

Université de Montréal

**Caractérisation de la régulation des nouvelles cibles transcriptionnelles du
facteur de transcription *ETV6* dans la leucémie lymphoblastique aiguë**

par

Benjamin Neveu

Département de biochimie

Faculté de médecine

Thèse présentée à la Faculté de médecine
en vue de l'obtention du grade de Doctorat
en biochimie

Octobre 2017

© Benjamin Neveu, 2017

Université de Montréal

Cette thèse intitulée:

**Caractérisation de la régulation des nouvelles cibles transcriptionnelles du
facteur de transcription *ETV6* dans la leucémie lymphoblastique aiguë**

Présentée par: Benjamin Neveu

A été évaluée par un jury composé des personnes suivantes:

Dr. Gerardo Ferbeyre, président-rapporteur

Dr. Daniel Sinnett, directeur de recherche

Dr. Josée Hébert, membre du jury

Dr. Richard Marcotte, examinateur externe

Dr. François Robert, représentant du doyen

Résumé

La leucémie aiguë lymphoblastique de type B (B-LAL) représente environ 22% des cas de cancers pédiatriques. Parmi les lésions génétiques identifiées dans la B-LAL, la translocation $t(12;21)(p13;q22)$ est la plus fréquente. Cette translocation génère une protéine chimérique jumelant les facteurs de transcription *ETV6* et *RUNX1*. Bien que certains phénotypes oncogéniques aient été liés à *ETV6-RUNX1*, sa présence est insuffisante pour enclencher de manière indépendante la leucémogénèse. La délétion du locus 12p13 où est localisé le gène *ETV6* est observée dans 70% des cas de B-LAL $t(12;21)$ positives, suggérant que l'inactivation complète d'*ETV6* est un événement complémentaire nécessaire à la transformation leucémique.

Malgré la forte incidence d'altérations génétiques d'*ETV6* observées dans les désordres hématologiques, peu d'études ont été faites pour élucider sa fonction moléculaire. Pour être en mesure de mieux comprendre l'impact de l'inactivation d'*ETV6* dans la B-LAL, je me suis intéressé à l'identification de ses cibles transcriptionnelles par l'analyse de transcriptomes de lignées cellulaires et de patients leucémiques. Avec la cartographie des sites de liaison d'*ETV6* élaborée à partir d'expériences d'immunoprécipitation de la chromatine, j'ai pu construire son réseau de régulation et mettre en lumière la complexité de sa fonction transcriptionnelle. Des études fonctionnelles ont permis d'identifier une nouvelle fonction pour le gène cible *CLIC5* dans la résistance au stress oxydatif.

Toujours dans l'optique de caractériser *ETV6* au niveau moléculaire, j'ai mis en place une approche de criblage pangénomique par ARN interférant visant l'identification de gènes impliqués dans sa fonction transcriptionnelle. L'impact de certains de ces gènes sur le réseau transcriptionnel d'*ETV6* a permis de confirmer leur rôle de modulateurs. En ce sens, les gènes *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* et *SRP72* sont des modulateurs généraux de la fonction d'*ETV6* alors que d'autres agissent spécifiquement sur certaines cibles. Cette étude est la première à s'être penchée sur les gènes impliqués dans la régulation fonctionnelle d'*ETV6* dans un contexte leucémique.

Les résultats obtenus au cours de cette thèse ont permis d'améliorer considérablement nos connaissances du facteur de transcription *ETV6* dans le cadre de la B-LAL. L'intégration de plusieurs approches a mis en évidence la complexité de sa régulation et permettra de mieux comprendre les mécanismes sous-jacents à son implication dans les désordres hématologiques.

Mots-clés: leucémie aiguë lymphoblastique de l'enfant, facteur de transcription *ETV6*, régulation transcriptionnelle, liaison à l'ADN, modulateurs fonctionnels.

Abstract

B-cells acute lymphoblastic leukemia (B-ALL) accounts for 22% of all childhood cancer cases. Among the genetic aberrations identified in B-ALL, the translocation t(12;21)(p13;q22) is the most frequent one. This translocation generates a chimeric protein between the transcription factors *ETV6* and *RUNX1*. Although some oncogenic phenotypes have been associated to ETV6-RUNX1, its presence alone is not sufficient to induce leukemogenesis. The deletion of the 12p13 locus bearing the *ETV6* gene is observed in approximately 70% of t(12;21) positive B-ALL cases, suggesting that the complete inactivation of *ETV6* is a complementary event required for leukemic transformation.

Despite the high occurrence of *ETV6* abnormalities observed in hematological disorders, few studies have focused on its molecular function. To better understand the impact of *ETV6* inactivation in B-ALL, I used expression data from leukemic cell lines and patients to identify its transcriptional targets. Together with the mapping of *ETV6* binding sites obtained from chromatin immunoprecipitation experiments, I constructed its regulatory network and highlight the complexity of its transcriptional function. Functional studies have unveiled a new function for the *CLIC5* target gene in protecting cells against oxidative stress.

To further characterize *ETV6* at the molecular level, I designed a genome wide shRNA screen and identified genes involved in its transcriptional function. The impact of some of these genes on *ETV6* transcriptional network confirmed their

role as modulators. In this regard, *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* and *SRP72* are broad modulators of *ETV6* function whereas some others act specifically on certain targets. This study is the first to have focused on genes involved in the functional regulation of *ETV6* in leukemia.

The results obtained throughout this thesis have greatly improved our knowledge of the *ETV6* transcription factor in B-ALL. The integration of various approaches has highlighted the complexity of its regulation and will allow a better understanding of the mechanisms underlying its role in hematological disorders.

Key words: childhood acute lymphoblastic leukemia, *ETV6* transcription factor, transcriptional regulation, DNA binding, functional modulators.

Table des matières

Résumé	I
Abstract	III
Table des matières	V
Liste des figures	XI
Liste des tableaux	XIV
Liste des abréviations	XVI
Remerciements	XXI
Chapitre 1: Introduction	1
1.1 Les cancers au Canada.....	1
1.1.1. Généralités	1
1.1.2. Les cancers pédiatriques.....	2
1.2 La leucémie de l'enfant	4
1.2.1. Hématopoïèse et biologie de la leucémie	4
1.2.2. Facteurs de risque	10
1.2.3. Présentation clinique	13
1.2.4. Traitements basés sur le risque de rechute.....	14
1.3 Génétique de la leucémie aiguë lymphoblastique de l'enfant	16
1.3.1. Modèle en deux étapes	16
1.3.2. L'aneuploïdie	19
1.3.3. Les translocations chromosomiques	20
1.3.3.1. Surexpressions	21
1.3.3.2. Les protéines kinases; chromosome de Philadelphie	22
1.3.3.3. Les facteurs de transcription	24
1.3.3.3.1. TCF3-PBX1	25
1.3.3.3.2. Réarrangements <i>MLL</i>	25

1.3.3.3.3. ETV6-RUNX1	27
1.4 <i>ETV6</i>	31
1.4.1. <i>ETV6</i> et le modèle en deux étapes	31
1.4.2. La famille ETS	33
1.4.3. Structure d' <i>ETV6</i>	35
1.4.4. Fonction d' <i>ETV6</i>	37
1.4.4.1. <i>ETV6</i> dans l'hématopoïèse	37
1.4.4.2. <i>ETV6</i> comme suppresseur de tumeur.....	37
1.4.5. Altérations d' <i>ETV6</i> dans les désordres hématologiques	38
1.4.5.1. Réarrangements chromosomiques d' <i>ETV6</i>	38
1.4.5.2. Mutations germinales d' <i>ETV6</i>	40
1.4.6. La régulation médiée par <i>ETV6</i>	41
1.4.6.1. Mécanismes de répression transcriptionnelle	41
1.4.6.2. Les cibles transcriptionnelles d' <i>ETV6</i>	43
1.4.7. La régulation d' <i>ETV6</i>	45
1.4.7.1. Régulation transcriptionnelle d' <i>ETV6</i>	45
1.4.7.2. Régulation post-traductionnelle d' <i>ETV6</i>	46
1.5 Rationnelle et hypothèses de recherche.....	48
1.6 But principal.....	49
Chapitre 2: Premier article - <i>CLIC5</i> dans la B-LAL	50
2.1 Préface	50
2.2 Author contributions.....	52
2.3 Abstract	53
2.4 Introduction.....	54
2.5 Methods.....	56
2.5.1. Expression profiling by RNA-sequencing	56
2.5.2. Quantitative real-time PCR.....	57
2.5.3. Chromatin immunoprecipitation.....	57
2.5.4. Apoptosis assays	57
2.5.5. Immunofluorescence microscopy	58
2.5.6. Statistical tests	58

2.5.7. Ethics statement.....	58
2.6 Results.....	59
2.6.1. <i>ETV6</i> represses the expression of 6 genes in pre-B ALL cell lines and patient samples	59
2.6.2. <i>CLIC5</i> , <i>BIRC7</i> , <i>ANGPTL2</i> and <i>WBP1L</i> are direct targets of <i>ETV6</i>	60
2.6.3. <i>CLIC5A</i> reduces hydrogen peroxide-induced apoptosis	61
2.6.4. <i>CLIC5A</i> is an endosomal ionic channel involved in lysosome-mediated apoptosis.....	62
2.7 Discussion	66
2.8 Declarations.....	71
2.8.1. Availability of data and material.....	71
2.8.2. Competing interests.....	71
2.8.3. Funding	71
2.8.4. Acknowledgments	72
2.9 Figures.....	73
2.10 Tables.....	80
2.11 References	81
2.12 Supplementary material.....	86
2.12.1. Supplemental methods.....	86
2.12.2. Supplemental figures.....	96
2.12.3. Supplemental tables	105
2.13 Reprint permission.....	113
Chapitre 3: Deuxième article - <i>La liaison d'ETV6 à l'ADN</i>	114
3.1 Préface	114
3.2 Author contributions.....	116
3.3 Abstract	117
3.4 Introduction.....	118
3.5 Methods.....	120
3.5.1. Constructs	120
3.5.2. Cell culture	120
3.5.3. Lentiviral production	120

3.5.4. Lentiviral infection.....	121
3.5.5. Western blotting	121
3.5.6. Chromatin immunoprecipitation.....	121
3.5.7. ChIP-seq data analysis.....	122
3.6 Results.....	124
3.6.1. Identification of genome wide <i>ETV6</i> -bound regions	124
3.6.2. <i>ETV6</i> distribution across hematopoietic cell lines genomes	125
3.6.3. Genomic features of <i>ETV6</i> -bound regions	126
3.6.4. Motifs enrichment of <i>ETV6</i> -bound regions	127
3.7 Discussion	130
3.8 Conclusions	133
3.9 Declarations.....	134
3.9.1. Availability of data and material.....	134
3.9.2. Competing interests.....	134
3.9.3. Funding	134
3.10 Figures.....	135
3.11 References	141
3.12 Supplementary material.....	145
3.12.1. Supplementary figures.....	145
3.12.2. Supplementary tables.....	150
Chapitre 4: Troisième article - Les modulateurs fonctionnels d'<i>ETV6</i>	165
4.1 Préface	165
4.2 Author contributions.....	168
4.3 Summary	169
4.4 Background	170
4.5 Results.....	173
4.5.1. Generating <i>ETV6</i> -dependent Blasticidin sensitive cells.....	173
4.5.2. Identification of putative <i>ETV6</i> modulators through a genome wide shRNA screen	175
4.5.3. Validation of putative <i>ETV6</i> modulators on known endogenous transcriptional targets	178

4.5.4. <i>ETV6</i> -modulated transcription of t(12;21)-associated genes.....	180
4.5.5. Implications of <i>ETV6</i> and its modulators in various cancers.....	182
4.6 Discussion	186
4.7 Conclusions	192
4.8 Methods.....	193
4.8.1. Constructs	193
4.8.2. Lentiviral production and transduction	193
4.8.3. Establishment and control of the screening cell line	194
4.8.4. shRNAs preparation	195
4.8.5. Genome wide shRNA screening	196
4.8.6. Data analysis.....	198
4.8.7. Validation of candidate shRNAs by targeted RNA sequencing	199
4.8.8. Targeted RNA sequencing analysis	200
4.8.9. Protein extraction	201
4.8.10. Western blotting	201
4.8.11. qRT-PCR.....	201
4.8.12. Transcriptome analysis of cancer patients	202
4.9 Declarations.....	204
4.9.1. Ethics approval and consent to participate	204
4.9.2. Availability of data and material.....	204
4.9.3. Competing interests.....	204
4.9.4. Funding	204
4.10 Figures.....	205
4.11 Tables.....	214
4.12 References	215
4.13 Supplementary material.....	221
4.13.1. Supplemental figures.....	221
4.13.2. Supplemental tables	233
Chapitre 5: Discussion	254
5.1 Récapitulatif et mise en contexte	254
5.2 Le choix du modèle cellulaire	256

5.3 Régulation directe et indirecte	258
5.4 L'interaction d' <i>ETV6</i> avec ETV6-RUNX1	261
5.5 Les cibles d' <i>ETV6</i> dans la leucémogénèse	263
5.5.1. <i>ETV6</i> , <i>CLIC5</i> et le stress oxydatif dans l'hypothèse virale	263
5.5.2. Les cibles d' <i>ETV6</i> associées à la signature t(12;21)	267
5.6 La régulation de la fonction d' <i>ETV6</i>	270
5.6.1. Le criblage fonctionnel pangénomique par ARN interférant	270
5.6.2. Validation et complexité de la régulation médiée par d' <i>ETV6</i>	272
5.6.3. Pathologies associées à <i>ETV6</i> et ses modulateurs	276
5.7 Conclusions générales et perspectives	278
Références	281

Liste des figures

Chapitre 1: Introduction

Figure I: Distribution des cancers sur 5 ans (période 2006-2010) selon l'âge.....	4
Figure II: Schéma simplifié de l'hématopoïèse.....	6
Figure III: Maturation de la lignée lymphoïde B.....	8
Figure IV: Fréquence des altérations chromosomiques pour les LAL pédiatriques de type B (B-ALL) et T (T-ALL).....	9
Figure V: Modèle en deux étapes des leucémies pédiatriques.....	18
Figure VI: Fréquence des gains de chromosomes chez 414 patients (<18 ans) atteints de LAL de type B.....	20
Figure VII: Fusions BCR-ABL1.....	23
Figure VIII: Protéine de fusion ETV6-RUNX1.....	28
Figure IX: Voies de signalisations et phénotypes cellulaires associés à l'expression de la protéine de fusion ETV6-RUNX1 (E/R).....	29
Figure X: Classification des facteurs de transcription de la famille ETS.....	34
Figure XI: Structure d' <i>ETV6</i>	36
Figure XII: Points de cassure des translocations impliquant le gène <i>ETV6</i> avec ses divers partenaires de fusion.....	39

Chapitre 2: Article 1

Figure 1: Schematic representation of the transcriptome-based design to detect putative direct <i>ETV6</i> target genes.....	73
Figure 2: Expression of putative <i>ETV6</i> targets in pediatric pre-B ALL patients...	74
Figure 3: Quantitative real-time PCR and chromatin immunoprecipitation validation of putative <i>ETV6</i> target genes.	75
Figure 4: <i>CLIC5A</i> protects cells from hydrogen peroxide-induced apoptosis.....	76
Figure 5: <i>CLIC5A</i> is implicated in lysosome-mediated cell death.	77
Figure 6: Co-localization of <i>CLIC5A</i> and transferrin receptors.....	78
Figure 7: Proposed mechanism of <i>CLIC5A</i> involvement in <i>ETV6</i> -associated childhood pre-B ALL.....	79
Figure S1: Protein expression levels of ETV6-His and ETV6 Δ ETS_NLS-His in Reh cells.....	96
Figure S2: Expression of ETV6 and ETV6-HA proteins in Reh cells.....	97
Figure S3: <i>ETV6</i> specifically binds and represses the <i>CLIC5A</i> isoform.....	98
Figure S4: <i>CLIC5</i> expression level in <i>CLIC5A</i> overexpressing Reh cells and leukemic patients samples.....	99
Figure S5: Migration of Reh cells to SDF-1.....	100

Figure S6: Mitochondrial membrane potential (MMP) in hydrogen peroxide-treated cells	101
Figure S7: Localization of <i>CLIC5A</i> and lysosomes	102
Figure S8: Expression profile of <i>CLIC5</i> in normal and matched tumor tissue pairs	103
Figure S9: <i>ETV6</i> constructs	104

