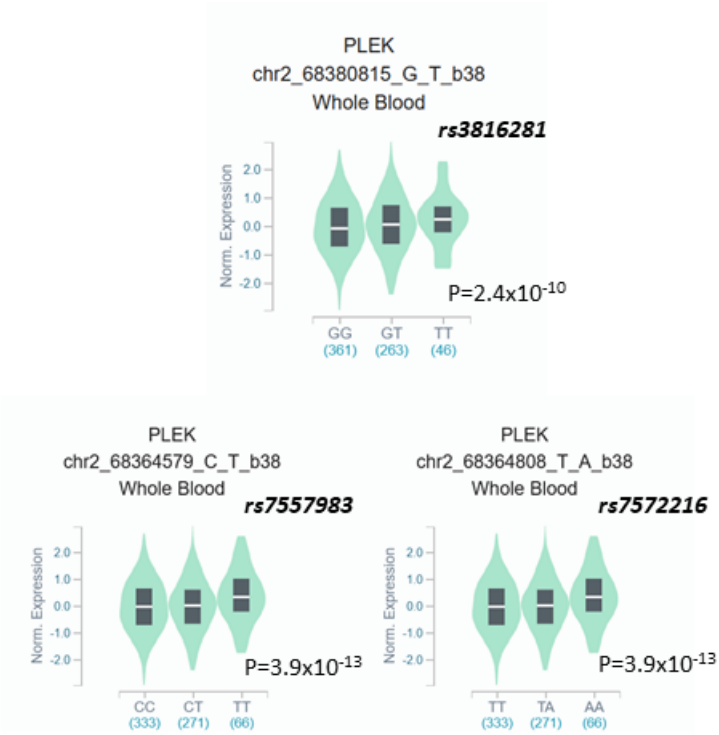
Supplemental Figure 1. Cumulative incidence of aGVHD 2-4 plotted for top-ranking loci identified through exome-wide association study of discovery cohort.

Cumulative incidence of acute graft versus host disease (aGVHD), grades 2-4, plotted for indicated genotype groups according to the genetic model that best fits the data (dominant, additive, or recessive). The gene name and SNP rs number are given at the bottom of the panels. Total number of patients in each group followed by the number of patients with aGVHD 2-4, and p value derived by the log-rank test are depicted on all plots.



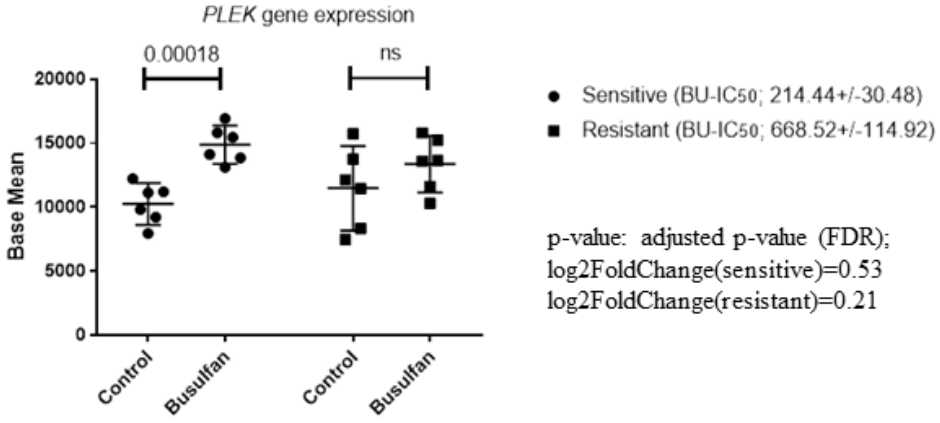
Supplemental Figure 2. Cumulative incidence of aGVHD 2-4 plotted for top ranking loci identified through exome-wide association study of discovery cohort using competing event analyses.

Cumulative incidence of acute graft versus host disease (aGVHD), grades 2-4, plotted for indicated genotype groups using competing event analyses and same genetic models as in Supplemental Figure 1. The gene name, SNP rs number, and p values are indicated on the plots. Further details of the analyses are also summarized in Supplemental Table 2. Labeling of the curves: the first number indicates genotype groups, and the second number indicates aGvHD (1) or competing event (death, 2).



Supplemental Figure 3. Relationship of the *PLEK rs3816281* and two promoter SNPs in LD with the mRNA expression.

Representative examples depicting significant change in expression according to the *PLEK rs3816281* genotype, as well as *PLEK rs7557983* and *rs7572216* (two promoter SNPs in linkage disequilibrium), as obtained from public database: The Genotype-Tissue Expression (GTEx) Project.



Supplemental Figure 4. Relationship of the *PLEK* expression with Busulfan (BU) IC50 in lymphoblastoid cell lines (LCLs).

Change in the *PLEK* gene expression as obtained by RNAseq data after Bu treatment of LCLs that are sensitive (circles) or resistant (squares) to treatment. Mean change relative to the expression prior to treatment (controls) is indicated along with the P value adjusted for multiple testing by false discovery rate. Log2Fold change is indicated below the graph.

BU-IC50 (concentration reducing the starting population of cells to 50%), defining sensitive and resistant LCLs, are indicated next to the graph.

8.9.7. Item S1

Demographic and treatment characteristics of replication cohort related to Table 1.

Hematologic Malignancies include Acute Lymphoblastic Leukemia (n=3); Acute Myeloid Leukemia (n=27); Myelodysplastic Syndrome (n=13); Chronic Myeloid Leukemia (n=2). Non-Malignancies include Hemoglobinopathy (n=13), Immunodeficiency (n=12), Metabolic Disease (n=5), Chronic Granulomatous Disease (n=7) and Hemophagocytic Syndrome (n=5).

Among 42 patients who received a cord blood transplant, there were 3 patients who received double cord blood.

Patients who received myeloablative regimen with reduced toxicity (n=16), received a Bu/Flu combination with a cumulative dose of Bu >8mg/kg.

Prophylaxis of GvHD include cyclosporine (CSA, n=1), CSA and steroids (n=30), CSA and methotrexate (MTX, n=30), CSA and Mycophenolate mofetil (MMF, n=21), CSA, MMF and steroids (n=2) and MTX and tacrolimus (n=2).
None of patients received total body irradiation (TBI).

Demographic and treatment characteristics of replication cohort related to Table 1.

Hematologic Malignancies include Acute Lymphoblastic Leukemia (n=26); Acute Myeloid Leukemia (n=37); Biphenotypic Acute Leukemia (n=2); Myelodysplastic Syndrome (n=30); Chronic Myeloid Leukemia (n=2); Lymphoma (3).

Non-Malignancies include Bone Marrow Failure Syndrome (n=5), Hemoglobinopathy (n=12), Immunodeficiency (n=19), Metabolic Disease (n=22), Chronic Granulomatous Disease (n=1), Hemophagocytic Syndrome (n=10) and Osteoporosis (n=3).

Among 63 patients who received a cord blood transplant, there were 3 patients who received the double cord blood.

Patients who received myeloablative regimen with reduced toxicity (n=50) received a Bu/Flu combination with a cumulative dose of Bu >8mg/kg.

Prophylaxis of GvHD include (CSA, n=16), CSA and steroids (n=39), CSA and MTX (n=77), CSA and tacrolimus (n=1), CSA, MTX and tacrolimus (n=1), CSA, MTX, tacrolimus and steroids, (n=1) CSA and MMF (n=18), CSA, MMF and tacrolimus (n=2), CSA, MMF, MTX and tacrolimus (n=1) CSA, MMF and steroids (n=2), MTX and tacrolimus (n=4), MMF and tacrolimus (n=2), data not available (n=4).