Chapitre 3: Article 2

Figure 1: ChIP-seq analysis of <i>ETV6</i> binding sites in pre-B leukemia cells	135
Figure 2: Genomic annotations of <i>ETV6</i> -bound regions	137
Figure 3: Consensus binding motifs enrichment analyses	139
Figure S1: Examples of identified <i>ETV6</i> peaks	145
Figure S2: Genomic annotations of <i>ETV6</i> peaks in GM12878 and K562 cells ..	146
Figure S3: Density of the top 8 over-represented motifs in Reh <i>ETV6</i> peaks ...	147
Figure S4: Fold enrichment of expression-correlated <i>ETV6</i> peaks in Reh cells according to each motif groups	148
Figure S5: Motif enrichment analysis of ETS/RUNX free <i>ETV6</i> peaks in Reh cells	149

Chapitre 4: Article 3

Figure 1: Overview of the genome wide shRNA screening design	205
Figure 2: Establishment of an <i>ETV6</i> -dependent Blastocidin-sensitive leukemia cell system	206
Figure 3: Genome wide shRNA screening to identify <i>ETV6</i> modulators	208
Figure 4: Impact of putative <i>ETV6</i> modulators on its transcriptional regulatory network	209
Figure 5: Impact of the top <i>ETV6</i> modulators on the transcription of t(12;21)-associated overexpressed genes	211
Figure 6: Top <i>ETV6</i> modulators in childhood pre-T ALL	212
Figure 7: <i>ETV6</i> modulators in glioblastomas and gliomas	213
Figure S1 (Additional file 1): Identification of EBS ₃ tk BlastR insertion site using ChIP-seq data	221
Figure S2 (Additional file 2): Experimental design of the genome wide shRNA screen	222
Figure S3 (Additional file 3): shRNAs distribution	223
Figure S4 (Additional file 8): <i>DYRK4</i> knockdown	224
Figure S5 (Additional file 9): <i>ETV6</i> modulators on t(12;21)-associated genes ..	225
Figure S6 (Additional file 10): Top <i>ETV6</i> modulators on t(12;21)-associated genes	226

Figure S7 (Additional file 11): Knockdown of the top <i>ETV6</i> modulators in a pre-T ALL cell line	227
Figure S8 (Additional file 12): <i>ETV6</i> modulators in glioblastomas and gliomas	228
Figure S9 (Additional file 13): <i>ETV6</i> modulators in breast cancer.....	229
Figure S10 (Additional file 15): Impact of <i>SRP72</i> on the <i>ETV6</i> protein.....	230
Figure S11 (Additional file 17): Assessment of the multiplicity of infection (MOI) for each shRNA pool.....	231
Figure S12 (Additional file 19): Technical reproducibility of the quantification of shRNAs from the genome wide screen.....	232

Chapitre 5: Discussion

Figure XIII: Hématopoïèse induite par des effets directs et indirects de l'infection	265
--	-----

Liste des tableaux

Chapitre 1: Introduction

Tableau I: Les principaux facteurs de risque environnementaux de la leucémie et leur contribution	11
Tableau II: Exemples de polymorphismes associés à la leucémie	12
Tableau III: Symptômes des leucémies pédiatriques.....	13
Tableau IV: Facteurs influençant le pronostic des enfants atteints de LAL.....	15
Tableau V: Altérations génétiques secondaires communes (fréquence >5%) de type CNA (<i>Copy number alterations</i>) répertoriées dans les cas de B-LAL ETV6-RUNX1 positive.....	32

Chapitre 2: Article 1

Table 1: Expression status of the 6 putative direct <i>ETV6</i> target genes.....	80
Table S1: Oligonucleotide sequences used for quantitative real-time PCR validation.....	105
Table S2: Oligonucleotide sequences used for chromatin immunoprecipitation qPCR	106
Table S3: Downregulated transcripts in ETV6-His expressing Reh cells vs pLENTI cells	107
Table S4: Downregulated transcripts in ETV6-His expressing Reh cells vs ETV6 Δ ETS_NLS-His cells	110
Table S5: Oligonucleotide sequences used for cloning	111
Table S6: Antibodies used in western blotting	112

Chapitre 3: Article 2

Table S1: Oligonucleotide sequences used for chromatin immunoprecipitation qPCR	150
Table S2: Summary statistics of Reh-derived <i>ETV6</i> peaks using different peak fold enrichment thresholds.....	151
Table S3: Characteristics of the 1,931 prioritized Reh-derived <i>ETV6</i> peaks.....	153
Table S4: Details of 74 Reh-derived <i>ETV6</i> peaks associated with a differentially expressed gene	154
Table S5: Reh and GM12878 overlapping peaks and correlated gene expression	157
Table S6: Motifs enrichment of <i>ETV6</i> -bound regions in Reh, GM12878 and K562 cell lines	158
Table S7: Peak fold enrichment per motifs and expression	163

Table S8: Motifs enrichment of ETS/RUNX motifs free <i>ETV6</i> -bound regions in Reh cells	164
---	-----

Chapitre 4: Article 3

Table 1: Data summary of the 5 top <i>ETV6</i> modulators	214
Table S1 (Additional file 4): List of the 1,241 selected shRNAs	233
Table S2 (Additional file 5): shRNAs selected for validation	234
Table S3 (Additional file 6): Struct2Net prediction analysis.....	238
Table S4 (Additional file 7): <i>ETV6</i> targets and control genes used in the validation process	240
Table S5 (Additional file 14): Proteomic analysis of <i>ETV6</i> -bound proteins	249
Table S6 (Additional file 16): Oligonucleotide sequences	250
Table S7 (Additional file 18): Raw and normalized counts of SHC002 and SHC202 added plasmids	251
Table S8 (Additional file 20): Probes and target sequences of genes monitored by Targeted RNA sequencing (Illumina "Manifest")	252
Table S9 (Additional file 21): Antibodies used for immunodetection	253

Chapitre 5: Discussion

Tableau VI: Sommaire des motifs consensus identifiés au niveau des séquences liées par <i>ETV6</i> ayant été associées à des gènes réprimés dans la lignée Reh et à la signature transcriptionnelle des patients t(12;21)(p13;q22) positifs	260
---	-----

Liste des abréviations

AD: Activation Domain

ALL / LAL: Acute lymphoblastic leukemia / Leucémie aiguë lymphoblastique

AML: Acute myeloid leukemia

B-ALL / B-LAL: B-cells ALL / LAL de type B

BCP: B-cell progenitor

BlastR: Blastocidin resistance gene

BlastS: Blastocidin

c.-à-d.: c'est-à-dire

cDNA: Complementary DNA

CÉM: Champs électromagnétiques

ChIP: Chromatin immunoprecipitation / Immunoprécipitation de la chromatine

ChIP-MS: Chromatin-interacting protein mass spectrometry

ChIP-seq: Chromatin immunoprecipitation sequencing

CID: C-terminal inhibitory domain

CLIC: Chloride intracellular channel

CLP: Common lymphoid progenitor

CMP: Common myeloid progenitor

CNA: Copy number alteration

CNS / SNC: Central nervous system / Système nerveux central

CPT: Camptothecin

CTD: C-terminal domain / Domaine C-terminal

CTL: Cytotoxic T lymphocytes / Lymphocytes T cytotoxiques

DEG: Differentially expressed genes

DFO: Deferoxamine mesylate salt

DOXO: Doxorubicin

E/R: ETV6-RUNX1

EBS: ETS binding site

ECL: Enhanced chemiluminescence

EMSA: Electrophoretic mobility shift assay

EP: Erythroid progenitor

ETP: Early T-cell precursor

ETS: E26 transformation-specific

FAC: Ferric ammonium citrate

FACS: Fluorescence activated cell sorting

FBS: Fetal bovine serum

gDNA: Genomic DNA

GMP: Granulocyte–macrophage progenitor

GWAS: Genome wide association study / Étude d'association à l'échelle du génome

HDACs: Histone deacetylases / Histones déacétylases

HLH: Helix-loop-helix

HRP: Horseradish peroxidase

HSC: Hematopoietic stem cell

HSPCs: Hematopoietic stem and progenitor cells / Progéniteurs et cellules souches hématopoïétiques

IAA-94: Indanyloxyacetic acid 94

iAmp21: Intrachromosomal amplification of chromosome 21

IFNs: Interférons

ISRE: IFN-stimulated response element

KD: Knockdown

LID: Linker inhibitory damper

LMP: Lysosomal membrane permeabilization

LOH: Loss of heterozygosity

MEP: Megakaryocyte erythroid progenitor

MkP: Megakaryocyte progenitor

MMP: Mitochondrial membrane potential

MOI.: Multiplicity of infection

MOMP: Mitochondrial outer membrane permeabilization

MRM: Maladie résiduelle minimale

MSCs: Mesenchymal stromal cells / Cellules stromales mésenchymateuses

N: Sample size / Échantillonnage

NAM: Nicotinamide

NGS: Normal goat serum

NK: Natural killer

NLS: Nuclear localization signal

OR: Odds ratio

PAMPs: Pathogen-associated molecule patterns / Motifs moléculaires associés aux pathogènes

PBS: Phosphate-buffered saline

PD: Pointed domain

p. ex.: par exemple

Ph-like: Philadelphia-like

PI: Propidium iodide

PNC: Peripheral nervous cells

PNT: Pointed

pre-B ALL: Precursor B-cell ALL

PRRs: Pattern recognition receptors / Récepteurs de reconnaissance de motifs moléculaires

PRXD: Hydrogen peroxide

qPCR: Quantitative PCR

qRT-PCR: quantitative real-time PCR

RD: Repression Domain

RIME: Rapid immunoprecipitation mass spectrometry of endogenous protein

RNA-seq: Ribonucleic acid sequencing

ROS: Reactive oxygen species / Espèces d'oxygènes réactives

RPKM: Reads Per Kilobase per Million mapped reads

rRNA: Ribosomal RNA

SAM: Sterile alpha motif

SD: Standard deviation

shRNA: Small hairpin RNA

SNP: Single nucleotide polymorphism

T-ALL / T-LAL: T-cells ALL / LAL de type T

TCGA :The cancer genome atlas

TNK: T-cell natural killer cell progenitor

TSA: Trichostatin A

TSS: Transcription start site

WT: Wild type

Remerciements

Je voudrais tout d'abord remercier Daniel de m'avoir initié au monde de la recherche et donné par la suite les moyens et la liberté de mener à terme mes projets ambitieux. Merci également de m'avoir mis en contact avec des gens exceptionnels sans qui ces années d'études auraient été encore plus difficiles.

Durant les années passées, j'ai vu plusieurs personnes transiter par le laboratoire et remercie chacune d'elles d'avoir contribué, à leurs façons, à la personne que je suis aujourd'hui.

Je remercie Camille, Josiane et Sylvie d'avoir eu la lourde tâche de former un stagiaire et surtout de m'avoir donné le goût de continuer. Pauline et Karine, pour leur aide inestimable quand venait le temps de mener de front des quantités d'expérience trop grandes pour un seul homme. Merci de ne jamais m'avoir fait regretter d'avoir eu confiance en vous. Manon, pour son efficacité légendaire; la rapidité de ses commandes m'a fait gagner un temps précieux. Un mot particulier pour Chantal, sans qui mon passage dans le monde de la recherche aurait été très éphémère. Non seulement a-t-elle été la personne derrière toutes mes réussites, mais elle a su me partager son enthousiasme toujours débordant à chacune de nos innombrables discussions. Merci également à l'équipe de bio-informatique d'avoir pris part à plusieurs aspects de mes travaux; une mention spéciale à Maxime avec qui j'ai eu énormément de plaisir à travailler. Je remercie tous ceux qui ont eu à affronter mes premières versions de manuscrits.

Merci à Stéphane pour son implication majeure dans mon projet le plus important, de sa mise en place jusqu'à l'écriture. Je le remercie également pour son criticisme, sa détermination et sa patience mais surtout, surtout, pour son honnêteté.

Un énorme merci à Jean-François qui a été un mentor pendant mes années au Doctorat. Mes recherches ont grandement profité de son criticisme et de sa franchise. Mais surtout, cette même honnêteté sans finesse, en science comme ailleurs, aura fait naître une grande amitié. Merci aussi à Pamela d'avoir pris part à nos délires de fin de soirée.

Merci à mes amis de longue date, Antoine, Éric, Simon et Yan et à leur copine, Véronique, Marie-Christine, Josée et Charlotte, d'avoir apporté une dimension humaine à ma vie. À Kate, qui a été une témoin de la vie palpitante d'un étudiant en rédaction de thèse. Tu as rendu ces journées tellement plus agréables.

Et finalement, merci à mes parents, Anne et Sylvain, à mon second père, Clermont, et à mes soeurs Gabrielle, Julia et Delphine. Votre soutien inconditionnel et vos conseils me donnent la force de progresser dans tout ce que j'entreprends.

*À Chantal,
cette thèse est en grande partie la tienne*

Chapitre 1: Introduction

1.1 Les cancers au Canada

1.1.1. Généralités

Près du quart des Canadiens mourront d'un cancer, ce qui en fait la première cause de mortalité au Canada en 2011 devant les maladies cardiovasculaires (*Canadian Cancer Statistics, 2015*). Malgré d'importants progrès au niveau du diagnostic, du traitement et de la prise en charge ainsi qu'une diminution d'incidence de certains types de cancers (p. ex. le cancer du poumon attribué au tabac), le vieillissement de la population contribuera à hausser le nombre de nouveaux cas dans les prochaines années. Bien que la mortalité associée aux cancers soit à la baisse de façon générale, les cancers continueront d'être la première cause de mortalité. Les conséquences des cancers eux-mêmes ainsi que les effets secondaires associés aux traitements sont lourds sur la santé humaine. L'augmentation des cas de cancers combinée à une meilleure survie auront comme impact d'engendrer un stress sur le système de santé qui devra composer avec un bassin plus important de survivants.

Pour diminuer les impacts sociaux et financiers des cancers, la prévention, l'efficacité diagnostique et l'amélioration des traitements sont les points clés sur lesquels miser. Selon l'Organisation mondiale de la Santé, la prévention serait la méthode la plus efficace et rentable à long terme pour contrôler les coûts

associés aux cancers (<http://www.who.int/cancer/prevention/en/>). La détection précoce de certains types de cancers a également eu des effets bénéfiques notables dans les dernières années. La diminution marquée de mortalité associée au cancer du sein illustre bien l'importance d'un dépistage précoce et efficace à grande échelle dans la tranche d'âge à risque [1]. Finalement, l'amélioration et la personnalisation des traitements permettent non seulement un meilleur taux de rémission, mais aussi la réduction des effets secondaires néfastes. L'identification et l'élaboration des mécanismes moléculaires impliqués dans le développement de cancers permettent la mise en place de thérapies ciblées et personnalisées dont bénéficieront les patients.

1.1.2. Les cancers pédiatriques

En général, les enfants sont très peu atteints par les cancers. Les jeunes dans la tranche d'âge 0-19 ans représentent un peu moins de 1% de l'ensemble des cas de cancers diagnostiqués. Les tumeurs pédiatriques constituent toutefois la principale cause de décès par maladie dans cette population. Bien que les nouvelles thérapies ont globalement permis d'améliorer le taux de survie des patients atteints de cancers, il n'y a eu que peu d'améliorations observées dans cette tranche d'âge depuis les dernières années. La complexité génétique sous-jacente à ces cancers pédiatriques combinée aux lourdes répercussions à long terme des traitements chez les enfants rendent plus difficile cette progression.

Il existe plusieurs différences de présentation clinique et moléculaire entre les cancers adultes et pédiatriques. La distribution des sites tumoraux primaires est dépendante de l'âge. Par exemple, les leucémies représentent la plus grande

proportion des nouveaux diagnostics de cancer chez les 0-14 ans (32%) tandis qu'elles représentent environ 3% des nouveaux cas dans la population générale. Inversement, le cancer du poumon représente 17% des cancers chez les 70 ans et plus alors qu'il est pratiquement inexistant dans la population pédiatrique (**Figure I**). Ces particularités sont caractéristiques des étiologies diverses de chacun des cancers. Dans le cas du cancer du poumon, le tabagisme est le facteur de risque par excellence et est directement corrélé avec son incidence dans la population. L'exposition prolongée à de tels carcinogènes a assurément moins d'impact dans l'étiologie des cancers pédiatriques ayant même dans certains cas une origine prénatale (*in utero*). De plus, au sein d'un même tissu d'origine, les atteintes moléculaires sont également variables. Dans les rares cas de leucémies adultes, la translocation t(9;22)(q34;q11) (ou *BCR-ABL1*) est observée dans environ le quart des cas, mais jusqu'à 95% dans les cas de leucémie myéloïde chronique [2]. En pédiatrie, les précurseurs lymphoïdes sont majoritairement touchés [3] et cette même translocation est observée dans seulement 3% des cas [4]. La sélection des atteintes génétiques est donc fortement dépendante du contexte cellulaire car leur contribution aux cancers est spécifique et variable.

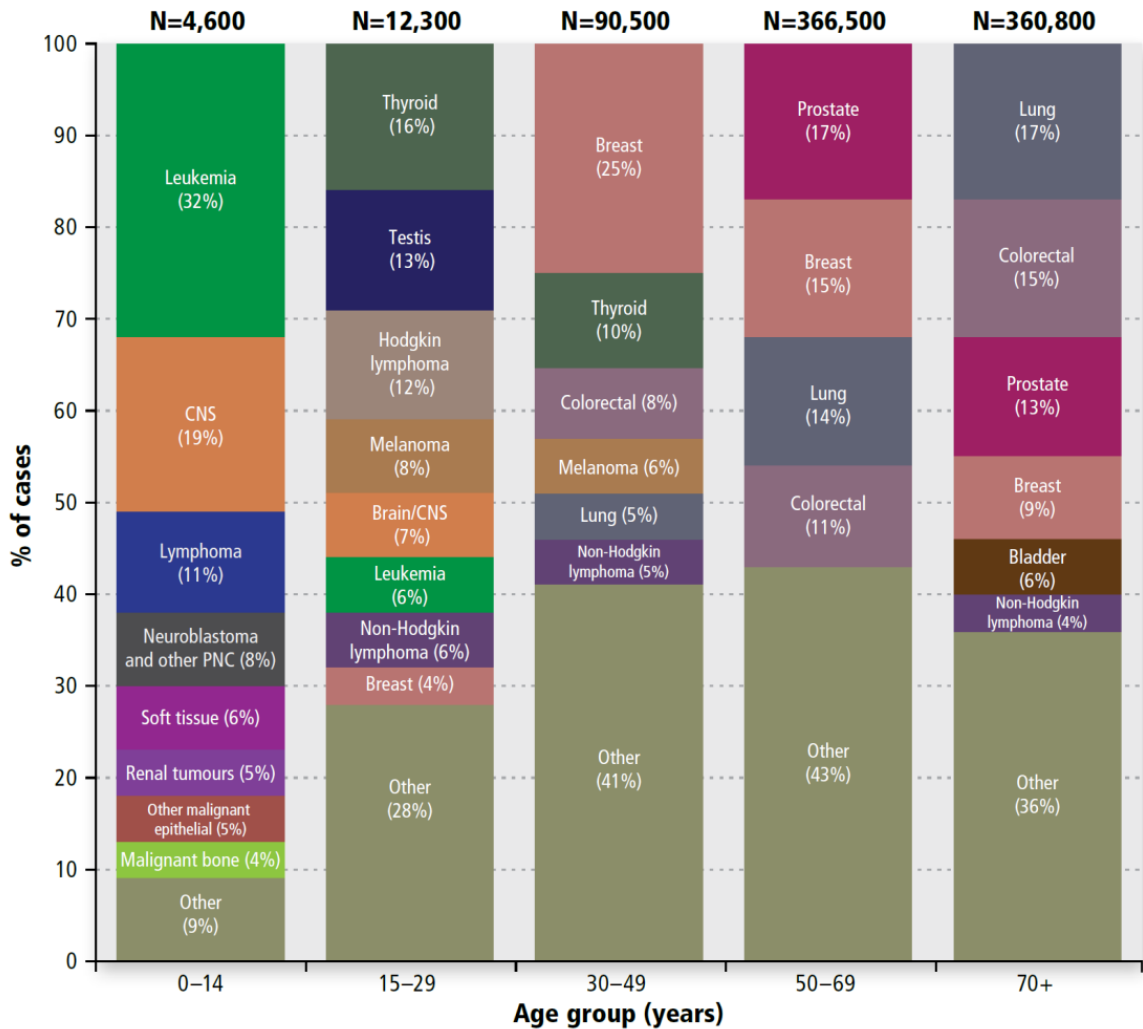


Figure I: Distribution des cancers sur 5 ans (période 2006-2010) selon l'âge. CNS, *central nervous system*; PNC, *peripheral nervous cells*. Adaptée de *Canadian Cancer Statistics*, 2015.

1.2 La leucémie de l'enfant

1.2.1. Hématopoïèse et biologie de la leucémie

L'hématopoïèse fait référence au processus de production de l'ensemble des cellules sanguines. Pour ce faire, les cellules souches hématopoïétiques (ou

multipotentes) de la moelle osseuse sont en constante prolifération et division cellulaire. Cette division est particulière au sens où une des cellules filles demeure pluripotente alors que l'autre emprunte une voie de maturation (ou différenciation); ce phénomène est appelé division asymétrique et permet de maintenir la population de cellules souches. La différenciation de ces cellules permet de générer la totalité des cellules sanguines matures par un processus complexe de régulation de l'expression de gènes induits par des facteurs de transcription clés modulés par des cytokines spécifiques à une niche cellulaire (**Figure II**) [5-7]. La première grande classification est la dichotomie lignage myéloïde-lymphoïde. Chacun de ces deux lignages comprend divers types cellulaires. La lignée myéloïde englobe les monocytes et les cellules dendritiques, les granulocytes (neutrophiles, éosinophiles, basophiles et mastocytes), les érythrocytes (globules rouges) ainsi que les mégacaryocytes et les plaquettes. La lignée lymphoïde comprend les *natural killer* (cellules NK) ainsi que les lymphocytes B et T. D'un point de vue simpliste, les leucémies sont la conséquence d'un dérèglement du processus d'hématopoïèse.

Les types de leucémies sont associés à la lignée affectée ainsi qu'au niveau de sévérité du développement du cancer. La terminologie "aiguë" et "chronique" fait référence à une accumulation rapide de cellules immatures et non fonctionnelles (appelées aussi blastes) par opposition à une déficience ou surproduction modérée de cellules plus matures, respectivement. Les leucémies chroniques peuvent être asymptomatiques pendant plusieurs années, mais peuvent ensuite évoluer en une phase aiguë appelée crise de blastes [8]. Au moment de cette

transformation, la leucémie présente certaines caractéristiques des leucémies aiguës telles que la déficience de production des cellules matures, mais demeure génétiquement et cliniquement distincte des leucémies aiguës *de novo*.

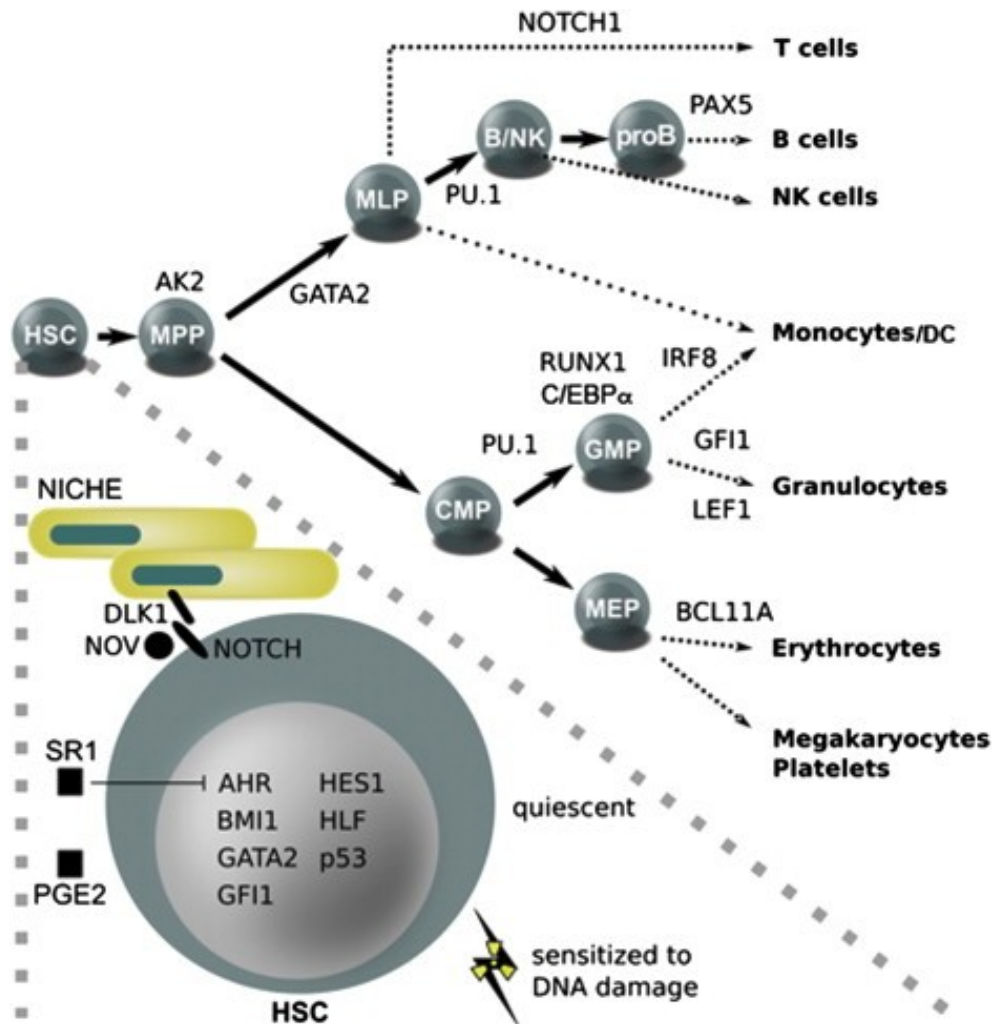


Figure II: Schéma simplifié de l'hématopoïèse. Représentation du processus de différenciation des cellules souches hématopoïétiques par l'effet de facteurs de transcription (non exhaustifs). B/NK, *B-cell and natural killer progenitor*; CMP, *common myeloid progenitor*; DC, *dendritic cells*; HSC, *hematopoietic stem cell*; GMP, *granulocyte and macrophage progenitor*; MEP, *megakaryocyte and erythroid progenitor*; MLP, *multilymphoid progenitor*; MPP, *multipotent progenitor*; proB, *B-cell progenitor*. Reproduite de S. Doulatov *et al.*, 2012 [7].

Les leucémies pédiatriques sont très majoritairement (~75%) de type aiguë lymphoblastique et affectent préférentiellement (85%) la lignée des lymphocytes B (B-LAL) par rapport aux lymphocytes T (15%; T-LAL) [9]. De façon générale, les cellules leucémiques sont bloquées à une étape précoce de leur développement causée par la défaillance des programmes transcriptionnels nécessaires à leur maturation, mais conservent leur capacité de prolifération. Dans le lignage B, les facteurs de transcription IKAROS (*IKZF1*), E2A (*TCF3*), *EBF1*, *FOXO1*, *PAX5* et *IRF4*, pour n'en nommer que quelques-uns, sont à la tête des réseaux de régulation impliqués dans les différentes étapes de différenciation [10, 11]. Ces mêmes facteurs sont d'ailleurs communément la cible de pertes de fonction dans la B-LAL, démontrant bien l'importance que prend le processus de différenciation dans la pathophysiologie de la leucémie [12]. Dans la majorité des cas de B-LAL, les blastes présentent des marqueurs de surface caractéristiques des cellules pré-B ($CD34^+CD10^+CD19^+$), indicatif d'un blocage de maturation à ce stade (**Figure III**) [13]. Ainsi, ces cellules immatures rivaliseront avec les cellules saines et prendront ultimement une place importante dans la niche hématopoïétique pour ultimement engendrer des symptômes chez la personne atteinte.

Comme les voies moléculaires impliquées dans la maturation des lymphocytes B et T sont majoritairement différentes, les altérations génétiques observées dans les cas de LAL B et T sont, pour la plupart, également distinctes (**Figure IV**) [9]. Pour les T-LAL, l'expression aberrante du facteur de transcription *TAL1* et de son cofacteur *LMO2* est retrouvée dans plus de la moitié des cas en combinant les

différentes altérations génétiques à ces locus [14]. La surexpression de ces facteurs engendre des boucles d'autorégulation avec des facteurs clés de l'hématopoïèse notamment *GATA3*, *RUNX1* et *MYB* [15]. Pour les B-LAL, la haute hyperdiploïdie (plus de 50 chromosomes par cellule) et la translocation t(12;21)(p13;q22) impliquant les facteurs de transcription *ETV6* et *RUNX1* (*AML1*) se partagent également la moitié des cas.

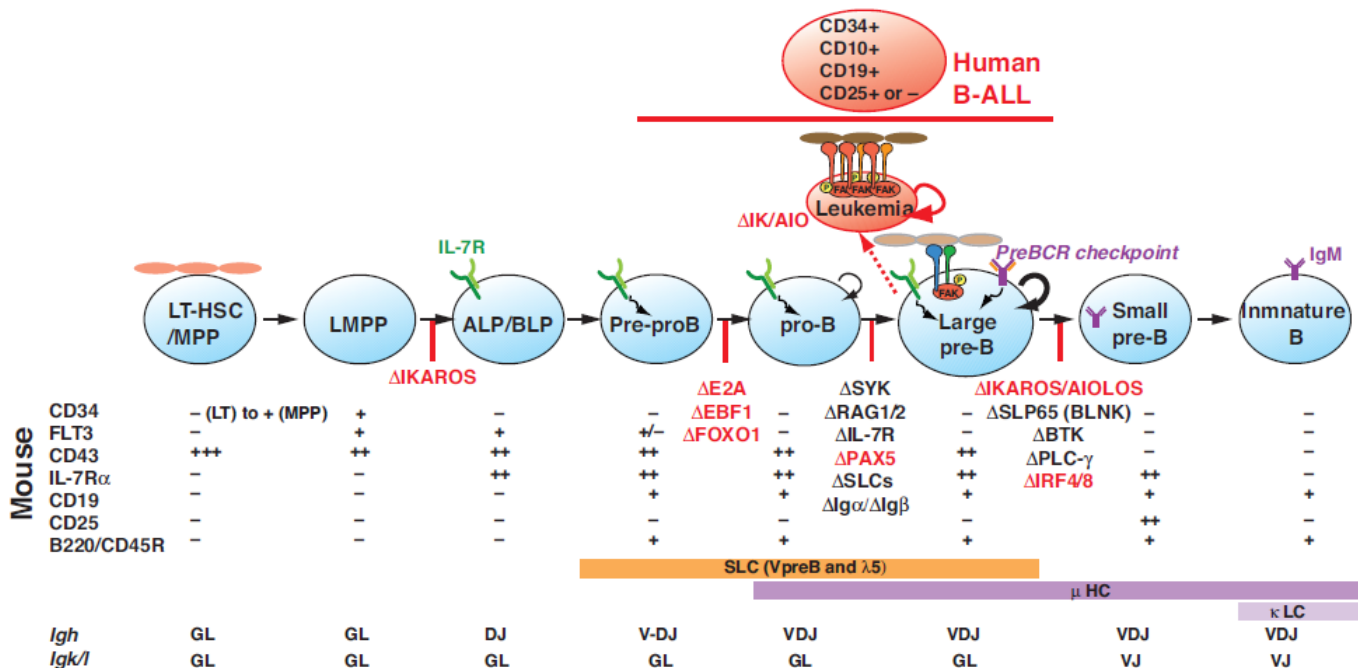


Figure III: Maturation de la lignée lymphoïde B. L'expression des marqueurs de surface (murins et humains) et l'état des réarrangements des gènes d'immunoglobulines (*IgH* et *IgK*) sont indiqués pour chaque stade de différenciation. Les gènes majeurs impliqués dans les étapes de transition sont indiqués entre chaque stade. ALP/BLP, *All/B-cell-biased lymphoid progenitor*; B-ALL, *B-cell acute lymphoblastic leukemia*; LMPP, *lymphoid multipotent progenitor*; LT-HSC, *long-term hematopoietic stem cell*; MPP, *multipotent progenitor*. Reproduite de Y. Hu et al., 2017 [13].