Patients from Alberta Children's Hospital, Calgary (n=28) received a TBI, 4 Gy given in 2 fractions of 2 Gy each. Of these, 24 patients received myeloablative regimen with reduced toxicity and 4 patients received non-myeloablative regimen.

Section D

Chapter 9

General Discussion

In this chapter, I highlight the most important insights that can be gleaned from the various sections presented in this thesis. I point out the limitations that may have influenced the obtained results. In conclusion, I outline the most important research findings presented in this dissertation.

9.1. General Discussion

9.1.1. Discussion of Section A

Genetic variations in the DNA sequence affect the risk of developing many diseases and/or complex traits. Early studies that investigated genetic variations underlying rare conditions with clear Mendelian patterns of family segregation (e.g., Huntington's disease, cystic fibrosis) were very successful in detecting these genetic variations, primarily because these variations carried a 100% risk and were the only cause of the disease[1]. Typically, difficulties arise when a simple match between genotype and phenotype is violated, as either the same genotype can lead to different phenotypes (due to chance, incomplete penetrance, the environment, or the interaction with other genes) or different genotypes can lead to the emergence of different phenotypes[2]. Thus, the term "complex trait" was proposed by geneticists to denote any phenotype that does not exhibit recessive or dominant Mendelian inheritance attributable to a single gene[2]. It is therefore not surprising that finding genetic variants underlying common "complex diseases" (e.g., type 2 diabetes, cardiovascular disease, many types of cancer) has proven to be much more challenging. This was later explained by the fact that each variant individually is only one of many genetic and environmental causal factors, each of which is neither necessary nor sufficient to cause disease[1]. Therefore, they predispose rather than directly lead to its development[1].

Genetic variants that result in a slight increase in the risk of a common condition can still have significant public health implications (for example, in terms of the number of people affected by them); thus, identifying such variants is very important. Moreover, such findings may reveal new causal relationships worthy of further investigation[1].

Next-generation sequencing (NGS) platforms and large-scale genome-wide association studies (e.g., microarrays for copy number alterations), as well as whole transcriptome analysis, exhibit

nowadays an unprecedented ability to classify new molecular subgroups of currently known conditions based on their gene expression profiles[3]. For instance, this approach has revealed new oncogenic drivers of leukemogenesis. Many of them have been shown to have prognostic and/or therapeutic value[3]. From a diagnostic standpoint, elucidating a molecular diagnosis can provide clinical utility by predicting prognosis, anticipating future symptoms, allowing early intervention, and identifying treatment options, while avoiding inappropriate interventions[4]. For example, since the early 2000s, large-scale genomic research on ALL, greatly advanced by the NGS, has led to the development of a new taxonomy for ALL, identifying new subsets of ALL (characterized by "driver" oncogenic changes) previously hidden for conventional methods of karyotyping[3].

On the other hand, the cure rate for many pediatric cancers (including ALL) has reached a remarkable incidence (has exceeded 90% in some contemporary clinical trials)[5]. However, the dose intensity of conventional chemotherapy has been pushed to its limit; therefore, further improvement in outcomes will need to rely more on molecular therapeutic approaches (including immune and cell therapy) as well as on precise risk stratification[5-7].

Furthermore, a significant number of genes have been identified through association studies of treatment-related outcomes in cancer patients. These results may well lead to predictive models capable to identify patients who might be at increased risk for specific treatment-related side effects (both acute and long-term) and therefore require treatment adjustment or closer monitoring (or have lower susceptibility, therefore may not need additional follow-ups).

In addition, special attention is currently being paid to the contribution of rare and low-frequency variants to human traits and diseases[8]; in turn, they could contribute significantly to our understanding of the role of the genetic component in the pathophysiology of toxicities associated with chemotherapy treatment.

9.1.2. Discussion of Section B

The neurotoxic effects of cancer treatments consist of peripheral neuropathy and central neurotoxicity, characterized by encephalopathy and/or neurodevelopmental cognitive deficits, are well documented in childhood ALL survivors[9]. These effects have been mainly associated with cranial radiotherapy (CRT) and intrathecal methotrexate (MTX)[10]. Furthermore, several psychiatric syndromes have been associated with cranial radiotherapy (CRT)[11,12] and intensive treatment[13-15]. With regard to psychoaffective status CRT alone or in combination with MTX has been associated with higher incidence of depression/anxiety, degraded social skills, attention deficits and antisocial behaviours[16]. With different dosages depending on the risk status, the DFCI protocols implicate the administration of corticosteroids (CS) in the induction phase for intensification and continuation of treatment. An intensified CS treatment could also affect emotional regulation (depression and suicidal tendencies,[17] anxiety [17,18], and externalizing problems[19 20]), in the short term and to have longer-term effects[21]. For instance, treatments with high-dose CS increase the neurotoxicity of MTX and CRT[22,23]. By their action on the hippocampus and limbic system, CS treatments yield a dysregulation of levels of endogenous CS and might affect the hypothalamic-pituitary-adrenal axis response (HPA axis), and have an impact on the actual structure of the HPA axis[24,25] giving rise to chronic stress [26] and depression[17,18]. These detrimental effects seem to vary depending on treatment intensity and the nature of the CS. For example, the most important psychoaffective effects among young and older children were noted with *Dexamethasone* (vs. *Prednisone*)[19,27]. Therefore, it is possible that adjustments introduced in therapies over time (changes in CRT doses, as well as CS: substitution of *Dexamethasone* for *Prednisone*)[28] might differentially impact the level of emotional adjustment in the long-term[29]. It was also suggested that manifestation of long-term treatment-related side effects (rather than the

remote cancer diagnosis or treatment history) could affect, the emotional well-being of childhood cancer survivors[30], with a cancer-related pain[31,32] and learning or memory problems[33,34] being between the most relevant late effects, but potentially amenable to intervention[30].

In addition, available epidemiological and neuroimaging results indicate that ALL survivors who received contemporary therapeutic protocols (which consist of intensified intravenous and intrathecal administration of chemotherapeutic drugs for standard risk patients[35]) were still at risk of neurocognitive problems[36]. Future research should then move from simply describing and quantifying the prevalence of these adverse outcomes to detailing the underlying pathophysiological processes so that early detection and preventive measures can be assessed[36].

In our study, we addressed the genetic component in the neurocognitive impairments along with anxiety and depression by first interrogating the relationship between the above-mentioned complications and genotypic profiling of candidate genes obtained through whole exome sequencing (WES) of childhood ALL survivors. We identified a panel of common variants that were associated with deficits in neurocognitive tests performance and variants that were associated with anxiety and depression. Then, the association analyses were complemented by an exome-wide association study, which identified several additional genes as potential modulators of the risk of developing treatment-related neurocognitive complications as well as anxiety and depression.

Few genes are of particular interest since these associations were validated in the replication SJLIFE Cohort (USA). For instance, the *MTR* gene that is involved in the metabolic pathway of MTX. The detected common functional variant *rs1805087* potentially affects enzymatic activity, therefore increasing the level of homocysteine [10,37,38], which is the essential in methionine production[39]. This variant along with other polymorphisms implicated in the Hcy pathway were

already studied in the context of MTX long-term neurotoxicity and has been found to affect neurocognitive function in childhood ALL survivors[40,41]. Another gene identified by the candidate gene approach is the *CACNB2* gene that is highly expressed in hippocampal pyramidal neurons and linked to the Bipolar disorders[42], autism spectrum disorders[43] and social cognition in schizophrenia patients[44]. Moreover, although this gene has not been confirmed in the replication SJLIFE Cohort, the *SLCO1B1* gene also deserves further attention since previously shown to affect MTX clearance and short-term toxicity in ALL patients. Additionally, the association between deficit in the trail making test and variant rs61732180 in the *ZNF382* gene is also worth mentioning since association was found across the entire replication SJLIFE Cohort (despite a non-significant male-specific association). All confirmed associations identified by candidate genes and exome-wide approaches are summarized in **Table 1a**. Furthermore, our analysis also suggests that synergistic interactions might exist between the variants identified in the entire discovery cohort (rs740965 *PTPRZ1*, rs2748431 *MUC16*, and rs2748431 *TNRC6C-AS1*) and in the group of male participants (rs7285167 *APOL2*, and rs61731441 *C6orf165*) in relation to moderate-severe anxiety; which could explain the markedly significant associations in the combined variants model.