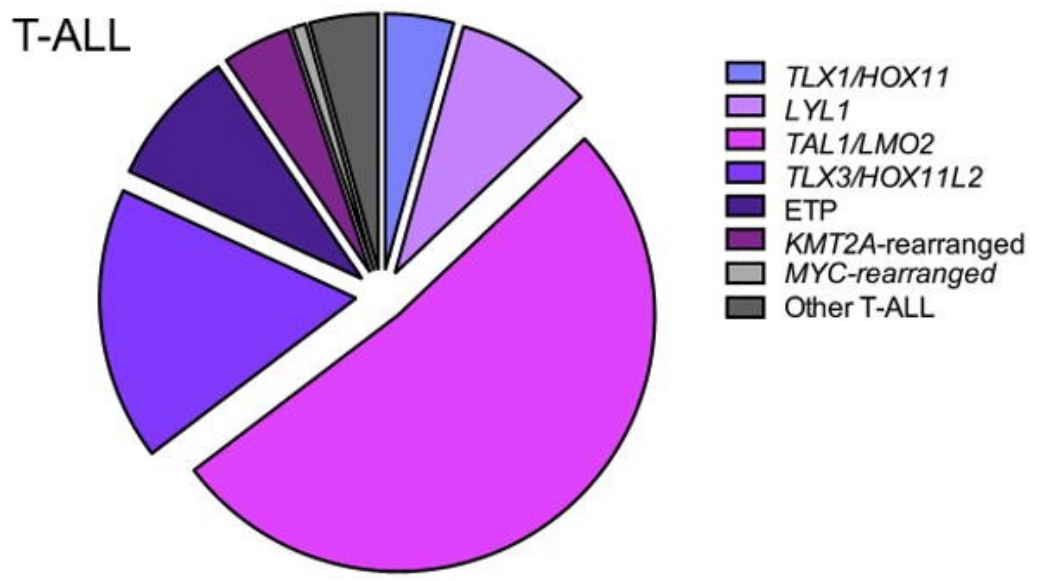
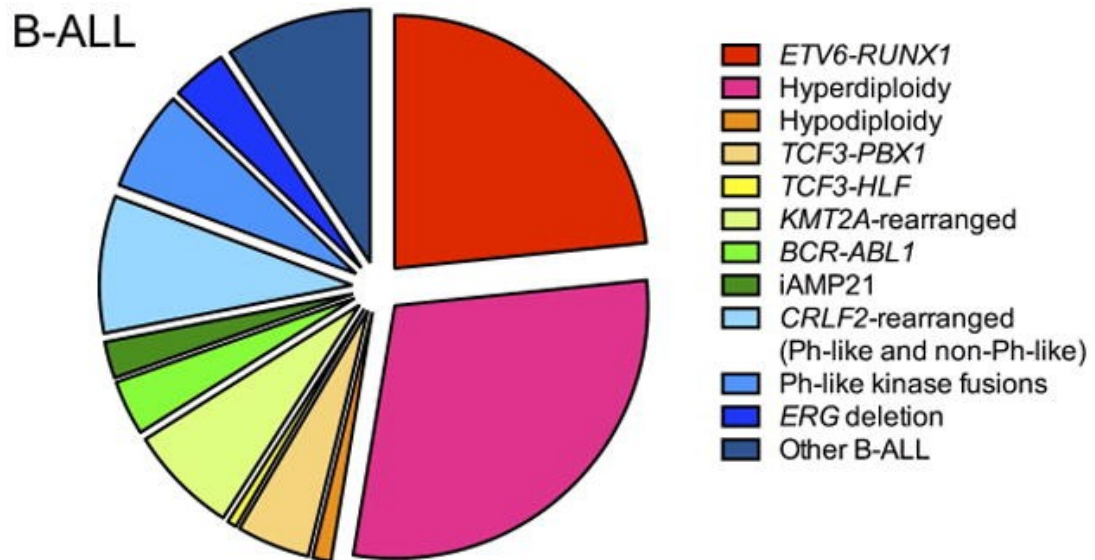


Figure IV: Fréquence des altérations chromosomiques pour les LAL pédiatriques de type B (B-ALL) et T (T-ALL). iAmp21, *Intrachromosomal amplification of chromosome 21*; Ph-like, *Philadelphia-like*; ETP, *Early T-cell precursor*. Reproduite de S.K. Tasian *et al.*, 2015 [9].

1.2.2. Facteurs de risque

En pédiatrie, l'impact de l'environnement sur l'initiation et le développement des cancers est moindre que chez l'adulte de par la courte période d'exposition. Néanmoins, certains groupes de recherche ont démontré certaines corrélations entre le développement de la leucémie et l'exposition à certains types de radiations ionisantes, de champs électromagnétiques, de produits chimiques comme le benzène ou certains pesticides et la prise de tabac ou d'alcool par la mère pendant la grossesse [16]. Le **tableau I** présente un sommaire des principaux facteurs de risque environnementaux et leur potentielle contribution [17]. Il est à noter qu'un consensus n'a toujours pas été atteint quant à l'implication de ces expositions sur le développement de la leucémie.

La cause environnementale la plus fortement associée au développement d'une leucémie serait le profil d'infection virale (voir **Tableau I**). Bien qu'aucune évidence expérimentale ne vienne appuyer ces observations, il a été suggéré qu'une forte réponse immunitaire pourrait réduire la maturation des cellules hématopoïétiques et promouvoir l'apparition ou la sélection d'un clone plus résistant et adapté à ce nouvel environnement pro-inflammatoire [18]. Selon cette hypothèse, un manque d'expositions aux infections communes en bas âges pourrait engendrer, plus tard dans la vie de l'enfant, une réponse démesurée due à un système immunitaire naïf. Les associations (faibles) entre le risque de leucémie et l'allaitement et la présence en garderie pourraient s'inscrire dans cette optique de développement du système immunitaire (voir **Tableau I**) [17].

Tableau I: Les principaux facteurs de risque environnementaux de la leucémie et leur contribution. Adapté de J. Schüz et F. Erdmann, 2016 [17].

Facteurs de risque	Convaincant	Moyen	Faible	Contribution*
Pesticides postnataux			X	
Pesticides parentaux		X		Petite
Autres agents chimiques parentaux			X	
Pollution atmosphérique			X	
Radiations & CÉM				
Radiations thérapeutiques (fortes doses)	X			Petite
Radiations diagnostiques		X		Modeste
Radon		X		Modeste
Énergie nucléaire (opération normale)			X	
Ionisation parentale (lieu de travail)			X	
CÉM de radiofréquences			X	
CÉM de très basses fréquences		X		Petite
CÉM parentaux (lieu de travail)			X	
Style de vie, comportements et infections				
Tabagisme maternel			X	
Tabagisme paternel		X		Modeste
Consommation d'alcool maternelle			X	
Diète maternelle			X	
Prise d'acide folique maternelle [†]		X		Modeste
Présence en garderie [†]		X		Modeste
Faible stimulation immunitaire		X		Modeste
Profil d'infection virale	X			Marquée
Infections spécifiques			X	
Allaitement [†]		X		Modeste

Les facteurs de risques sont catégorisés (convaincants, moyen et faible) selon leur association avec la leucémie. CÉM, champs électromagnétiques. [†]Corrélation inverse (protection). *Petite (fraction attribuable estimée <5%), modeste (fraction attribuable estimée 5-15%), ou marquée (fraction attribuable estimée >15%).

Une contribution génétique importante est suggérée par la présence de certains syndromes, notamment ceux associés avec une instabilité génétique, qui montrent un risque plus élevé de développer une leucémie. À titre d'exemple, jusqu'à 9% des patients atteints d'anémie de Fanconi développeront une

leucémie, soit environ 2000 fois plus que dans la population générale [19]. Les enfants atteints du syndrome de Down ont quant à eux environ 2,5% (500 fois plus) de chance de développer une leucémie [20, 21]. De plus, des études d'association à l'échelle du génome (GWAS) ont pu relever certains polymorphismes hérités qui augmentent le risque de leucémie (**Tableau II**) [22], démontrant ainsi la composante génétique inhérente de cette maladie. Finalement, des mutations rares dans *PAX5* et *ETV6* ont été liées à un historique familial de leucémie et autres désordres hématologiques [23-28]. Bien qu'elles ne soient associées qu'à une infime partie des cas de leucémies, elles démontrent bien la forte pénétrance que peuvent avoir certaines altérations génétiques.

Tableau II: Exemples de polymorphismes associés à la leucémie. Adapté de T. Moriyama *et al.*, 2015 [22].

Gène	SNP ID	Phénotype	N	OR	P-value
<i>ARID5B</i>	rs7089424	Risque LAL	907	1.65 (1.54-1.76)	6.7×10^{-19}
	rs10821936		441	1.91 (1.6-2.2)	1.4×10^{-15}
<i>IKZF1</i>	rs4132601	Risque LAL	907	1.69 (1.58-1.81)	1.2×10^{-19}
	rs11978267		441	1.69 (1.4-1.9)	8.8×10^{-11}
<i>CEBPE</i>	rs2239633	Risque LAL	907	1.34 (1.22-1.45)	2.9×10^{-7}
<i>CDKN2A</i>	rs17756311	Risque LAL	2450	1.36 (1.18-1.56)	1.4×10^{-5}
<i>PIP4K2A</i>	rs7088318	Risque LAL	2450	1.40 (1.28-1.53)	1.1×10^{-11}
<i>GATA3</i>	rs3824662	Risque LAL	3107	1.31 (1.21-1.41)	8.6×10^{-12}
		Risque LAL <i>Ph-like</i>	511	3.85 (2.7-5.4)	2.2×10^{-14}
<i>TP63</i>	rs17505102	Risque LAL <i>ETV6-RUNX1</i>	1370	0.65 (0.52-0.75)	8.9×10^{-9}

SNP, *Single nucleotide polymorphism*; N, Échantillonnage; OR, *Odds ratio* (calculé avec un intervalle de confiance de 95%); LAL, leucémie aiguë lymphoblastique; Ph-like, *Philadelphia-like*.

1.2.3. Présentation clinique

La présentation clinique des leucémies est généralement associée à une déficience de cellules sanguines matures. La gravité des symptômes est variable selon le type de leucémie et reflète normalement le niveau de remplacement des cellules normales par les blastes dans la moelle osseuse. Ainsi, les patients leucémiques affichent dans plus de 75% des diagnostics les symptômes typiques associés à une anémie (diminution importante des érythrocytes), à une neutropénie (déficience en neutrophiles circulants) ainsi qu'à une thrombocytopénie (diminution des plaquettes). Dans certains cas plus rares (10-15% des diagnostics), une infiltration de blastes au niveau du système nerveux central (SNC) peut provoquer des maux de tête, des vomissements et des problèmes d'acuité visuelle. Une liste plus exhaustive des symptômes et signes cliniques est présentée dans le **tableau III** [29].

Tableau III: Symptômes des leucémies pédiatriques. Adapté de M.F. Colby-Graham et C. Chordas, 2003 [29].

Leucémie aiguë lymphoblastique	Leucémie aiguë myéloïde
Anémie: malaise, fatigue et pâleur	Pâleur
Thrombocytopénie: pétéchies et purpura	Fatigue
Saignements	Pétéchies
Neutropénie: fièvre	Fièvre
Hépatosplénomégalie	Infections
Lymphadénopathie	Lésions cutanées
Douleurs osseuses	Lymphadénopathie
Affectation du SNC (<10% des cas au diagnostic): augmentation de la pression intracrânienne, maux de tête, vomissements et troubles visuels	Affectation du SNC (<15% des cas au diagnostic): augmentation de la pression intracrânienne, maux de tête, vomissements et troubles visuels

SNC, Système nerveux central.

1.2.4. Traitements basés sur le risque de rechute

Le traitement de la leucémie est fait par une chimiothérapie (et radiothérapie dans certains cas) basée sur le niveau de risque de rechute. Cette classification permet d'adapter le traitement donné aux patients de manière à maximiser le taux de rémission tout en diminuant le plus possible les effets secondaires. Les patients susceptibles de bien répondre reçoivent donc un traitement standard avec des doses minimales (et donc moins toxiques) d'agents chimiothérapeutiques alors que ceux ayant une leucémie à plus haut risque entreprennent une thérapie plus soutenue. Ces protocoles intensifiés comprennent généralement des doses plus élevées d'agents chimiothérapeutiques en plus d'être complétés d'agents additionnels visant à mieux répondre à l'agressivité de la leucémie [30].

Plusieurs facteurs sont pris en considération pour déterminer le niveau de risque et le protocole de traitement de chaque patient (**Tableau IV**) [31, 32]. Il y a entre autres l'âge au diagnostic, le sexe, le compte initial de cellules blanches, l'immunophénotype, l'infiltration du SNC, les lésions chromosomiques macroscopiques (par analyses cytogénétiques) ainsi que la maladie résiduelle après la première phase du traitement. Plusieurs altérations génétiques ayant été identifiées récemment semblent avoir un impact prédictif sur la réponse aux traitements et le risque de rechute. Cependant, peu de ces altérations sont actuellement utilisées pour la stratification des patients [33, 34], mais ceci est appelé à changer dans un avenir rapproché [9].

Tableau IV: Facteurs influençant le pronostic des enfants atteints de LAL.
Adapté de K.R. Rabin et D.G. Poplack, 2011 [31].

Facteurs de risque	Favorable	Défavorable
Âge	1-10 ans	< 1 ou ≥ 10 ans
Sexe	Femelle	Mâle
Ethnicité	Caucasien, Asiatique	Noir, Hispanique
Compte de cellules blanches au diagnostic	< 50 000/μL	≥ 50 000/μL
Immunophénotype	Précurseurs-B	Lymphocytes T Lymphocytes B matures
Altérations génétiques	Hyperdiploïdie t(12;21) (ETV6-RUNX1)	t(9;22) (BCR-ABL1) Réarrangements <i>MLL</i> Hypodiploïdie iAmp(21) Surexpression de <i>CRLF2</i>
Atteintes extramédullaires (SNC, testicules)	Non	Oui
Réponse précoce aux traitements		
Prednisone	Compte absolu de blastes périphériques < 1 000/μL	Compte absolu de blastes périphériques > 1 000/μL
MRM dans le sang périphérique au jour 8 de l'induction	< 0.01-1%	> 1%
Morphologie de la moelle osseuse aux jours 8 et 15 de l'induction	M1	M2 ou M3

SNC, Système nerveux central; MRM, Maladie résiduelle minimale. M1, M2 et M3 font référence à la proportion de blastes dans la moelle osseuse: moins de 5%, de 5% à 25% et plus de 25%, respectivement.

Indépendamment du groupe de risque, tous les patients passent par une chimiothérapie en 3 étapes. Pour la LAL, la première étape dite d'induction dure 4 semaines et vise à éliminer rapidement la majorité des cellules cancéreuses afin d'atteindre un état de rémission. Dans la majorité des protocoles, les médecins administrent un régime composé de vincristine, de L-asparaginase, de glucocorticoïdes (prednisone ou dexaméthasone) et d'antracyclines (doxorubicin ou daunorubicin) [35-37]. La rémission initiale est atteinte dans environ 95% des cas de LAL. De la fraction restante, environ la moitié décède durant l'induction et

l'autre moitié est considérée comme résistante et aura un mauvais pronostic [38]. La deuxième étape dite de consolidation (ou d'intensification) dure de 6 à 8 mois et vise à empêcher l'apparition d'une rechute tout en prévenant les atteintes du SNC. À cet égard, plusieurs protocoles prévoient une chimiothérapie intrathécale à cette étape du traitement. En plus des agents chimiothérapeutiques donnés lors de l'étape d'induction, les médecins prescrivent notamment de la mercaptopurine et de fortes doses de méthotrexate. La troisième et dernière étape dite de maintenance peut s'étaler sur une période allant jusqu'à 2 ans. Au cours de cette étape, le patient prend oralement et de manière quotidienne de la mercaptopurine en plus d'une administration hebdomadaire de méthotrexate. La majorité des protocoles continuent également la chimiothérapie intrathécale pour restreindre les atteintes au SNC. Avec les avancées des traitements actuels, un enfant diagnostiqué avec la LAL a environ 85% de chance de survie sur 5 ans.

1.3 Génétique de la leucémie aiguë lymphoblastique de l'enfant

Environ le quart des nouveaux cas de cancers diagnostiqués avant l'âge de 15 ans sont des LAL, ce qui en fait le cancer le plus prévalent dans la population pédiatrique.

1.3.1. Modèle en deux étapes

Le très jeune âge auquel certains diagnostics de leucémie sont faits laisse présumer une initiation *in utero*. Dans environ 80% des cas de leucémie infantile

(diagnostiquée avant l'âge de 1 an), des analyses cytogénétiques confirment la présence d'une translocation impliquant le gène *MLL* (*KMT2A*; locus 11q23) avec divers partenaires [39, 40]. Le gène *AFF1* (*AF4*; locus 4q21) est retrouvé dans la moitié des cas des fusions impliquant *MLL* dans cette population. La fusion $t(4;11)(q21;q23)$ a donc fait l'objet d'un intérêt particulier visant à en faire une identification précoce. L'analyse de cas rares de jumeaux identiques étant tous deux atteints d'une leucémie infantile a démontré que les points de cassure de la translocation $t(4;11)(q21;q23)$ étaient identiques chez les pairs, suggérant fortement une incidence *in utero* [41, 42]. De plus, une méthode PCR a permis de confirmer la présence de cette translocation dans les taches de sang néonatales (*blood spots*; cartes de Guthrie) d'enfants ayant eu un diagnostic précoce (5 mois à 2 ans) de leucémie [43]. Ces études ont permis de mettre en évidence l'occurrence prénatale de cette translocation.