Interestingly, although several validated genes described in this thesis are linked to cancer treatment, only two confirmed variants - rs1805087 *MTR* and rs4149056 *SLCO1B1*, have a direct impact on the pharmacodynamics and pharmacokinetics of the anticancer agent MTX, respectively (**Table 1a)**.

Multiple evidence has been collected nowadays for potential genetic and epigenetic risk markers of the long-term treatment-related neurocognitive and emotional complications in survivors of childhood cancer. In addition, accumulating data suggest that genetic factors contribute

significantly to resilient responses to trauma and stress[45]. However, large genome-wide association studies on the genetic architecture of mental disorders indicate its polygenic nature, which means that the contribution of genetic factors is due to small effects of several genetic variants distributed throughout the genome[46-48]. Therefore, future studies will be required not only for independent replications, but also for evaluation of the potential prognostic and/or therapeutic value of the reported findings in order to implement them into clinical practice.

In a similar fashion to our first study in the PETALE cohort presented earlier, we carried another analysis that focused on treatment-related cardiac complications. Cardiovascular diseases represent the main cause of secondary morbidity and mortality in childhood cancer patients [49-51]. Anthracyclines are an essential component of childhood cancer therapy, with *Doxorubicin* being the most widely used agent, administered in a wide spectrum of hematologic and solid malignancies, including leukemia, lymphoma, and sarcomas[52-54]. Approximately 50–60% of childhood cancer survivors have been treated with an anthracycline regimen[49,54-56]. Major complication of the Anthracycline toxicity (ACT) include cardiomyopathy, coronary artery disease, and atherosclerosis[57]. Furthermore, at the present time, there is no imaging approach or specific guidelines to prospectively identify patients who are at risk of ACT[58].

We performed candidate gene analyses regarding treatment-related cardiac complications which identified several novel markers associated with these toxicities in the. In addition, through an exome-wide association study, the contribution of the rare variants genes was linked to the risk of developing treatment-related cardiotoxicity in survivors of childhood ALL.

It is worth mentioning that even though identified genetic markers were not validated in the SJLIFE cohort (all the details of the replication results were discussed earlier), several identified loci have already been associated with treatment-related cardiotoxicity whereas others have been

shown to play an important role in cardiovascular biology. For example, we detected a cardioprotective effect of the three common independent variants in the *TTN* gene (a gene strongly associated with cardiomyopathy) for both LVEF and FS M-mode outcomes.

In addition, in this study, we demonstrated the endowment of rare genetic variants to long-term treatment-related cardiotoxicity in childhood ALL survivors; contributing to knowledge about the implication of rare variants that usually remains limited in candidate gene and GWAS studies. Therefore, the current results nevertheless provide further evidence for the genetic contribution to treatment-related cardiovascular outcomes in patients with ALL.

Furthermore, we reviewed published pharmacogenomic markers related to ACT affecting childhood cancer patients and included a brief overview of the applicability of reported findings to the PETALE cohort, validating several of them. Interestingly, almost all evaluated genes had at least one of the variants significantly associated with tested outcomes in the PETALE cohort. At the same time, it should be mentioned that only few of them passed a false discovery rate threshold (FDR lower or equal to 5%). Among such variants, are common SNPs in *ABCC1, SLC22A7* and *SPG7* genes (more details can be found in **Table 1b**). It is worth noting that the *ABCC1* gene requires special attention, as the associations reported in our study lend additional weight to this gene, which seems to play an important role in ACT in pediatric cancer patients (as described in **Section b, Chapter 6**). Moreover, our results conducted in leukemia patients treated with *doxorubicin* replicated the risk variant rs2229774 in the *RARG* gene (currently referred for pharmacogenomic testing in the Evidence-Based Clinical Practice Guidelines). Therefore, these results not only validated the genes identified by previous studies but, in some cases, pointed to potentially causal variants.

Mortality from childhood ALL has dropped dramatically since the introduction of effective chemotherapy combinations. The main challenge now is to further minimize complications from the disease and treatment, as well as improve the quality of life for ALL survivors[36]. The identification of genetic markers associated with high or low risk of treatment-related complications in childhood cancer survivors (together with other known risk factors), could enable their mitigation through individualized treatment approaches or development of new prevention, intervention, and follow-up strategies. In addition, given the complex polygenic structure of cardiovascular, neurocognitive, and emotional disorders, it could be important to evaluate the synergistic interactions between the reported variants and genes in order to identify the most significant risk and/or protective combinations.

9.1.2.1. Table 1. Summary table of risk-associations confirmed by replication analysis presented in the context of the thesis (PETALE Cohort).

a. Neurocognitive deficits, anxiety, and depression, PETALE Cohort.

Study design	Outcome	Variant	Gene	Gene Function	Effect on PK/PD
Neurocognitive deficits, anxiety, and depression					
Candidate genes	Digit span	rs58225473	***CACNB2***	auxiliary voltage-dependent subunit of L-type calcium-channel, mainly expressed in brain and heart tissue	n/a
	Verbal fluency	rs1805087	***MTR***	B12 dependent methionine synthase involved in remethylation of homocysteine (Hcy), is required to maintain adequate intracellular methionine, intracellular folate, and normal homocysteine levels; studied in the context of MTX long-term neurotoxicity	rs1805087 variant may reduce enzymatic activity (affects PD)

	Moderate-severe depression	rs4149056*	***SLCO1B1***	liver-specific member of the organic anion transporter family involved in hepatic uptake of MTX	rs4149056 is associated with low MTX clearance (affects PK)
Exome-wide association study	Trail making test	rs61732180	***ZNF382***	transcription inhibitor, inhibits the activating protein 1 (AP-1) and nuclear factor kappa-B (NF-kB) signaling	n/a

b. Chemotherapy-induced cardiotoxicity reported common variants replicated in the PETALE Cohort.

Study design	Outcome	Variant	Gene	Gene Function	Effect on PK/PD
Chemotherapy-induced cardiotoxicity					
Replication of reported findings	LVEDD z score, LVEF 2D (%) binary	rs246232	***ABCC1***	ATP Binding Cassette (ABC) transporter-coding gene mediating efflux of multiple anticancer drugs and other xenobiotics, their metabolites, and multiple other bioactive molecules	n/a
	LVEF Mmode (%) binary	rs70953680	***SLC22A7***	Solute carrier (SLC), sodium-independent organic anion transmembrane transporter	n/a
	LVEF Mmode (%) binary	rs4149056	***SPG7***	regulator of the mitochondrial permeability transition pore (PTP)	n/a

PK: pharmacokinetics; PD: pharmacodynamics; n/a: data on the role of the variant on PK or PD are not available.