Des études similaires sur des jumeaux identiques et des cartes de Guthrie ont été faites sur la translocation $t(12;21)(p13;q22)$, la plus fréquente dans la LAL de type B (discutée plus en détail dans la section 1.3.3.3.3). Une fois encore, des points de cassure identiques ont été identifiés dans les paires de jumeaux. Cependant, certains ont développé la leucémie jusqu'à 9 ans plus tard que leur frère ou sœur [44]. La résultante de la $t(12;21)$, la fusion *ETV6-RUNX1*, a également été retrouvée dans environ 1% de la population, soit environ 100 fois plus que l'incidence de la B-LAL [45]. Ces observations suggèrent qu'à l'inverse de la fusion *MLL-AF4*, la fusion *ETV6-RUNX1* n'est pas suffisante pour induire à elle seule la leucémie. Certaines mutations ont dû s'ajouter au contexte

préleucémique initié par ETV6-RUNX1 pour que la leucémie se développe chez ces patients. Cette hypothèse est appuyée par des études chez la souris démontrant que la surexpression d'ETV6-RUNX1 n'engendre pas de phénotypes hématologiques délétères [46] à moins d'être combinée à un contexte hypermutateur [47, 48]. Ainsi, certaines leucémies pédiatriques suivraient un modèle en deux étapes (*two-hit hypothesis*; **Figure V** [18]) où un clone préleucémique serait généré *in utero* par l'acquisition de certaines altérations génétiques (p. ex. ETV6-RUNX1), suivie d'une période de latence jusqu'au moment d'être la cible de mutations coopératives qui enclencheraient le processus leucémogénique [49].

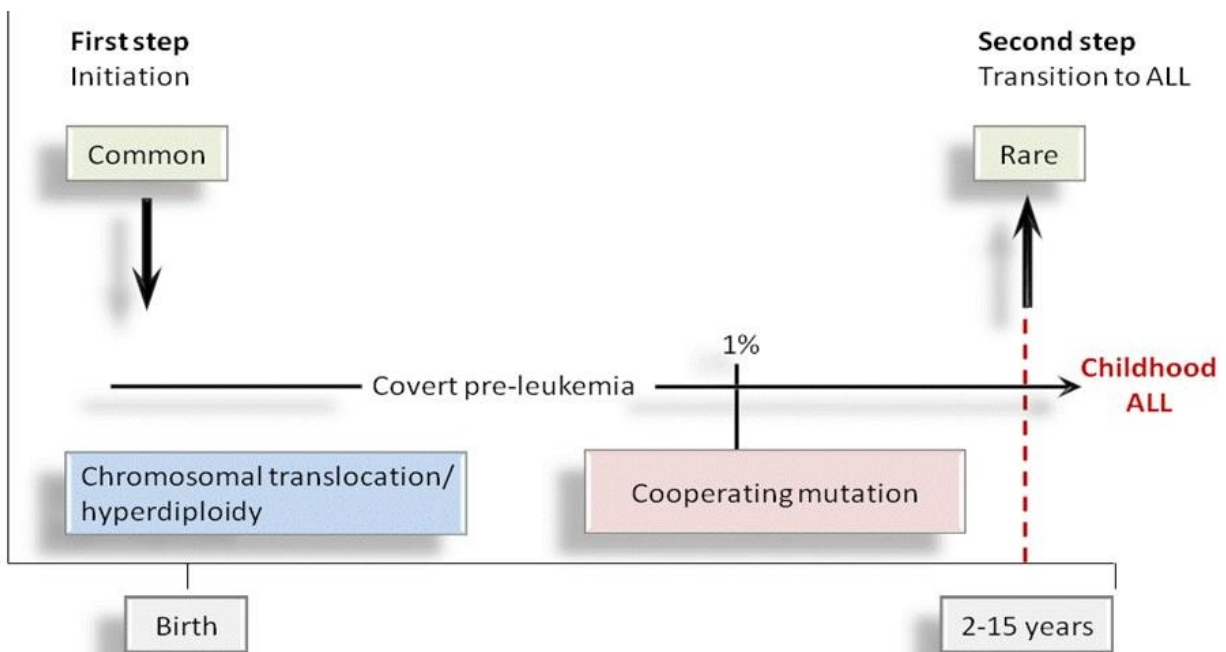


Figure V: Modèle en deux étapes des leucémies pédiatriques. ALL, *Acute lymphoblastic leukemia*. Adaptée de M. Greaves, 2006 [18].

1.3.2. L'aneuploïdie

Dans la LAL de type B, l'hyperdiploïdie est observée dans environ 25% des cas. Tout comme la translocation t(12;21)(p13;q22), il a été démontré que l'acquisition de copies complètes de chromosomes avait une origine prénatale et ne serait pas suffisante pour initier de façon indépendante la leucémogénèse [50]. Toutefois, le gain de chromosomes se fait de façon non aléatoire, suggérant que ces duplications procurent un avantage sélectif à la cellule leucémique. On observe de manière préférentielle la duplication des chromosomes 4, 6, 10, 14, 17, 18, 21 et X (**Figure VI**) [51], mais on ignore encore comment elles contribuent à la transformation cellulaire. Parmi les hypothèses émises, la plus probable serait un déséquilibre de l'expression des gènes (*gene dosage*) retrouvés sur les chromosomes dupliqués. Des études d'expression ont démontré que la majorité des gènes présents sur ces chromosomes étaient surexprimés et permettaient une classification de ce sous-type moléculaire de B-LAL [52-54]. Il est facilement concevable que seulement une fraction de ces gènes surexprimés contribuent réellement au phénotype préleucémique, mais leur identité demeure toujours inconnue.

La perte de chromosomes (l'hypodiploïdie; <44 chromosomes) est également retrouvée chez les patients atteints de B-LAL mais dans une proportion beaucoup plus faible, soit dans environ 1% des cas. Comme pour l'hyperdiploïdie, la perte de chromosomes est non aléatoire. Les chromosomes 8, 10, 14, 18, 21 et chromosomes sexuels demeurent diploïdes dans la quasi-totalité des cas [55]. L'hypodiploïdie est souvent associée à des mutations au

niveau des gènes de la voie de signalisation RAS (p. ex. *NRAS*, *KRAS*, *MAPK1*, *FLT3* et *PTPN11*) et de *TP53* [56, 57]. Dans le cas de *TP53*, les mutations sont retrouvées dans le matériel normal (non tumoral) chez près de la moitié des patients, suggérant donc que cette hypodiploïdie serait une conséquence du syndrome de Li-Fraumeni [57]. Par conséquent, le pronostic de ce sous-type est défavorable (voir **Tableau IV**).

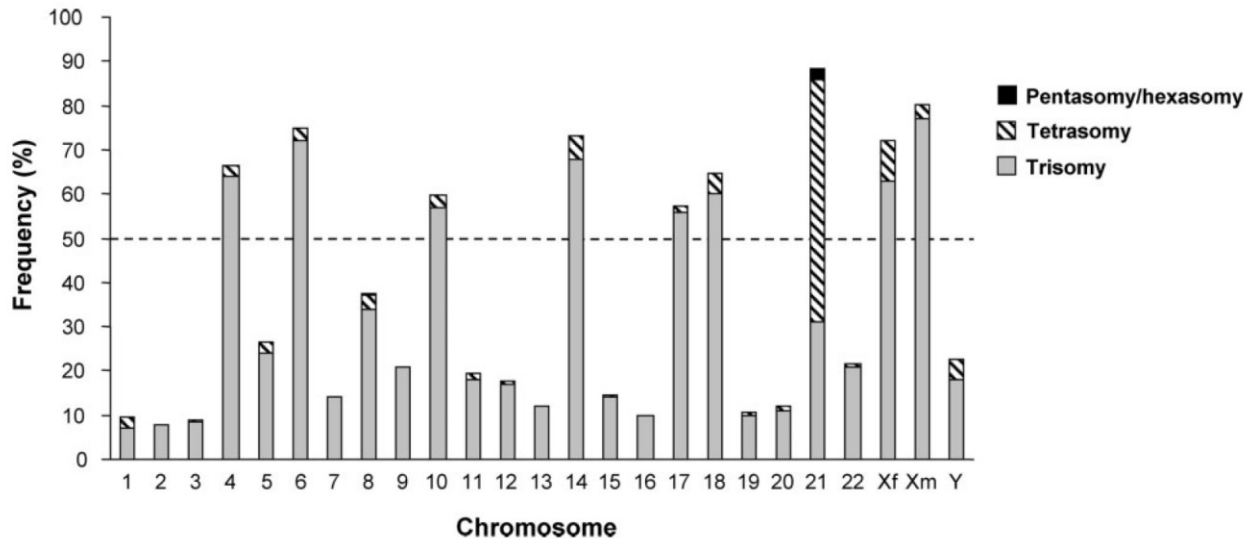


Figure VI: Fréquence des gains de chromosomes chez 414 patients (<18 ans) atteints de LAL de type B. Xf est le gain du chromosome X chez les filles; Xm chez les garçons. Reproduite de K. Paulsson et B. Johansson, 2009 [51].

1.3.3. Les translocations chromosomiques

Les translocations sont un échange réciproque de matériel génétique entre chromosomes (homologues ou non). Une translocation est dite "équilibrée" s'il n'y a aucune perte de matériel chromosomique ou à l'inverse "déséquilibrée" s'il y a perte (ou gain) de parties de chromosomes. Les translocations sont

retrouvées dans plusieurs types de cancers, dont les leucémies. Les translocations les plus fréquemment retrouvées engendrent soit la surexpression d'un gène par la juxtaposition à un promoteur fort ou bien la génération d'une protéine de fusion (chimère). Ces chimères sont souvent associées à une fonction kinase aberrante (constitutivement active) ou bien à des fonctions de régulation de l'expression des gènes.

1.3.3.1. Surexpressions

L'expression aberrante du facteur de transcription *TAL1* est fréquemment retrouvée dans la LAL de type T [14]. Cette surexpression est souvent associée à la translocation t(1;14)(p33;q11) qui juxtapose *TAL1* en aval du promoteur des gènes *TRA* ou *TRD* (T cell receptor- α/δ) [58, 59].

Dans la LAL de type B, le gène *CRLF2* (*cytokine receptor-like factor 2*) est surexprimé chez environ 8% des patients, dont environ la moitié est associée avec le sous-type *BCR-ABL1-like* (voir la section 1.3.3.2). Le gène *CRLF2* est situé sur la partie pseudoautosomale des chromosomes sexuels Xp22.3/Yp11.3. Ce locus est trouvé recombinaison aux régions *enhancers* des chaînes lourdes des immunoglobulines (*IGH@*) au niveau 14q32.33 favorisant ainsi son expression. *CRLF2* compose, avec le *IL7RA* (*IL7 receptor alpha*), le récepteur au *TSLP* (*Thymic stromal-derived lymphopoietin*). Cette cytokine possède un effet pro-prolifératif sur les cellules précurseur-B via une activation de *STAT5* [60, 61]; la surexpression de son récepteur favoriserait donc ce processus. De façon similaire à *CRLF2*, la surexpression des gènes *ID4* (*inhibitor of DNA binding 4*) et *EPOR* (*erythropoietin receptor*) est observée (beaucoup plus rarement) dans

certains cas de B-LAL suite à leur réarrangement avec le locus *IGH@* par les translocations t(6;14)(p22;q32) et t(14;19)(q32;p13), respectivement [62].

1.3.3.2. Les protéines kinases; chromosome de Philadelphie

L'activation constitutive de kinases par des réarrangements chromosomiques est un processus oncogénique commun. Plusieurs récepteurs tyrosine kinases sont d'ailleurs la cible de translocations dans les cancers hématologiques [63]. Dans le cas de la B-LAL pédiatrique, *ABL1* est la protéine kinase la plus souvent impliquée dans un réarrangement (~ 3% des cas) [4]. *ABL1* se trouve fusionné avec le gène *BCR* par la translocation t(9;22)(q34;q11). Il est à noter que cette même translocation (appelée aussi le chromosome de Philadelphie) est retrouvée dans environ 95% des cas de leucémie myéloïde chronique [2], mais génère un produit de fusion différent.

En effet, il existe 3 formes de la protéine de fusion selon les points de cassure dans le gène *BCR* (**Figure VII**). La forme la plus usuelle (M-BCR ou p210; associée à la leucémie myéloïde chronique) est généralement constituée de la partie 5' de *BCR* comprenant les 13-14 premiers exons et de la partie 3' de *ABL1* à partir de l'exon 2. Dans la LAL, la forme la plus fréquente (m-BCR ou p190) comprend seulement le premier exon de *BCR* [64, 65]. Le domaine *coiled-coil* de *BCR* (encodé par l'exon 1 et donc présent dans ces 2 formes) permet à la protéine de fusion BCR-ABL1 de former un tétramère qui active constitutivement la fonction kinase de *ABL1* contenue dans son domaine SH1 (*src homology domain*) [66, 67]. L'activité kinase de BCR-ABL1 favorise la prolifération cellulaire tout en inhibant la différenciation et l'apoptose des précurseurs hématopoïétiques

en modulant entre autres les voies de signalisation JAK/STAT, PI3K/AKT et MAPK [68]. L'imatinib, un inhibiteur du site actif du domaine kinase SH1 de BCR-ABL1, a permis depuis son introduction en clinique de réduire de façon significative la mortalité associée à ce sous-type de B-LAL [69, 70].

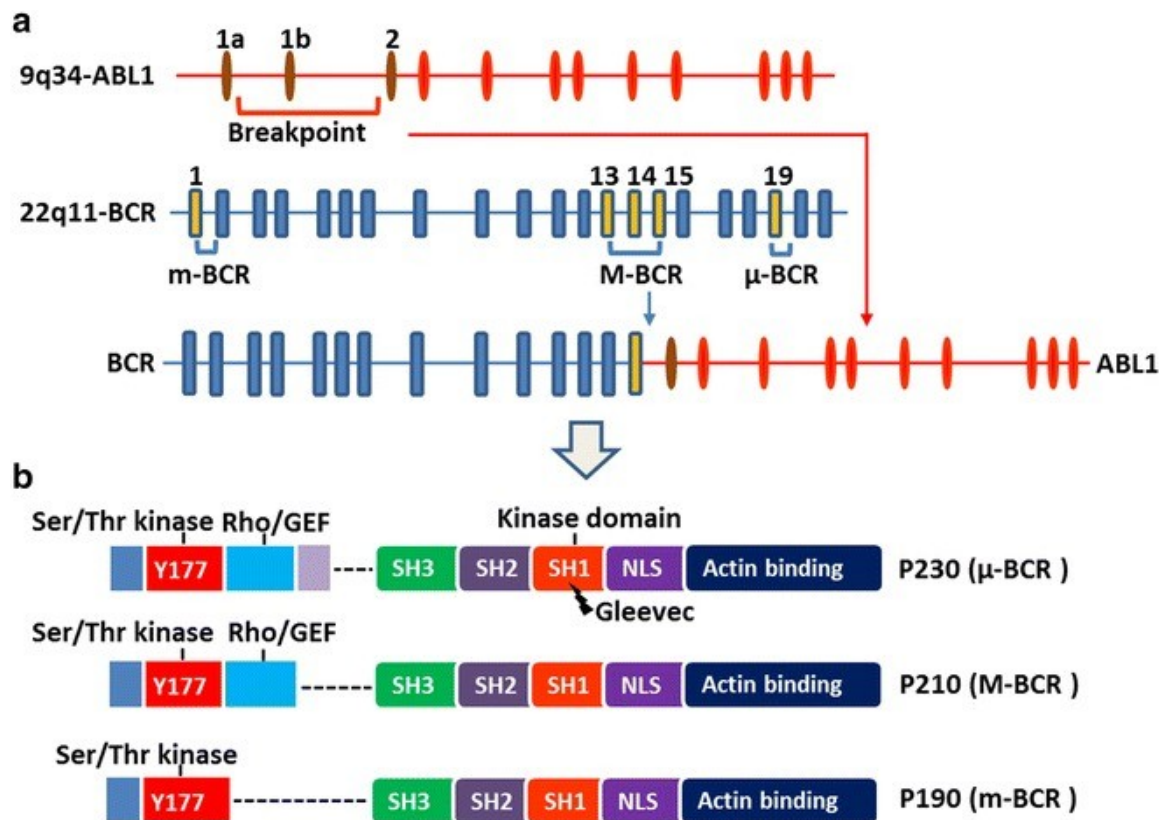


Figure VII: Fusions BCR-ABL1. **A.** Points de cassures associés aux 3 formes majeures de la t(9;22)(q34;q11) indiquées m-BCR, M-BCR et μ-BCR. **B.** Domaines fonctionnels présents dans les protéines de fusions. Le domaine *coiled-coil* en N-terminal est marqué en bleu (non identifié sur la figure). Le domaine ciblé par l'imatinib (Gleevec) est également montré. Reproduite de Z.J. Kang *et al.*, 2016 [68].