CACNB2: Calcium Voltage-Gated Channel Auxiliary Subunit Beta2; *MTR*: 5-Methyltetrahydrofolate-Homocysteine Methyltransferase; *SLCO1B1*: Solute Carrier Organic Anion Transporter Family Member 1B1. *ABCC1: ATP Binding Cassette Subfamily C Member 1; SLC22A7: Solute Carrier Family 22 Member 7; SPG7: Spastic paraplegia type 7.*

* This association was not replicated in the SJLIFE cohort, but was added to the summary table due to its proven association with individual variability in high-dose MTX clearance. Its role as a predictor of short-term toxicity after MTX treatment has been confirmed in other independent cohorts.

9.1.3. Discussion of Section C

HSCT usually involves preparatory or conditioning regimens that include chemotherapy and/or high-dose radiation therapy[16]. These regimens, as well as other pre- and post-HSCT treatments (such as immunosuppressants to prevent graft-versus-host disease (GVHD)), can damage a patient's organs and tissues and induce both acute and long-term complications[59]. The most common post-HSCT complications include mucositis, sepsis, haemorrhagic cystitis, endothelial damage, pulmonary toxicity, sinusoidal obstruction syndrome and acute graft versus host disease (aGVHD) among others[59,60].

Using WES data of pediatric patients who underwent allogeneic HSCT, we performed the genetic association studies and addressed two of the most life-threatening complications – hepatic sinusoidal obstruction syndrome (SOS), also known as veno-occlusive disease (VOD) and aGVHD. A summary table of variants supported by replication studies (for both SOS and aGVHD) can be found in **Table 2**.

Our analyses identified few significant loci, notably in the *BHLHE22*, *KIAA1715,* and *UGT2B10* genes, that were associated with an increased risk of SOS; also, through the interaction with non-genetic factors variants *ERC1* rs1046473*, PLEK* rs3816281, *NOP9* rs2332320 and *SPRED1* rs11634702 were associated with higher aGVHD incidence; all these variants were validated in the independent replication cohort of pediatric patients who underwent allogeneic or autologous HSCT (**Table 2**). Moreover, the combined effect model of risk alleles of *KIAA1715, UGT2B10* and *GSTA1* (GSTA1 haplotype groups classified as slow vs. fast and normal metabolizers, were included in the model based on the previous studies showing the higher risk of SOS for GSTA1 diplotypes) was tested in a more intensive conditioning regimen, demonstrating the increased SOS in additive manner from 3.9 (95% CI=1.4-10.4, p=0.007) in patients with one risk genotype at any

locus (group 1), to 9.8 (95% CI=2.8-33.8, p=0.0003) in patients with two risk genotypes (group 2) when compared to those with no risk genotypes (group 0).

It is important to note that dose adjustment of BU after therapeutic drug monitoring contributes to a better outcome of HSCT[61]. Genotype-guided BU dose adjustments can contribute to further improvement. Since BU is eliminated by conjugation with glutathione catalyzed by glutathione S-transferase enzymes (GST) primarily by the Glutathione S Transferase Alpha1 (GSTA1) isoform followed, to a lesser extent, by the Glutathione S Transferase Mu 1 (GSTM1) and Glutathione S Transferase Pi 1 (GSTP1)[62]; several genetic polymorphisms have been studied, including variants within the genes for glutathione S transferase[61]. One of the recent reports revealed that *GSTA1* diplotypes (conferring slow metabolizing capacity) were associated with a higher incidence of SOS, aGVHD, and combined treatment-related toxicity[61]. However, it is unclear whether *GSTA1* affects SOS beyond its influence on BU clearance[61,63].

In our study variants in the *KIAA1715,* and *UGT2B10* genes were retained in multivariable model, while controlling for other risk factors such as intensity of conditioning regimen, disease indication, Bu exposure, TBI and *GSTA1* haplotypes. It is worth mentioning that most of the published association studies for SOS have used a candidate-gene approach[63]; while our study is the only one currently known that used an exome-wide approach and had a validation in the independent replication cohort[63]. In conclusion, the genetic markers uncovered in our study, along with other known risk factors, may lead to prediction models capable to identify children who might be highly susceptible to SOS in a HSCT setting and could potentially benefit from early prophylactic intervention.

It is believed that several genes may be involved in GVHD, however the most characterized genetic system is that of the human major histocompatibility complex (MHC) system located on

chromosome 6[64]. In addition, particular HLA genotypes also serve as prognostic indicators of GVHD[64]. Moreover, the limited utility of a single SNP as a predictive biomarker was observed; whereas models that included clinical and genetic variables from the cytokine genes predicted severe aGVHD significantly better than models including only clinical variables or only genetic variables[65]. Therefore, it would be interesting to assess whether the polygenic risk score for aGVHD can be further improved with the loci discovered in this study.

9.1.3.1. Table 2. Summary table of risk-associations for Sinusoidal Obstruction Syndrome (SOS) and acute graft versus host disease (aGVHD) confirmed by replication analysis, Busulfan cohort.

Study design	Outcome	Variant	Gene	Gene Function	Effect on PK/PD
Exome-wide/UTR adjacent association study	**Sinusoidal obstruction syndrome (SOS)**				
	Sinusoidal obstruction syndrome (SOS)	rs17146905	***UGT2B10***	N-glucuronidation of amine-containing compounds (mostly expressed in liver), involved in the detoxification of a variety of endogenous and exogenous compounds, including many hormones, drugs and carcinogens	n/a
		rs16931326	***BHLHE22***	basic helix-loop-helix (bHLH) family of transcription factors that regulate cell fate determination, proliferation, and differentiation45	n/a
		rs2289971	***KIAA1715***	endoplasmic reticulum (ER) junction formation factor, ubiquitously expressed in number of tissues	n/a
	acute graft versus host disease (aGVHD)				
	acute graft versus host disease (aGVHD)	rs1046473	***ERC1***	participates in the canonical DNA damage response (DDR) signaling pathway	n/a
		rs3816281	***PLEK***	the major protein kinase C substrate of platelets	n/a

		rs2332320	***NOP9***	nuclear protein similar to the PUF family of RNA-binding proteins	n/a
		rs11634702	***SPRED1***	tumor suppressor and is a negative regulator of the RAS MAPK pathway38	n/a

PK: pharmacokinetics; PD: pharmacodynamics; n/a: data on the role of the variant on PK or PD are not available.

UGT2B10: UDP Glucuronosyltransferase Family 2 Member B10; BHLHE22: Basic Helix-Loop-Helix Family Member E22; KIAA1715: Lunapark, ER Junction Formation Factor; ERC1: ELKS/RAB6-interacting/CAST family member 1; PLEK (pleckstrin); NOP9: Nucleolar Protein 9; SPRED1 (sprouty-related, EVH1 domain-containing protein 1).

9.2. Limitations

The present studies have certain limitations:

- First, the limited sample size of the discovery cohorts (especially of the Busulfan project) may affect the accuracy of the results (for example increase the possibility of finding false-positive associations), particularly in the context of the stratified analysis and different candidate gene groups.

- The association results obtained for rare variants should be taken with caution given their low number.

- Other unmeasured factors in these studies, for example, inflammation and oxidative stress, could modulate or potentiate associations with genetic factors.

- Furthermore, among associations detected in the PETALE cohort only few showed a similar trend in the replication SJLIFE cohort. This could be explained by several reasons. Despite the application of similar inclusion/exclusion criteria, matching outcomes, and patients' characteristics between the two cohorts, differences in treatment protocols or time of ALL diagnosis[67-70] may contribute to the observed discrepancies. Likewise, in some cases, only a limited number of discovery rare variants within a locus of interest, particularly those with extremely low minor allele frequencies, passed quality control, were not monomorphic in SJLIFE and were therefore considered for replication.

- Although the analyses were corrected for multiple testing (in addition, in the PETALE cohort confounding was reduced due to homogeneous population and uniform treatment) we cannot disregard the possibility that some of the associations observed in the discovery cohorts could have been obtained by chance.

- Finally, it should be noted that the functional role of the reported genetic markers, in particular those that have been successfully validated in replication cohorts, must be evaluated.

9.3. Prospective studies

Further research is needed to confirm whether the described genetic markers may be useful in identifying patients at increased risk for treatment-related complications (both acute and long-term). Besides, it would certainly be useful to consider groups of genes and pathways instead of focusing on individual variants. Indeed, the current understanding of the genetic contribution to treatment-related complications is that in most cases it is of a polygenic nature, comprising a number of genetic variants (or polymorphisms), each of which has a small effect (possibly with varying degrees of effect) on the outcome of interest. Polygenic scores are commonly used to assess the cumulative effect of an identified genetic variation on a phenotype of interest. Therefore, the confirmed genetic markers presented in this dissertation can be evaluated along with already confirmed markers from various studies to build better predictive models capable of identifying high-risk and low-risk patients who may benefit from alternative treatment options. Thus, the validation of current results may include not only additional independent replications and functional studies, but also a multi-level integration of several approaches, including, as mentioned above, polygenic risk assessment models (taking into account also other non-genetic factors such as gender, severity of symptoms and etc.).

The inability to replicate the rapidly emerging results of association studies has shifted the focus from single-locus studies to the study of the complex effects of multiple genes (including gene-gene and gene-treatment interactions) on drug response phenotypes[66-69]. Epistasis or gene-gene interactions occur when the effects of alleles on the phenotype depend on the genetic background[70]. Epistasis is an important component of genetic variation and may be causally related in certain complex diseases or traits[70]. For example, one of the recent studies identified epistatic networks that could be applied to risk estimation for *On-Statin* major adverse cardiovascular events[67].

However, measurement of epistasis requires an adequate mathematical or statistical model capable of analyzing multiple patterns of genomic interactions[71]. Furthermore, the calculation of epistatic interactions in general is a very complex computational problem due to the lack of simple, fast and accessible methods for their evaluation[70].

Quantification of epistasis might be of particular interest in the context of anthracyc line-induced cardiomyopathy, where markers of both risk and protective effect have been reported (as described in **Chapters 5** and **6**) and may represent one of the future directions of the research.

At the same time, the individual contribution of each associated genetic variant should not be disregarded, as it can be valuable in identifying a gene or pathway of biological significance.

In order to translate genetic findings into clinical practice and develop new individual preventive approaches, well-designed prospective studies in larger, well-defined populations are needed to accurately confirm the predictive value of reported genetic biomarkers. It should also be noted that implementation of the treatment approaches using genetic information (combined with data on non-genetic factors of interpatient variability in drug response), as well as an accurate and systematic quantification of drug response phenotypes such as toxicity, resistance, etc., can be a much more difficult task than identification of the pharmacogenomic associations. Furthermore, despite biotechnological advances, most healthcare systems currently lack the infrastructure to screen even well-established PGx variants[72], as well as the consensus guidelines that could help health professionals to interpret the results of testing and guide their decision-making process.

9.4. Concluding remarks

The presence of common and rare genetic polymorphisms in patients' genetic constitution can contribute to the substantial variability in terms of the occurrence and severity of treatment-related toxicities in a post HSCT period in pediatric patients, as well as of treatment-related late-adverse effects in survivors of childhood ALL. The application of two different approaches – a candidate genes and an exome-wide association study, using common and rare genetic variants derived through whole-exome sequencing (WES), permitted the discovery of new genetic markers associated with the risk of developing treatment-related complications both during treatment and in the long-term.

Although a few candidate genes studies and GWAS have reported the input of common genetic variants, the knowledge about the contribution of rare variants remains limited. In this study (in the context of the PETALE project), besides identifying the panel of common variants as potential modulators of the risk of developing late adverse effects of the treatment, we also demonstrated the important contribution of rare genetic variants on long-term treatment-related complications in childhood ALL survivors.

In addition, we used WES data to perform whole exome/adjacent UTR analysis of the genetic component of SOS and aGVHD in pediatric patients undergoing HSCT. Several reported genetic markers were successfully replicated in independent patient groups, while also demonstrating dependence on non-genetic factors. These loci have not been previously identified as potential risk predictors and may help understand the molecular mechanisms behind these acute potentially life-threatening therapeutic complications.

This work contributes to the general knowledge of the influence of genetic factors on the risk of developing treatment-related complications both during treatment and in the long-term.

9.5. References

1. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. *Nat Protoc.* 2007;2(10):2492-2501.
2. Morris-Rosendahl DJ. Are there anxious genes? *Dialogues Clin Neurosci.* 2002;4(3):251-260.
3. Santiago R, Tran TH. Genomics and precision medicine in pediatric acute lymphoblastic leukemia. *Journal of Translational Genetics and Genomics.* 2021;5(4):380-395.
4. Krier JB, Kalia SS, Green RC. Genomic sequencing in clinical practice: applications, challenges, and opportunities. *Dialogues in clinical neuroscience.* 2016;18(3):299-312.
5. Pui CH. Precision medicine in acute lymphoblastic leukemia. *Front Med.* 2020;14(6):689-700.
6. The Genome Aggregation Database (gnomAD). In.
7. Teachey DT, Pui C-H. Comparative features and outcomes between paediatric T-cell and B-cell acute lymphoblastic leukaemia. *The Lancet Oncology.* 2019;20(3):e142-e154.
8. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biology.* 2017;18(1):77.
9. Kadan-Lottick NS, Zeltzer LK, Liu Q, et al. Neurocognitive functioning in adult survivors of childhood non-central nervous system cancers. *Journal of the National Cancer Institute.* 2010;102(12):881-893.
10. Robaey P, Krajinovic M, Marcoux S, Moghrabi A. Pharmacogenetics of the neurodevelopmental impact of anticancer chemotherapy. *Developmental disabilities research reviews.* 2008;14(3):211-220.
11. Barrera M, Atenafu E, Pinto J. Behavioral, social, and educational outcomes after pediatric stem cell transplantation and related factors. *Cancer.* 2009;115(4):880-889.
12. Janson C, Leisenring W, Cox C, et al. Predictors of marriage and divorce in adult survivors of childhood cancers: a report from the Childhood Cancer Survivor Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2009;18(10):2626-2635.
13. Clarke SA, Skinner R, Guest J, et al. Clinical outcomes and health-related quality of life (HRQOL) following haemopoietic stem cell transplantation (HSCT) for paediatric leukaemia. *Child: care, health and development.* 2011;37(4):571-580.
14. Kazak AE, Derosa BW, Schwartz LA, et al. Psychological outcomes and health beliefs in adolescent and young adult survivors of childhood cancer and controls. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2010;28(12):2002-2007.
15. Stuber ML, Meeske KA, Krull KR, et al. Prevalence and predictors of posttraumatic stress disorder in adult survivors of childhood cancer. *Pediatrics.* 2010;125(5):e1124-1134.
16. Schultz KA, Ness KK, Whitton J, et al. Behavioral and social outcomes in adolescent survivors of childhood cancer: a report from the childhood cancer survivor study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2007;25(24):3649-3656.