Récemment, un nouveau sous-type de B-LAL a pu être identifié sur la base d'études d'expression. Environ 9% des cas de B-LAL affichent un profil d'expression génique similaire à celui des patients BCR-ABL1 positifs sans toutefois être porteurs de cette translocation. Dans la majorité des cas, ces patients possèdent des délétions ou mutations dans des gènes impliqués dans le développement des lymphocytes B comme *IKZF1*, *TCF3* et *PAX5* [71]. On observe également la surexpression de *CRLF2* (section 1.3.3.1) dans la moitié des cas. Un séquençage en profondeur de 15 patients BCR-ABL1-like a récemment permis d'identifier une série d'altérations génétiques pouvant contribuer à l'activation des voies de signalisation des kinases (comme la voie JAK/STAT) caractéristiques des profils d'expression observés chez ces patients. Des réarrangements impliquant les gènes *PDGFRB*, *ABL1*, *JAK2* et *EPOR* font partie de ces altérations [72]. Ce nouveau sous-type de B-LAL (*Philadelphia-like*) est le sujet de plusieurs initiatives de recherche visant à identifier les kinases impliquées pouvant être ciblées par des agents thérapeutiques de façon similaire à l'imatinib dans le traitement des patients BCR-ABL1 positifs.

1.3.3.3. Les facteurs de transcription

Dans la B-LAL, les translocations impliquant des facteurs de transcription sont les plus fréquentes. Dans ces cas, la translocation génère une protéine de fusion ayant de nouvelles propriétés transcriptionnelles uniques menant à une régulation aberrante de l'expression de certains gènes impliqués dans la prolifération et la maturation des précurseurs B.

1.3.3.3.1. TCF3-PBX1

La translocation t(1;19)(q23;p13) est retrouvée dans environ 4% des cas de B-LAL [9]. Elle fusionne les deux facteurs de transcription *TCF3* et *PBX1* de manière à conserver le domaine de transactivation de *TCF3* et le domaine *homeobox* de liaison à l'ADN de *PBX1* [73]. Le domaine de transactivation de *TCF3* modifie drastiquement l'activité de *PBX1*; les cibles du complexe transcriptionnel HOX/*PBX1* sont conséquemment surexprimées par TCF3-PBX1 [74]. Comme ces cibles sont impliquées dans la différenciation hématopoïétique (notamment *WNT16* [75]), leur dérégulation peut avoir un impact sur l'initiation et le développement de la B-LAL. De façon similaire à la fusion ETV6-RUNX1, la TCF3-PBX1 favoriserait un état préleucémique qui devra être complété d'une/de mutation(s) additionnelle(s) pour aboutir à un phénotype leucémique complet [76] (voir section 1.3.1).

1.3.3.3.2. Réarrangements *MLL*

Le gène *MLL* (*KMT2A*) n'est pas un facteur de transcription au sens propre car il ne possède pas la capacité de lier directement l'ADN. Il agit néanmoins de façon importante dans la régulation transcriptionnelle via l'activité méthyltransférase de son domaine SET sur la lysine 4 de l'histone 3 (H3K4) et par le recrutement de remodeleurs de la chromatine (p. ex. CBP/p300 et hMOF [77, 78]). Sa fonction est donc associée à un relâchement des nucléosomes permettant une activation de la transcription par d'autres facteurs. Son implication dans l'hématopoïèse se fait notamment par la régulation des gènes de la famille HOX [39].

MLL est impliqué dans une grande variété de réarrangements chromosomiques qui représentent environ 6% des cas de B-LAL [9]. Ils sont très fréquents chez les patients diagnostiqués avant l'âge de 1 an [39, 40]. Le gène *AFF1* (*AF4*) est le partenaire de fusion de *MLL* le plus fréquemment retrouvé dans la B-LAL. *AF4*, comme *MLL*, n'est pas un facteur de transcription, mais est directement impliqué dans la régulation de l'expression par l'ARN polymérase II. En ce sens, *AF4* interagit et active le *Pol II transcription elongation factor b* (*P-TEFb*) au sein du complexe coactivateur AEP (*AF4 family/ENL family/P-TEFb*) qui phosphoryle le domaine C-terminal (CTD) de l'ARN polymérase II pour initier l'élongation [79]. *AF4* peut également s'associer au *selectivity factor I* (*SL1*) pour recruter des *TATA-binding proteins* (TBP) au niveau des promoteurs et ainsi initier la transcription [80].

Ainsi, la protéine de fusion résultante de la t(4;11)(q21;q23) c.-à-d. *MLL-AF4* surexprimerait les gènes normalement régulés par *MLL* due à la forte activation de l'ARN polymérase II grâce à la portion *AF4* de la fusion. Un mécanisme similaire est proposé pour les fusions entre *MLL* et d'autres membres du complexe AEP notamment *AF9* dans la t(9;11)(p22;q23). La surexpression des gènes *HOX* caractéristiques des patients réarrangés pour *MLL* abonde d'ailleurs en ce sens [81]. Finalement, contrairement aux autres fusions retrouvées dans la B-LAL, la fusion *MLL-AF4* semble suffisante à elle seule pour le développement de la maladie. Bien que les phénotypes murins soient parfois peu représentatifs de ceux observés chez l'humain, l'expression de *MLL-AF4* dans les souris induit facilement l'apparition de maladies hématologiques comme des lymphomes et

des leucémies [82]. Aussi, très peu de mutations sont retrouvées dans les génomes des patients porteurs de cette translocation [12, 83], corroborant ainsi son fort potentiel oncogénique.

1.3.3.3.3. ETV6-RUNX1

La translocation t(12;21)(p13;q22) est observé dans environ 25% des cas de B-LAL, ce qui en fait le réarrangement le plus fréquent dans ce type de cancer [9]. Comme les autres translocations impliquant des facteurs de transcription, il y a génération d'une protéine de fusion. Celle-ci produit une chimère entre les facteurs de transcription *ETV6* et *RUNX1 (AML1)* [84], deux facteurs clés dans le développement du système hématopoïétique [85, 86]. Tout comme *ETV6* (discuté plus en détail dans la section 1.4.4.1), le rôle de *RUNX1* dans l'hématopoïèse a été considérablement étudié [87]. En plus d'occuper une fonction critique dans la mise en place de l'hématopoïèse définitive [88, 89] par la modulation de l'expression des gènes *GFI1* et *GFI1B* [90], *RUNX1* est nécessaire au maintien du niveau de cellules souches hématopoïétiques [91] et à la maturation adéquate des précurseurs myéloïdes [92] et lymphoïdes [93] dans la moelle osseuse. Pour ce faire, *RUNX1* induit des modifications épigénétiques aux loci de certains facteurs hématopoïétiques, notamment PU.1 (*SPI1*), permettant d'initier des programmes transcriptionnels impliqués dans la différenciation [94, 95].

La protéine de fusion résultante de la t(12;21)(p13;q22) contient les domaines d'homo/hétérodimérisation et de répression centrale d'*ETV6* (voir sections 1.4.2 et 1.4.3 pour plus de détails) en plus de la totalité des domaines fonctionnels

codés par *RUNX1* (**Figure VIII**) [96]. Seul le domaine de liaison à l'ADN d'*ETV6* n'est pas inclus dans la fusion. Ainsi, les cibles transcriptionnelles normalement activées par *RUNX1* sont réprimées par ETV6-RUNX1 via le recrutement de corépresseurs par la portion *ETV6* [97]. De plus, ETV6-RUNX1 possède un effet dominant négatif sur la forme sauvage d'*ETV6* [98]. La dérégulation de leurs cibles transcriptionnelles bouleverserait donc le processus normal d'hématopoïèse et participerait au développement de la leucémie. ETV6-RUNX1 régule également des cibles transcriptionnelles distinctes pouvant contribuer à la transformation cellulaire [99].

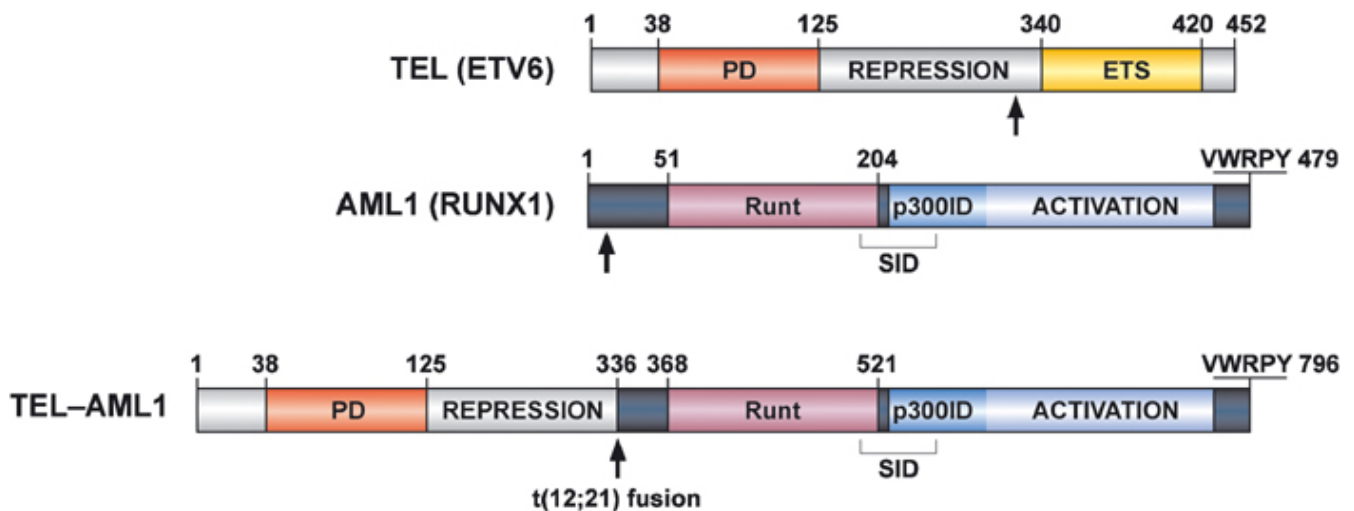


Figure VIII: Protéine de fusion ETV6-RUNX1. Les domaines fonctionnels de chaque protéine sont indiqués avec leur correspondance en acides aminés. Les flèches indiquent les points de cassure. Reproduite de A. Zelent *et al.*, 2004 [96].

La présence de la t(12;21)(p13;q22) dans une aussi grande proportion de cas de B-LAL [9] combinée avec la présence systématique de la protéine de fusion

ETV6-RUNX1 chez ces patients [100, 101] lui suggèrent un rôle important dans la leucémogénèse. Bien que la présence d'ETV6-RUNX1 ne soit pas suffisante pour induire de façon indépendante la leucémie chez l'humain [44, 45, 49] ou la souris [46], plusieurs dérèglements dans la prolifération et la maturation des précurseurs hématopoïétiques ont pu être démontrés (**Figure IX**) et sont conséquents avec l'état préleucémique associé à cette fusion [102].

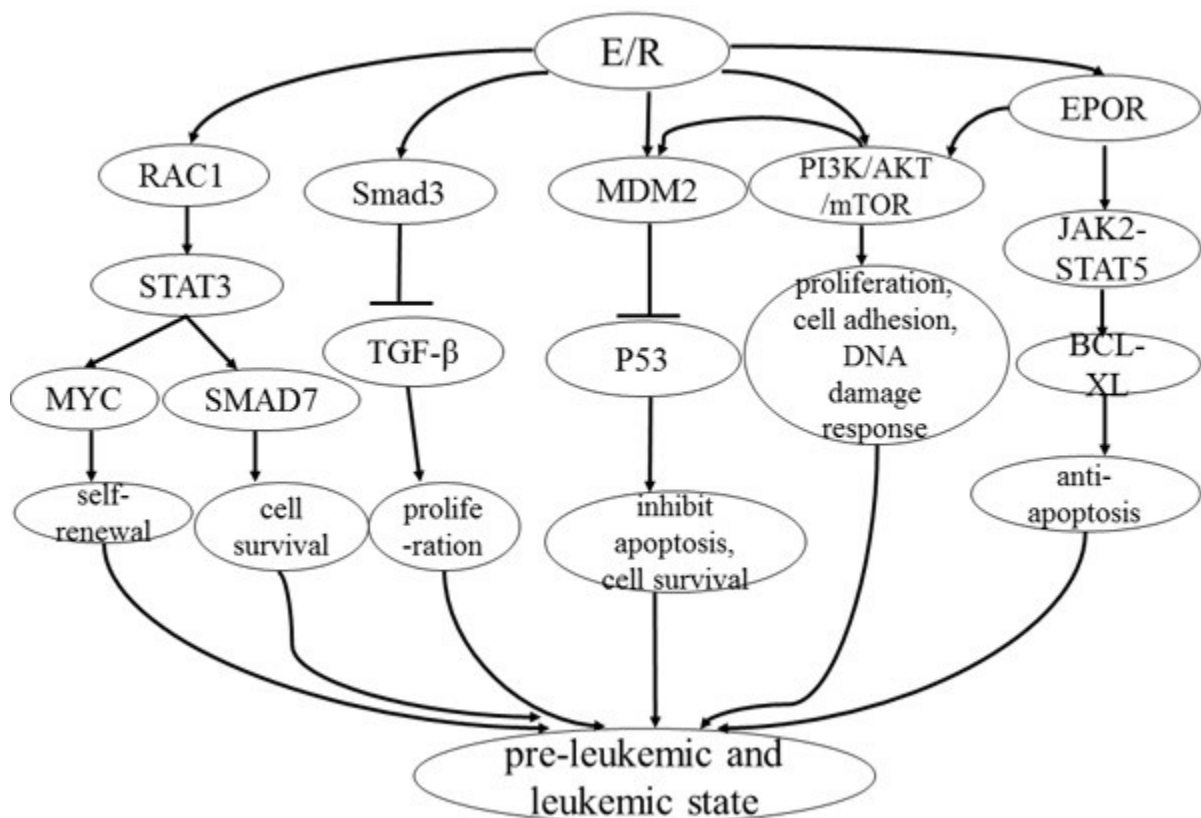


Figure IX: Voies de signalisations et phénotypes cellulaires associés à l'expression de la protéine de fusion ETV6-RUNX1 (E/R). Reproduite de C. Sun *et al.*, 2017 [103].

ETV6-RUNX1 induit l'activation de *RAC1* [104] qui, en complexe avec *RACGAP1*, favorise la phosphorylation du facteur de transcription *STAT3* par les kinases Janus associées aux récepteurs de l'interleukine-6 (*IL6R*) [105]. L'activation de *STAT3* promeut la transcription du proto-oncogène *MYC* qui aura une incidence sur la survie et l'autorenouvellement des cellules [104]. Comme ETV6-RUNX1 possède des propriétés transcriptionnelles, le mécanisme conduisant à l'activation de *RAC1* demeure inconnu.

ETV6-RUNX1 agit également sur l'intégration de la voie de signalisation TGF- β . En ce sens, ETV6-RUNX1 réduit l'impact de cette voie sur l'inhibition de la prolifération en se liant au régulateur transcriptionnel *SMAD3*, l'empêchant ainsi de transcrire l'inhibiteur du cycle cellulaire *p27^{KIP1}* malgré l'activation du récepteur au TGF- β [106]. La voie TGF- β est impliquée dans la réponse immunitaire et expliquerait la sélection de clones préleucémiques (ETV6-RUNX1 positifs) lors d'une infection, abondant dans le sens de l'hypothèse virale associée au développement de la LAL (voir section 1.2.2).

Bien que le mécanisme d'action d'ETV6-RUNX1 soit toujours nébuleux dans la transcription de *MDM2*, son implication dans la voie de p53 est néanmoins importante à mentionner. En effet, la présence d'ETV6-RUNX1 augmente l'expression de *MDM2*, un répresseur important de p53 [107]. L'inhibition de la signalisation médiée normalement par p53 peut sans aucun doute avoir un effet dans le développement de la LAL [108]

L'activation de la signalisation PI3K/AKT/mTOR est également associée avec la présence d'ETV6-RUNX1. L'implication de cette voie dans la LAL est importante puisque son inhibition réduit la prolifération et la survie cellulaire qui se traduit par la diminution de la fréquence et une augmentation de la latence des leucémies induites chez la souris [109]. Même si les mécanismes moléculaires ne sont pas encore bien compris, l'activation de la voie PI3K/AKT/mTOR par ETV6-RUNX1 pourrait être attribuable à l'activation de *EPOR* [110]. En plus de ce rôle, une forte activité de *EPOR* peut promouvoir la survie cellulaire via la cascade signalétique JAK/STAT [111].

1.4 ETV6

1.4.1. ETV6 et le modèle en deux étapes

Plus d'altérations de type CNA (*Copy number alterations*) sont retrouvées dans la B-LAL t(12;21)(p13;q22) positive que dans la majorité des autres sous-types moléculaires [112], renforçant les observations que ETV6-RUNX1 n'est pas suffisant pour induire la leucémie. Parmi ces CNA (**Tableau V**), la plus fréquente est la délétion du locus 12p13 où est localisé le gène *ETV6* [113, 114]. En plus de la haute fréquence (70%) de la del(12p), des études au niveau des patients B-LAL t(12;21)(p13;q22) positifs n'ont pu détecter la protéine *ETV6* sauvage, et ce malgré la présence de l'allèle résiduel [101, 115]. Ces résultats suggèrent que l'inactivation complète d'*ETV6* est un événement important dans la transition de l'état préleucémique à leucémique.