17. Stuart FA, Segal TY, Keady S. Adverse psychological effects of corticosteroids in children and adolescents. *Archives of disease in childhood.* 2005;90(5):500-506.
18. Hochhauser CJ, Lewis M, Kamen BA, Cole PD. Steroid-induced alterations of mood and behavior in children during treatment for acute lymphoblastic leukemia. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer.* 2005;13(12):967-974.
19. Marcoux S, Robaey P, Krajinovic M, Moghrabi A, Laverdiere C. Predictive factors of internalized and externalized behavioral problems in children treated for acute lymphoblastic leukemia. *Pediatric blood & cancer.* 2012;58(6):971-977.
20. Marcoux S, Robaey P, Gahier A, et al. Role of NOS3 DNA variants in externalizing behavioral problems observed in childhood leukemia survivors. *Journal of pediatric hematology/oncology.* 2013;35(4):e157-162.
21. Gaynon PS, Carrel AL. Glucocorticosteroid therapy in childhood acute lymphoblastic leukemia. *Advances in experimental medicine and biology.* 1999;457:593-605.
22. Mullenix PJ, Kernan WJ, Schunior A, et al. Interactions of steroid, methotrexate, and radiation determine neurotoxicity in an animal model to study therapy for childhood leukemia. *Pediatric research.* 1994;35(2):171-178.
23. Drigan R, Spirito A, Gelber RD. Behavioral effects of corticosteroids in children with acute lymphoblastic leukemia. *Medical and pediatric oncology.* 1992;20(1):13-21.
24. Packan DR, Sapolsky RM. Glucocorticoid endangerment of the hippocampus: tissue, steroid and receptor specificity. *Neuroendocrinology.* 1990;51(6):613-618.
25. Wust S, Van Rossum EF, Federenko IS, Koper JW, Kumsta R, Hellhammer DH. Common polymorphisms in the glucocorticoid receptor gene are associated with adrenocortical responses to psychosocial stress. *The Journal of clinical endocrinology and metabolism.* 2004;89(2):565-573.
26. Gordijn MS, van Litsenburg RR, Gemke RJ, et al. Hypothalamic-pituitary-adrenal axis function in survivors of childhood acute lymphoblastic leukemia and healthy controls. *Psychoneuroendocrinology.* 2012;37(9):1448-1456.
27. Pound CM, Clark C, Ni A, Athale U, Lewis V, Halton JM. Corticosteroids, behavior, and quality of life in children treated for acute lymphoblastic leukemia; a multicentered trial. *Journal of pediatric hematology/oncology.* 2012;34(7):517-523.
28. Silverman LB, Stevenson KE, O'Brien JE, et al. Long-term results of Dana-Farber Cancer Institute ALL Consortium protocols for children with newly diagnosed acute lymphoblastic leukemia (1985-2000). *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK.* 2010;24(2):320-334.
29. Waber DP, Shapiro BL, Carpentieri SC, et al. Excellent therapeutic efficacy and minimal late neurotoxicity in children treated with 18 grays of cranial radiation therapy for high-risk acute lymphoblastic leukemia: a 7-year follow-up study of the Dana-Farber Cancer Institute Consortium Protocol 87-01. *Cancer.* 2001;92(1):15-22.
30. Oancea SC, Brinkman TM, Ness KK, et al. Emotional distress among adult survivors of childhood cancer. *J Cancer Surviv.* 2014;8(2):293-303.
31. Lu Q, Krull KR, Leisenring W, et al. Pain in long-term adult survivors of childhood cancers and their siblings: a report from the Childhood Cancer Survivor Study. *Pain.* 2011;152(11):2616-2624.

32. Hudson MM, Mertens AC, Yasui Y, et al. Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *Jama.* 2003;290(12):1583-1592.
33. Moleski M. Neuropsychological, neuroanatomical, and neurophysiological consequences of CNS chemotherapy for acute lymphoblastic leukemia. *Arch Clin Neuropsychol.* 2000;15(7):603-630.
34. Mulhern RK, Fairclough D, Ochs J. A prospective comparison of neuropsychologic performance of children surviving leukemia who received 18-Gy, 24-Gy, or no cranial irradiation. *J Clin Oncol.* 1991;9(8):1348-1356.
35. Pui CH, Campana D, Pei D, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med.* 2009;360(26):2730-2741.
36. Cheung YT, Krull KR. Neurocognitive outcomes in long-term survivors of childhood acute lymphoblastic leukemia treated on contemporary treatment protocols: A systematic review. *Neurosci Biobehav Rev.* 2015;53:108-120.
37. Rozycka A, Jagodzinski PP, Kozubski W, Lianeri M, Dorszewska J. Homocysteine Level and Mechanisms of Injury in Parkinson's Disease as Related to MTHFR, MTR, and MTHFD1 Genes Polymorphisms and L-Dopa Treatment. *Curr Genomics.* 2013;14(8):534-542.
38. Linnebank M, Malessa S, Moskau S, et al. Acute methotrexate-induced encephalopathy--causal relation to homozygous allelic state for MTR c.2756A>G (D919G)? *J Chemother.* 2007;19(4):455-457.
39. Bellampalli R, Vohra M, Sharma K, et al. Acute lymphoblastic leukemia and genetic variations in BHMT gene: Case-control study and computational characterization. *Cancer Biomark.* 2017.
40. Krull KR, Bhojwani D, Conklin HM, et al. Genetic mediators of neurocognitive outcomes in survivors of childhood acute lymphoblastic leukemia. *J Clin Oncol.* 2013;31(17):2182-2188.
41. Kamdar KY, Krull KR, El-Zein RA, et al. Folate pathway polymorphisms predict deficits in attention and processing speed after childhood leukemia therapy. *Pediatr Blood Cancer.* 2011;57(3):454-460.
42. Liu F, Gong X, Yao X, et al. Variation in the CACNB2 gene is associated with functional connectivity of the Hippocampus in bipolar disorder. *BMC Psychiatry.* 2019;19(1):62.
43. Breitenkamp AFS, Matthes J, Nass RD, et al. Rare Mutations of CACNB2 Found in Autism Spectrum Disease-Affected Families Alter Calcium Channel Function. *PLOS ONE.* 2014;9(4):e95579.
44. Cosgrove D, Mothersill O, Kendall K, et al. Cognitive Characterization of Schizophrenia Risk Variants Involved in Synaptic Transmission: Evidence of CACNA1C's Role in Working Memory. *Neuropsychopharmacology.* 2017;42(13):2612-2622.
45. Wu G, Feder A, Cohen H, et al. Understanding resilience. *Frontiers in Behavioral Neuroscience.* 2013;7(10).
46. Musliner KL, Mortensen PB, McGrath JJ, et al. Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA Psychiatry.* 2019;76(5):516-525.

47. Power RA, Tansey KE, Buttenschøn HN, et al. Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biol Psychiatry.* 2017;81(4):325-335.
48. Schulze TG, Akula N, Breuer R, et al. Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World J Biol Psychiatry.* 2014;15(3):200-208.
49. Chow EJ, Leger KJ, Bhatt NS, et al. Paediatric cardio-oncology: epidemiology, screening, prevention, and treatment. *Cardiovasc Res.* 2019;115(5):922-934.
50. Fulbright JM. Review of cardiotoxicity in pediatric cancer patients: during and after therapy. *Cardiol Res Pract.* 2011;2011:942090.
51. Hudson MM, Ness KK, Gurney JG, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. *Jama.* 2013;309(22):2371-2381.
52. Octavia Y, Tocchetti CG, Gabrielson KL, Janssens S, Crijns HJ, Moens AL. Doxorubicin-induced cardiomyopathy: from molecular mechanisms to therapeutic strategies. *J Mol Cell Cardiol.* 2012;52(6):1213-1225.
53. Olson RD, Mushlin PS. Doxorubicin cardiotoxicity: analysis of prevailing hypotheses. *FASEB J.* 1990;4(13):3076-3086.
54. Chow EJ, Antal Z, Constine LS, et al. New Agents, Emerging Late Effects, and the Development of Precision Survivorship. *J Clin Oncol.* 2018;36(21):2231-2240.
55. Barry E, Alvarez JA, Scully RE, Miller TL, Lipshultz SE. Anthracycline-induced cardiotoxicity: course, pathophysiology, prevention and management. *Expert Opinion on Pharmacotherapy.* 2007;8(8):1039-1058.
56. Lipshultz SE, Alvarez JA, Scully RE. Anthracycline associated cardiotoxicity in survivors of childhood cancer. *Heart.* 2008;94(4):525-533.
57. Scully RE, Lipshultz SE. Anthracycline cardiotoxicity in long-term survivors of childhood cancer. *Cardiovasc Toxicol.* 2007;7(2):122-128.
58. Kim AS, Bergmann SR. Anthracycline-induced cardiomyopathy: The search continues. *Journal of Nuclear Cardiology.* 2017;24(1):265-267.
59. Afessa B, Peters SG. Major complications following hematopoietic stem cell transplantation. *Semin Respir Crit Care Med.* 2006;27(3):297-309.
60. Wallhult E, Quinn B. Early and Acute Complications and the Principles of HSCT Nursing Care. In: Kenyon M, Babic A, eds. *The European Blood and Marrow Transplantation Textbook for Nurses: Under the Auspices of EBMT.* Cham (CH): Springer