Tableau V: Altérations génétiques secondaires communes (fréquence >5%) de type CNA (Copy number alterations) répertoriées dans les cas de B-LAL ETV6-RUNX1 positive. Adapté de C. Sun *et al.*, 2017 [103].

Chromosome	Anomalie génétique	Fréquence (%)	Gènes candidats possibles
12	del(12p) del(12)(p11-13) del(12)(p13.2) del(12)(p21.3)	18-69	ETV6, CDKN1B, BCL2L14, BTG1
6	del(6q) del(6)(q21-27) der(6)t(X;6)	13-37	AIM1, PRDM1, FOXO3, CCNC, FYN
9	del(9p) del(9)(p13-24) del(9)(p21.3) del(9)(p13.2)	7-29	CDKN2A/B, PAX5, MTAP, JAK2, P14ARF, P16INK4a/ARF
21	*gain(21) gain(21q)	10-25	RUNX1
5	del(5)(q31.3-33.3)	23	NR3C1, EBF1
14	del(14)(q32.33) del(14)(q11.2)	3-21	DPF3
3	del(3)(p21) del(3)(q26.32) del(3)(q13.2) del(3)(p14.2)	3-21	LIMD1, ARPP-21, ULK4, FHIT, CD200, BTLA, TBL1XR1
7	del(7)(q34) del(7)(p14.1)	6-18	IKZF1
4	gain(4q) del(4)(q31.23)	6-17	NR3C2, YIPF7, ARHGAP10
11	del(11q) del(11)(q22-23) del(11)(q23-25) del(11)(p13)	6-13	CD44, RAG1, RAG2, BACL2, GNG3, HNRPUL2, TTC9C, ATM, KMT2A
19	del(19)(q13.11) del(19)(q13.32)	6-13	CEBPA, UHRF1, GRLF1, NPAS1, TMEM160
22	del(22)(q11.22)	13	-
X	loss(X) (femelles) gain(Xq) (mâles) gain(Xp)	4-11	SPANXB, HMGB3, FAM50A, HTATSF1
13	del(13q) del(13)(q14-34) del(13)(q12.2)	5-10	RB1, SERP2, DLEU2, STBP4, TRIM13, KCNRG, MIRN16-1, MIRN15A, DLEU1, DLEU7
1	del(1)(q31.3)	10	TROVE2, GLRX2, CDC73, B3GALT2, PDE4B
15	del(15)(q15.1) del(15)(q26.1)	10	LTK, MIRN626
2	del(2)(p25.3)	9	-
8	del(8p) del(8)(p11-23)	6-8	TOX
16	gain(16)	6	-

* Copie additionnelle du chromosome 21 ou du der(21)t(12;21)(p13;q22).

1.4.2. La famille ETS

ETV6 est un membre de la grande famille des facteurs de transcription ETS (E26 transformation-specific). Chez l'humain, cette famille est composée de 27 gènes subdivisés en 11 groupes (**Figure X**) [116]. Tous partagent le domaine ETS, un domaine conservé de liaison à l'ADN de type *winged helix-turn helix* qui reconnaît de façon spécifique la séquence consensus 5'-GGA(AT)-3' (séquence EBS; *ETS binding site*). Il existe cependant une spécificité des facteurs ETS basée sur leurs préférences au niveau des nucléotides flanquant cette séquence commune [117]; *ETV6* lie avec une forte affinité la séquence CCCGGAAGT. L'interaction des facteurs ETS avec d'autres facteurs de transcription ou cofacteurs leur confère également une grande spécificité [118]. Plusieurs des membres de cette famille, dont *ETV6*, contiennent un second domaine conservé: le domaine PNT (*Pointed*). Ce domaine, très comparable au domaine SAM (*Sterile alpha motif*), est responsable de certaines interactions protéines-protéines nécessaires à l'homo ou l'hétérodimérisation des facteurs ETS.

La surexpression et les translocations chromosomiques impliquant les facteurs ETS sont fréquemment observées dans une grande variété de cancers. Citons à titre d'exemples *ESE-1*, *PEA3*, *ETS-1*, *ETS-2*, *ERM* et *ER81* qui sont surexprimés dans les cas de cancers du sein et sont associés à la présence de métastases [119]. La surexpression de ces facteurs module l'expression de gènes cibles (p. ex. *HER2/NEU*, *MMPs*, *BCL2* et *VEGF*) ayant des effets notamment sur la prolifération, la migration, l'apoptose et l'angiogenèse [120]. Dans le cancer de la prostate, un marqueur important de la maladie est la

surexpression du facteur *ERG* via sa fusion avec le gène *TMPRSS2* [121]. Le promoteur de ce gène répond aux hormones androgènes et induit fortement l'expression d'*ERG*. Dans ces cellules, la surexpression d'*ERG* promeut l'expression de l'oncogène *c-MYC* et la répression des gènes *PSA* et *SLC45A3* associés à la différenciation épithéliale [122]. Dans la leucémie, le gène *ETV6* est le facteur ETS le plus fréquemment impliqué dans des réarrangements (voir section 1.4.5.1).

ETS family	Members	Structure	Binding specificities	
			Class	High affinity binding sequence
ETS	ETS1, ETS2		I	ACCGGAAGT
ERG	ERG, FLI1, FEV		I	ACCGGAAGT
GABPA	GABPa		I	ACCGGAAGT
ELF	ELF4, ELF2, ELF1		II	CCCGGAAGT
ESE	ESE1, ESE2, ESE3		II	CCCGGAAGT
ERF	ETV3, ERF		I	ACCGGAAGT
ETV6	ETV6, ETV7, YAN		II	CCCGGAAGT
PEA3	PEA3, ERM, ER81		I	ACCGGAAGT
SPI	SPII, SPIIB, SPIIC		III	AGAGGAAGT
TCF	ELK1, ELK4, ELK3		I	ACCGGAAGT
PDEF	PDEF/SPDEF		IV	ACCGGATGT

Figure X: Classification des facteurs de transcription de la famille ETS. La classe (I,II,III et IV) fait référence aux motifs consensus de liaison à l'ADN. AD, *Activation Domain*; RD, *Repression Domain*; PNT, *Pointed*. Reproduite de A. Kar et A. Gutierrez-Hartmann, 2013 [116].

1.4.3. Structure d'*ETV6*

Le gène *ETV6* est encodé par 8 exons distribués sur environ 250kb sur le bras court du chromosome 12p13.1 [123]. Bien que des isoformes aient déjà été identifiés (potentiellement par des mécanismes aberrants d'épissage) [124], *ETV6* encode généralement qu'un seul ARNm et n'est pas impliqué dans des événements d'épissage alternatif. Il existe cependant un codon initiateur alternatif qui génère une forme tronquée de la protéine *ETV6* commençant au codon 43 (*ETV6*-M43). La forme tronquée contient tous les domaines fonctionnels (**Figure XI**) mais montre des différences au niveau de sa localisation cellulaire [125] et de son activité transcriptionnelle [126]. Somme toute, très peu d'études se sont intéressées aux différences fonctionnelles entre les deux isoformes.

ETV6 est un facteur de transcription de la famille ETS défini par ses domaines conservés PNT et ETS qui sont encodés par les exons 3 et 4 et 6 à 8, respectivement (**Figure XI**). Contrairement à la majorité des facteurs ETS, *ETV6* est un répresseur transcriptionnel [127, 128]. Sa fonction particulière est attribuable à un domaine central codé par l'exon 5 qui permet une interaction avec des protéines responsables de la répression transcriptionnelle (voir section 1.4.6.1). Le domaine ETS de liaison à l'ADN est également responsable de la translocation nucléaire d'*ETV6*. En effet, l'ablation de la partie C-terminale d'*ETV6* [129] ainsi que plusieurs mutations localisées dans son domaine ETS [25-27] empêchent sa présence au noyau et conséquemment sa liaison à l'ADN. En plus de ses 3 domaines, *ETV6* possède un domaine d'autoinhibition, le C-

terminal inhibitory domain (CID), localisé entre les acides aminés 416 à 436 (**Figure XI**). La présence de ce domaine bloque (partiellement) l'accessibilité du domaine ETS et réduit l'affinité d'*ETV6* pour l'ADN. L'homodimérisation permet de contrevenir à cette autoinhibition par le CID et facilite la liaison à l'ADN [130, 131].

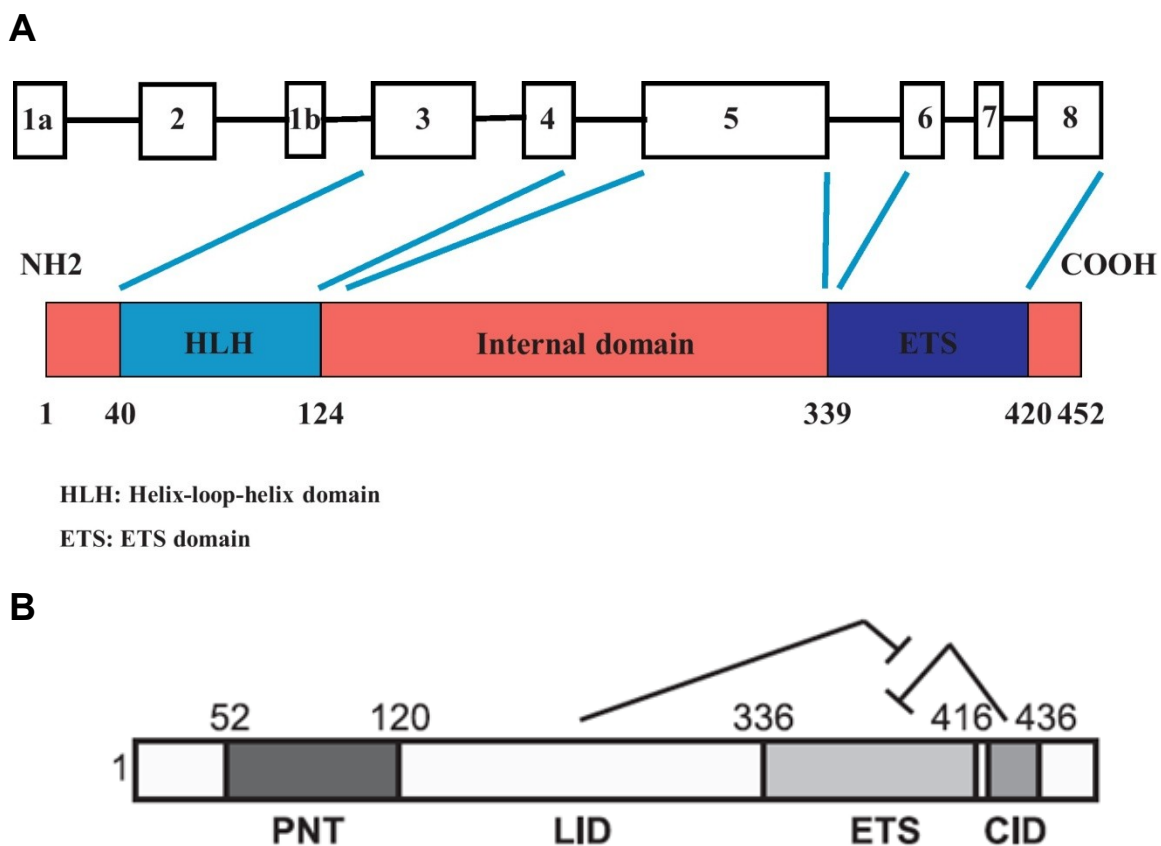


Figure XI: Structure d'*ETV6*. **A.** Structure du gène *ETV6* et sa correspondance au niveau protéique; les exons et les acides aminés sont indiqués. 1a fait référence au premier codon initiateur menant à la traduction de l'isoforme long *ETV6*-M1 et 1b au codon initiateur alternatif se traduisant par l'isoforme court *ETV6*-M43. HLH, *Helix-loop-helix domain* (ou domaine PNT). Reproduite de E. De Braekeleer *et al.*, 2012 [132]. **B.** Relation d'inhibition entre les domaines fonctionnels d'*ETV6*. PNT, *Pointed*; LID, *linker inhibitory damper*; CID, *C-terminal inhibitory domain*. Reproduite de S.M. Green *et al.*, 2010 [130].

1.4.4. Fonction d'*ETV6*

1.4.4.1. *ETV6* dans l'hématopoïèse

Les premières études visant à identifier le rôle biologique d'*ETV6* ont été réalisées avec des souris *knock-out*. La délétion complète d'*ETV6* dans ces souris provoquait la mort des embryons due à un défaut de l'angiogenèse du sac vitellin [133]. Par la suite, l'injection de cellules souches embryonnaires *ETV6*^{-/-} dans des blastocytes de souris a permis de créer des chimères viables permettant l'étude du rôle d'*ETV6* au-delà de ce stade de développement. Ces cellules *ETV6*^{-/-} n'ont pu générer les lignées hématopoïétiques (myéloïdes et lymphoïdes) normalement retrouvées dans la moelle osseuse, indiquant le rôle d'*ETV6* dans la mise en place et le maintien de l'hématopoïèse dans ce compartiment [86]. Finalement, Hock H. *et al.* ont pu étudier la fonction d'*ETV6* chez la souris adulte par des *knock-out* conditionnels utilisant le système Cre-Lox restreint à certaines lignées hématopoïétiques précises [134]. Bien que l'ablation d'*ETV6* dans les lignées commises ne semble pas avoir d'effet notable hormis sur la maturation des mégacaryocytes, ils ont confirmé l'importance d'*ETV6* pour la viabilité des cellules souches hématopoïétiques de la moelle osseuse.

1.4.4.2. *ETV6* comme suppresseur de tumeur

En plus de sa fonction de régulateur de l'hématopoïèse, des études fonctionnelles dans des lignées cellulaires ont proposé qu'*ETV6* pouvait agir comme suppresseur de tumeur. En effet, la surexpression d'*ETV6* prévient les

effets pro-prolifératifs et invasifs de l'expression de l'oncogène Ras [135, 136], diminue la prolifération induite par l'interleukine-3 (*IL3*) [137], augmente l'apoptose [137, 138] et facilite la différenciation en érythrocytes par rapport aux mégacaryocytes [139, 140]. Ces études, combinées aux altérations d'*ETV6* observées dans plusieurs désordres hématologiques (voir section suivante 1.4.5), supportent un rôle de suppresseur de tumeur.

1.4.5. Altérations d'*ETV6* dans les désordres hématologiques

ETV6 est exprimé dans la majorité des tissus et lignées cellulaires étudiés [141, 142], mais les altérations génétiques sont retrouvées préférentiellement dans les désordres hématologiques. Bien que quelques mutations ponctuelles ont été identifiées [143], la grande majorité de ces altérations sont des réarrangements chromosomiques.

1.4.5.1. Réarrangements chromosomiques d'*ETV6*

En plus de la fusion *ETV6*-*RUNX1* fréquemment retrouvée dans la B-LAL (voir section 1.3.3.3.3), plusieurs autres translocations impliquant le gène *ETV6* ont été identifiées et pas moins de 30 partenaires de fusion ont été caractérisés (**Figure XII**) [132]. Ces partenaires peuvent être classés en 5 catégories c.-à-d. les récepteurs tyrosine kinases, les tyrosine kinases (non-récepteurs), les facteurs de transcription, les gènes *homeobox* et les autres.