 2018:163-195.
61. Ansari M, Curtis PH, Uppugunduri CRS, et al. GSTA1 diplotypes affect busulfan clearance and toxicity in children undergoing allogeneic hematopoietic stem cell transplantation: a multicenter study. *Oncotarget.* 2017;8(53):90852-90867.
62. Czerwinski M, Gibbs JP, Slattery JT. Busulfan conjugation by glutathione S-transferases alpha, mu, and pi. *Drug Metab Dispos.* 1996;24(9):1015-1019.
63. Waespe N, Strebel S, Jurkovic Mlakar S, et al. Genetic Predictors for Sinusoidal Obstruction Syndrome-A Systematic Review. *J Pers Med.* 2021;11(5).
64. Petersdorf EW. Genetics of graft-versus-host disease: the major histocompatibility complex. *Blood Rev.* 2013;27(1):1-12.

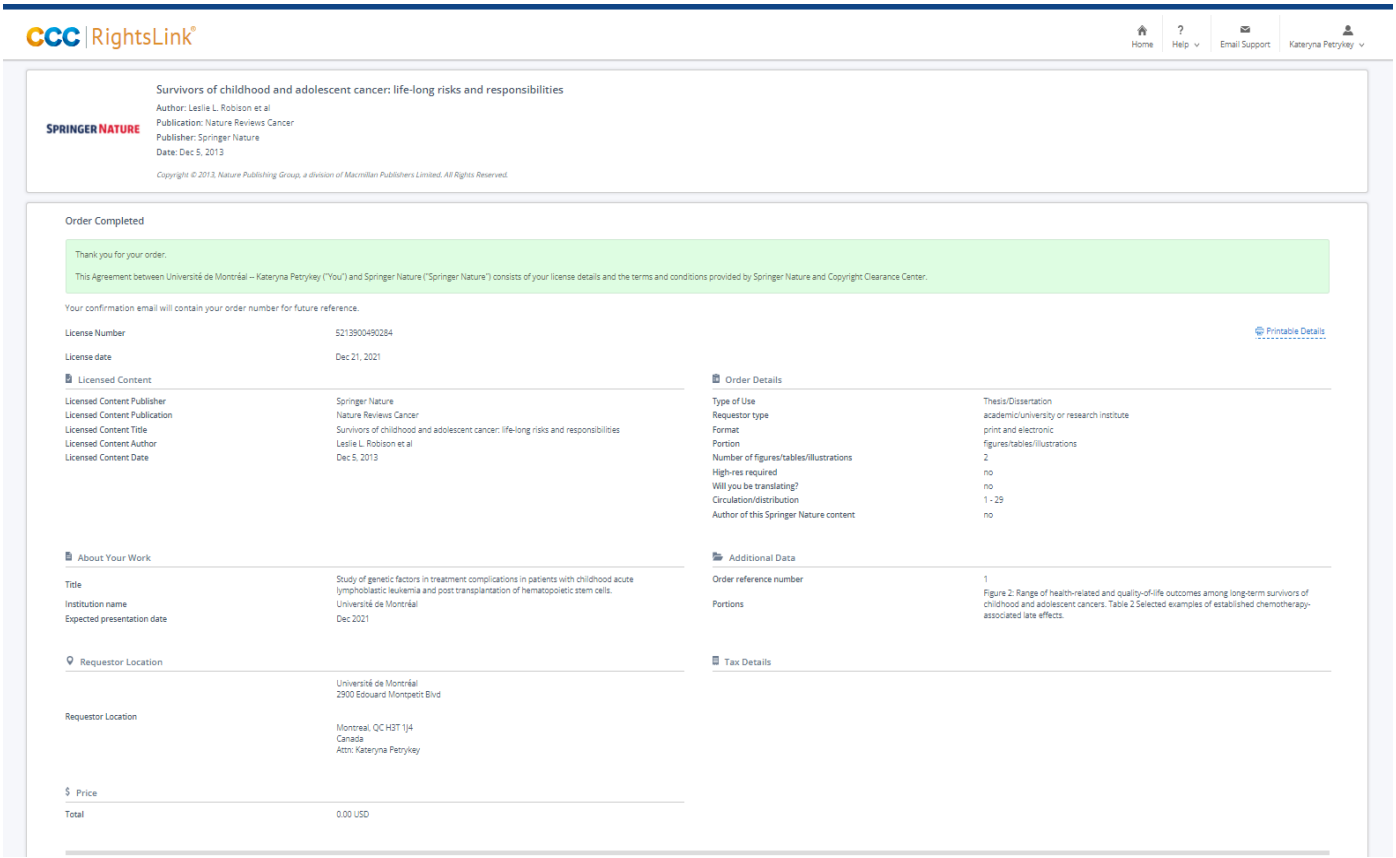
65. Martínez-Laperche C, Buces E, Aguilera-Morillo MC, et al. A novel predictive approach for GVHD after allogeneic SCT based on clinical variables and cytokine gene polymorphisms. *Blood Adv.* 2018;2(14):1719-1737.
66. Demchuk E, Yucesoy B, Johnson VJ, et al. A statistical model for assessing genetic susceptibility as a risk factor in multifactorial diseases: lessons from occupational asthma. *Environ Health Perspect.* 2007;115(2):231-234.
67. Adams SM, Feroze H, Nguyen T, Eum S, Cornelio C, Harralson AF. Genome Wide Epistasis Study of On-Statin Cardiovascular Events with Iterative Feature Reduction and Selection. *Journal of Personalized Medicine.* 2020;10(4):212.
68. Aung N, Vargas JD, Yang C, et al. Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure Development. *Circulation.* 2019;140(16):1318-1330.
69. Luizon MR, Pereira DA, Sandrim VC. Pharmacogenomics of Hypertension and Preeclampsia: Focus on Gene-Gene Interactions. *Front Pharmacol.* 2018;9:168.
70. Carmelo VAO, Kogelman LJA, Madsen MB, Kadarmideen HN. WISH-R– a fast and efficient tool for construction of epistatic networks for complex traits and diseases. *BMC Bioinformatics.* 2018;19(1):277.
71. Le Rouzic A. Estimating directional epistasis. *Front Genet.* 2014;5:198.
72. Auwerx C, Sadler MC, Reymond A, Kutalik Z. From pharmacogenetics to pharmaco-omics: Milestones and future directions. *HGG Adv.* 2022;3(2):100100.

Annex I

Copyright Authorisations

Wong SW, Lenzini S, Shin JW. Perspective: Biophysical regulation of cancerous and normal blood cell lineages in hematopoietic malignancies. APL Bioeng. 2018;2(3):031802. Published 2018 May 22. PMID: 31069313

Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. Nat Rev Cancer. 2014;14(1):61-70. doi:10.1038/nrc3634 PMID: 24304873



CCC | RightsLink®

Home | Help | Email Support | Kateryna Petrykey

SPRINGER NATURE

Survivors of childhood and adolescent cancer: life-long risks and responsibilities

Author: Leslie L. Robison et al
Publication: Nature Reviews Cancer
Publisher: Springer Nature
Date: Dec 5, 2013

Order Completed

Thank you for your order.

This Agreement between Université de Montréal -- Kateryna Petrykey ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5213900490284
License date	Dec 21, 2021

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	Survivors of childhood and adolescent cancer: life-long risks and responsibilities
Licensed Content Author	Leslie L. Robison et al
Licensed Content Date	Dec 5, 2013

Order Details

Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

About Your Work

Title	Study of genetic factors in treatment complications in patients with childhood acute lymphoblastic leukemia and post transplantation of hematopoietic stem cells.
Institution name	Université de Montréal
Expected presentation date	Dec 2021

Additional Data

Order reference number	1
Portions	Figure 2: Range of health-related and quality-of-life outcomes among long-term survivors of childhood and adolescent cancers. Table 2 Selected examples of established chemotherapy-associated late effects.

Requestor Location

Requestor Location	Université de Montréal 2900 Edouard Montpetit Blvd Montreal, QC H3T 1J4 Canada Attn: Kateryna Petrykey

Tax Details

$ Price

Total	0.00 USD

Creative Commons

The request you have made is considered to be non-commercial/educational. As the article you have requested has been distributed under a Creative Commons license (Attribution-Noncommercial), you may reuse this material for non-commercial/educational purposes without obtaining additional permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged (please see the article itself for the license version number). You may reuse this material without obtaining permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged, as per the terms of the license. For license terms, please see http://creativecommons.org/

Annex II

List of other scientific contributions

List of other scientific publications produced during Ph.D. studies not included in the body of the thesis:

1. Shalmiev A, Nadeau G, Aaron M, Ouimet-Grennan E, Drouin S, Bertout L, Beaulieu P, St-Onge P, Rauch F, Veilleux LN, Rezgui MA, **Petrykey K**, Laverdière C, Sinnett D, Alos N, Krajinovic M. Genetic Factors contributing to Late Adverse Musculoskeletal Effects in Childhood Acute Lymphoblastic Leukemia Survivors. *Accepted for publication in* ***The Pharmacogenomics Journal*** *(ISSN 1470-269X, eISSN 1473-1150) on August 10, 2021.*

2. Gagné V, Aubry-Morin A, Plesa M, Abaji R, **Petrykey K**, St-Onge P, Beaulieu P, Laverdière C, Alos N, Leclerc JM, Sallan SE, Neuberg D, Kutok JL, Silverman LB, Sinnett D, Krajinovic M. Genes identified through genome-wide association studies of osteonecrosis in childhood acute lymphoblastic leukemia patients. ***Pharmacogenomics.*** 2019 Nov;20(17):1189-1197.

3. Caru M, **Petrykey K**, Drouin S, Beaulieu P, St-Onge P, Lemay V, Bertout L, Laverdiere C, Andelfinger G, Krajinovic M, Sinnett D, Curnier D. Identification of genetic association between cardiorespiratory fitness and the trainability genes in childhood acute lymphoblastic leukemia survivors. ***BMC Cancer***. 2019 May 14;19(1):443.

4. Nadeau G, Ouimet-Grennan E, Aaron M, Drouin S, Bertout L, Shalmiev A, Beaulieu P, St-Onge P, Veilleux LN, Rauch F, **Petrykey K**, Laverdière C, Sinnett D, Alos N, Krajinovic M. Identification of genetic variants associated with skeletal muscle function deficit in childhood acute lymphoblastic leukemia survivors. ***Pharmgenomics Pers Med.*** 2019 Apr 11; 12:33-45.

5. Aaron M, Nadeau G, Ouimet-Grennan E, Drouin S, Bertout L, Beaulieu P, St-Onge P, Shalmiev A, Veilleux LN, Rauch F, **Petrykey K**, Laverdière C, Sinnett D, Alos N, Krajinovic M. Identification of a single-nucleotide polymorphism within CDH2 gene associated with bone morbidity in childhood acute lymphoblastic leukemia survivors. ***Pharmacogenomics.*** 2019 Apr;20(6):409-420.

Other contributions

NGALL is a MySQL database of genomic variants obtained by sequencing or genotyping. In addition, the database includes several types of annotations from local or external sources to characterize variations. The web interface, built using the *R Shiny* package, has been designed to provide data access to a user without programming skills.

NGALL was developed by the bioinformatics team from Dr. Sinnett's laboratory.

I am especially proud to have had the opportunity to apply my programming skills and contribute very little to the conception of this database. Specifically, I wrote code to convert sequencing data to PLINK format. The fragment of the code is presented in **Figure 1**.

```
data <- read.csv("ngall.csv")
data <- subset(data, select = c("POS","SAMPLE","GT"))
data$GT_res1 <- ifelse(data$GT == "0/0", "1", ifelse(data$GT == "0/1", "1", ifelse(data$GT == "1/0", "2",
                                                                                   ifelse(data$GT == "1/1", "2", "0"))))
data$GT_res2 <- ifelse(data$GT == "0/0", "1", ifelse(data$GT == "0/1", "2", ifelse(data$GT == "1/0", "1",
                                                                                   ifelse(data$GT == "1/1", "2", "0"))))
data <- recast(setDT(data), SAMPLE ~ POS + variable, measure.var=c("GT_res1","GT_res2"))
data1 <- subset(data, select = c(1,1))
data1$zeros1 <-rep(0,nrow(data))
data1$zeros2 <-rep(0,nrow(data))
data1$zeros3 <-rep(0,nrow(data))
data1$zeros4 <-rep(0,nrow(data))
data2 <- subset(data, select = -c(1))
data <- cbind(data1,data2)
write.table(data,sep=",",file="ped_ngall.csv",row.names=FALSE,col.names=TRUE)
```

Figure 1. Fragment of the R code for reading sequencing data.

R programming software was used to code the algorithm which reads the sequencing data from a vcf type data file and then transforms the selected columns into a result file containing the genotyping data in PLINK format.

Annex III

Honors and Awards

Honors and Awards

2021	**The Canadian Society of Pharmacology and Therapeutics** (CSPT), Annual Scientific Meeting 2021, Online edition. *Top Ten Trainee Oral Presentations*
2021	**Scholarship for completing doctoral studies** (**Bourse de fin d'études doctorales**), Faculty of Graduate and Postdoctoral Studies (FESP), Université de Montréal.
2019-2021	**The Cole Foundation Scholarship for Doctoral Fellows**
2019	**Presentation Award** from the **Cardiometabolic Health, Diabetes, And Obesity Research Network (CMDO)** during the 34th Annual research congress at the Sainte-Justine University Health Center
2018	**Maternity leave Scholarship,** Faculty of Graduate and Postdoctoral Studies (FESP), Université de Montréal.
2016	**Scholarship for the accelerated passage from master's to doctorate** (A), Faculty of Graduate and Postdoctoral Studies (FESP), Université de Montréal.
2016	**Recruitment Excellence Award,** Department of Pharmacology and Physiology, Faculty of Medicine, Université de Montréal.