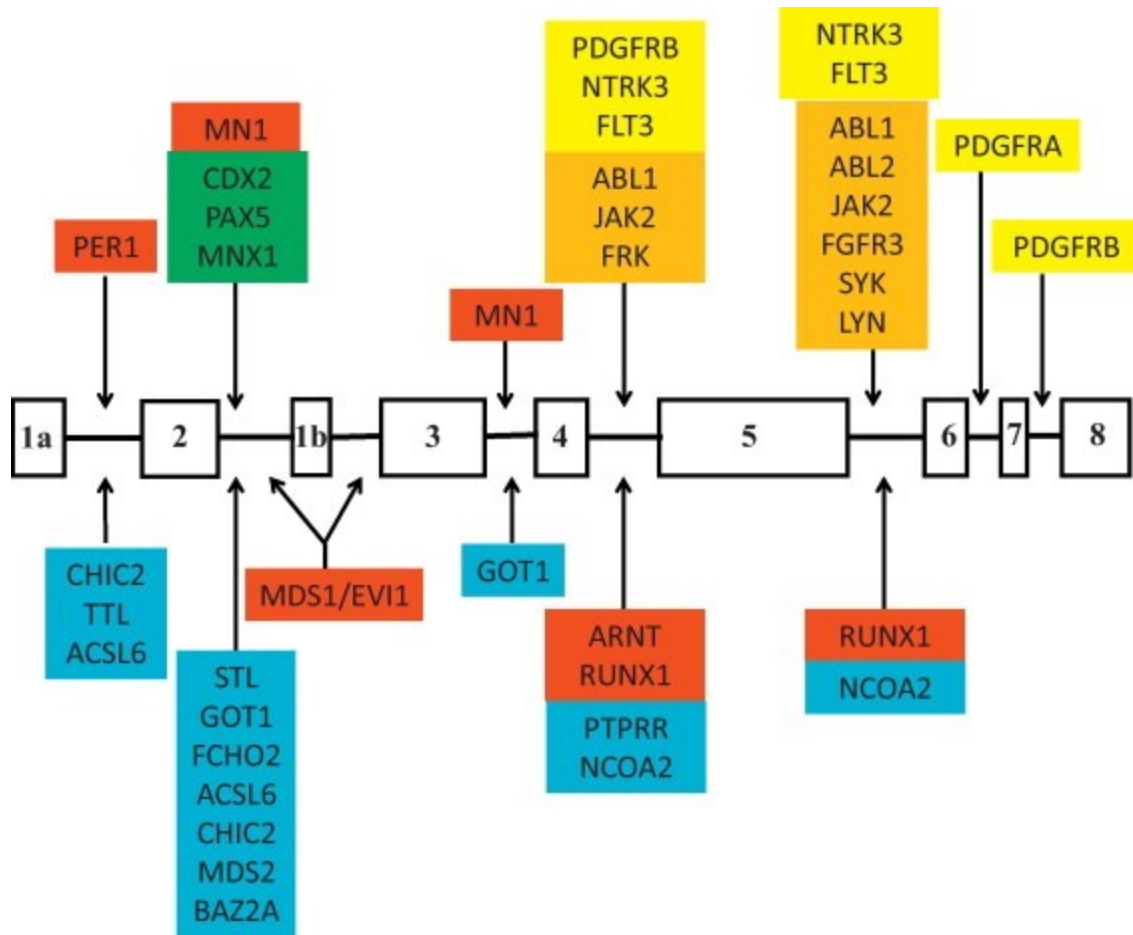


Figure XII: Points de cassure des translocations impliquant le gène *ETV6* avec ses divers partenaires de fusion. En jaune, récepteurs tyrosine kinases; en orange, tyrosine kinases; en rouge, facteurs de transcription; en vert, gènes homeobox et finalement en bleu tous les autres n'appartenant à aucune des catégories sus mentionnées. Reproduite de E. De Braekeleer *et al.*, 2012 [132].

Chacune des fusions agit selon 5 mécanismes distincts [132]. **1.** Tout comme la fusion BCR-ABL1 (voir section 1.3.3.2), les fusions entre *ETV6* et les tyrosine kinases (récepteurs ou non) engendrent l'activation constitutive de la fonction kinase des partenaires et enclenchent la suractivation de voies de signalisation oncogéniques. La fusion t(5;12)(q33;p13) entre *ETV6* et le récepteur tyrosine

kinase *PDGFRB* observée dans certains cas de leucémie myélomonocytaire chronique correspond à ce mécanisme. Il est à noter que l'identification du gène *ETV6* a d'ailleurs été faite dans le cadre de la caractérisation de cette fusion [141]. **2.** La fusion d'*ETV6* avec un autre facteur de transcription peut créer une protéine de fusion ayant des propriétés transcriptionnelles uniques et déréguler l'expression génique. La fusion *ETV6-RUNX1* dans la B-LAL et les rares cas de fusions *MN1-ETV6* dans la leucémie myéloïde aiguë partagent ce mécanisme d'action. **3.** Ces translocations avec d'autres facteurs de transcription peuvent également avoir un effet dominant négatif sur *ETV6*. Il est postulé que *ETV6-RUNX1* empêcherait la régulation médiée par la forme sauvage d'*ETV6* encodée par l'allèle résiduel (non réarrangé) par leur association via leur domaine PNT [98]. Le produit de fusion de la translocation t(1;12)(q21;p13) impliquant *ETV6* et le gène *ARNT* pourrait également interagir avec la forme native d'*ETV6* via le domaine PNT conservé dans la protéine chimère et inhiber sa fonction [132, 144, 145]. Les derniers cas de figure sont beaucoup plus rares et impliquent préférentiellement des gènes partenaires ayant des fonctions autres (ni tyrosine kinase ni facteur de transcription). **4.** Les translocations avec *ETV6* peuvent aussi être délétères et avoir comme impact une perte de fonction d'*ETV6* ou du partenaire de fusion. **5.** Elles peuvent finalement activer un proto-oncogène à proximité du point de cassure [132].

1.4.5.2. Mutations germinales d'*ETV6*

Des études récentes ont identifié des mutations germinales (transmises; héritées; *germline*) au niveau d'*ETV6* associées à un historique familial de

maladies hématologiques [25-28]. Au sein d'une même famille ayant une forte incidence de ce genre de maladies (p. ex. thrombocytopénie, syndrome myélodysplasique et LAL), ces mutations ségrèguent avec les individus atteints, suggérant un effet causal de celles-ci dans ces familles. Il est à noter que ces mutations sont retrouvées dans le domaine ETS de liaison à l'ADN d'*ETV6*, perturbant ainsi sa localisation nucléaire et son activité transcriptionnelle. L'identification de telles mutations renforce la notion qu'*ETV6* participerait à l'hématopoïèse et que son inhibition (réarrangements, mutations, délétions) contribuerait au développement de ces maladies comme la LAL.

1.4.6. La régulation médiée par *ETV6*

1.4.6.1. Mécanismes de répression transcriptionnelle

La répression de la transcription médiée par *ETV6* emprunte plusieurs mécanismes faisant intervenir des partenaires différents. Comme les autres facteurs ETS, *ETV6* reconnaît et lie les séquences EBS sur l'ADN via son domaine ETS. L'homodimérisation d'*ETV6* lui permet de lier avec une forte affinité des séquences EBS en tandem [130]. Une fois localisé sur l'ADN, *ETV6* recrute par l'entremise de son domaine central des corépresseurs classiques, notamment N-CoR, mSin3 et SMRT [146, 147], qui à leur tour peuvent recruter des histones déacétylases (HDACs). *HDAC3* et *HDAC9* ont d'ailleurs été identifiés dans les complexes de répression d'*ETV6* par co-immunoprécipitation [148, 149]. La présence des HDACs aux gènes ciblés par *ETV6* permet une répression active de l'expression génique. Il a d'ailleurs été démontré que l'inhibition chimique des HDACs par la trichostatin A (TSA) permet de prévenir la

répression médiée par le domaine central d'*ETV6* [147, 148]. Autrement, *ETV6* peut recruter le corépresseur du groupe *polycomb* *L3MBTL1* via son domaine PNT [150]. Inversement à la répression médiée par le domaine central d'*ETV6*, la répression par le domaine PNT et *L3MBTL1* ne fait pas intervenir les HDACs et est associée à une répression à long terme maintenue par la compaction de la chromatine.

Bien que la fonction répressive d'*ETV6* soit classiquement initiée par sa liaison directe à l'ADN, il peut également agir comme corépresseur. Par exemple, *ETV6* est un répresseur de la fonction transcriptionnelle du facteur de transcription *STAT3*. Le domaine PNT d'*ETV6* permet sa liaison avec *STAT3* et le recrutement subséquent de corépresseurs et HDACs aux cibles transcriptionnelles de *STAT3*. En ce sens, l'inhibition des HDACs par le TSA prévient la fonction répressive d'*ETV6* à l'égard de *STAT3*. De plus, la délétion du domaine ETS de liaison à l'ADN d'*ETV6* est sans conséquence sur son activité sur *STAT3*, suggérant qu'*ETV6* agit non pas en compétitionnant la liaison à l'ADN de *STAT3* mais plutôt par le recrutement d'un complexe répresseur [151]. *ETV6* agirait de façon similaire sur le facteur de transcription *ICSBP* (*IRF8*). Ce facteur est recruté à certains promoteurs (*IFN-stimulated response element*; ISRE) lorsque stimulé par l'interféron gamma ($IFN\gamma$). Or il a été démontré que la stimulation par l' $IFN\gamma$ induisait l'interaction entre *ICSBP* et *ETV6*. La présence d'*ETV6* aux promoteurs ISRE due à son interaction avec *ICSBP* promeut le recrutement de l'histone déacétylase *HDAC3* et l'inhibition de la transcription [152]. Ainsi, le rôle d'*ETV6* comme corépresseur est similaire à

celui tenu au sein de la protéine de fusion ETV6-RUNX1, soit le recrutement d'un complexe répresseur au niveau des gènes cibles d'un autre facteur de transcription.

1.4.6.2. Les cibles transcriptionnelles d'*ETV6*

Comme *ETV6* peut agir sur la transcription en se liant à l'ADN directement via son domaine ETS ou bien par l'entremise d'autres facteurs (p. ex. *STAT3* et *ICSBP*), son réseau de régulation est complexe. Les gènes régulés par *ETV6* peuvent être directs (c.-à-d. directement liés et réprimés) ou indirects (c.-à-d. modulés par des interactions avec d'autres facteurs).

L'expression du gène *stromelysin-1* (*MMP3*) a été montrée comme étant réprimée par *ETV6* et serait responsable du phénotype d'*ETV6* sur l'inhibition des effets invasifs de l'expression de l'oncogène Ras [135]. Des expériences de type EMSA (*electrophoretic mobility shift assay*) ont pu confirmer le potentiel de liaison d'*ETV6* à certaines séquences du promoteur proximal du gène, suggérant une cible directe. L'expression du gène anti-apoptotique *Bcl-X_L* a également été négativement corrélée à celle d'*ETV6* et expliquerait, au moins en partie, l'effet d'*ETV6* sur l'apoptose [138]. Des expériences d'immunoprécipitation de la chromatine (*ChIP*) ont permis d'associer cette régulation à la liaison d'*ETV6* au niveau de l'un des promoteurs responsables de l'expression de *Bcl-X_L*, suggérant une seconde cible directe. Il est à noter que ces deux études [135, 138] ont fait l'utilisation d'essais de gène rapporteur (luciférase) et n'ont pu observer la répression de ces promoteurs par une forme mutante d'*ETV6* sans son domaine ETS de liaison à l'ADN (Δ ETS). Bien que ceci puisse être interprété comme une

évidence additionnelle qu'*ETV6* doit lier directement ces promoteurs pour en réprimer l'expression, l'ablation du domaine ETS perturbe la localisation nucléaire d'*ETV6* [129] et par conséquent compromet une telle conclusion.

En plus des gènes ciblés par les facteurs *STAT3* (p. ex. α 2-macroglobulin; *A2M* [151]) et *ICSBP* (p. ex. 2',5'-oligoadenylate synthetase; *OAS1* [152]) qui sont réprimés indirectement par *ETV6* (voir section 1.4.6.1), l'expression d'*ETV6* corrèle également négativement avec les niveaux d'expression de certains gènes associés aux mégacaryocytes [140]. *Fli-1*, un autre membre de la famille ETS (**Figure X**), est directement impliqué dans la maturation des mégacaryocytes en activant la transcription des gènes *GPIX* (*GP9*), *GPIb α* (*GP1BA*) et *GPIIb* (*ITGA2B*) [153]. Or, bien que l'expression d'*ETV6* seul n'ait aucun effet sur l'activité transcriptionnelle de ces promoteurs, sa co-expression empêche leur activation par *Fli-1* en interagissant via leur domaine PNT [154]. De plus, comme aucune interaction entre ces promoteurs et *ETV6* n'a pu être démontrée, ces gènes sont probablement des cibles indirectes. De façon similaire, *ETV6* prévient l'activation du gène *MCSFR* (*CSF1R*) par les facteurs *CBFA2* (*RUNX1*) et *C/EBP α* (*CEBPA*) [155]. Finalement, l'expression de *TCL1* a été démontrée comme étant inversement proportionnelle à celle d'*ETV6*, mais aucun mécanisme n'a encore été proposé [156].

1.4.7. La régulation d'*ETV6*

1.4.7.1. Régulation transcriptionnelle d'*ETV6*

Comme *ETV6* est majoritairement étudié dans le contexte d'une perte de fonction (p. ex. les multiples réarrangements, délétions et mutations), peu d'études se sont intéressées aux facteurs impliqués dans sa transcription. *ETV6* est exprimé de façon ubiquitaire [141, 142] par un promoteur qui ne contient pas d'élément TATA [123]. Le complexe activateur de la transcription GABP (*GABP binding protein*), un dimère composé du facteur ETS GABP α et de la sous-unité activatrice GABP β , est responsable de la maturation de la lignée lymphoïde par la régulation du facteur de transcription *PAX5* [157] et de la chaîne α du récepteur à l'interleukine-7 (*IL7R*) [158]. La déficience en GABP chez la souris engendre d'ailleurs des phénotypes similaires à ceux observés en l'absence d'*ETV6* [134, 159]. En ce sens, la fonction de GABP sur le maintien et la différenciation des cellules souches hématopoïétiques a pu être associée à l'activation des facteurs de transcription *Zfx* et *ETV6* [159]. Les expériences de ChIP combinées aux analyses de données d'expression indiquent qu'*ETV6* est une cible directe du facteur GABP. Une autre étude chez la souris a pu identifier *ETV6* comme étant sous-exprimé suite à une déficience en *PAX6* [160]. Bien qu'aucun mécanisme n'ait été proposé, cela suggère que *PAX6* soit impliqué dans son activation transcriptionnelle. *ETV6*, en plus d'être impliqué dans la modulation de l'activité transcriptionnelle de *STAT3*, est aussi transcriptionnellement activé par ce dernier, créant ainsi une boucle d'autorégulation négative [151]. La régulation transcriptionnelle d'*ETV6* par

STAT3 n'a cependant pas été étudiée plus en détail et pourrait être médiée par le biais d'interactions indirectes.

1.4.7.2. Régulation post-traductionnelle d'*ETV6*

Des modifications post-traductionnelles peuvent également moduler l'activité d'*ETV6*. À cet égard, la sumoylation d'*ETV6* sur la lysine 99 (K99) a été bien caractérisée. L'interaction d'*ETV6* avec UBC9 (*UBE2I*) [161] permet de lier de façon covalente la protéine SUMO-1 à cette lysine [162]. La liaison de SUMO-1 ne mène pas à la dégradation d'*ETV6* [161], mais plutôt à son exportation du noyau vers le cytoplasme [163]. La mutation du résidu K99 d'*ETV6* pour une arginine (K99R) favorise d'ailleurs sa localisation nucléaire et conséquemment son activité de répresseur transcriptionnel. La sumoylation d'*ETV6* à la position K99 par *UBE2I* diminue donc son activité suite à sa délocalisation. Une seconde lysine (K11) a également été démontrée comme étant la cible de sumoylation [126]. Dans ce cas, l'ajout de SUMO-1 et SUMO-2 sur le résidu K11 d'*ETV6* se fait par son interaction avec la protéine *PIAS3*. Contrairement à la sumoylation K99, la sumoylation K11 n'induit pas de modification au niveau de sa localisation mais réduit sa capacité à lier l'ADN. Curieusement, l'inhibition de la liaison à l'ADN n'est pas médiée par une incapacité d'*ETV6* à dimériser comme le laisse supposer une sumoylation à proximité du domaine PNT. Bien que le mécanisme reste incomplet, il n'en demeure pas moins que cette diminution d'affinité pour l'ADN se traduit par une diminution de l'activité répressive d'*ETV6*. Il est à noter que l'inhibition de la fonction d'*ETV6* par la sumoylation du résidu K11 n'affecte que la forme complète (*ETV6*-M1) et non la forme tronquée (*ETV6*-M43; voir

section 1.4.3). Les conséquences et implications d'un tel mécanisme de régulation différentiel n'ont cependant toujours pas été clairement établies.

En plus de la sumoylation, *ETV6* peut être la cible de phosphorylations [142]. Outre la sérine 22 (S22) qui semble phosphorylée de manière constitutive, *ETV6* peut interagir directement avec la kinase p38 (*MAPK14*) qui le phosphoryle au niveau de la sérine 257 (S257) [164]. Des essais de gènes rapporteurs ont démontré que cette phosphorylation inhibait la répression médiée par *ETV6*. Finalement, l'activité kinase de v-src inhibe l'activité répressive d'*ETV6* en induisant son exportation nucléaire [125] de façon similaire à ce qui a été montré pour la sumoylation du résidu K99 par *UBE2I*. Bien que cette délocalisation ait pu être associée aux 52 premiers acides aminés d'*ETV6*, aucun des résidus potentiellement phosphorylables de cette région ne semble être la cible de l'activité kinase de v-src, suggérant un mécanisme plus indirect. Comme cette régulation par v-src nécessite les 52 premiers acides aminés d'*ETV6*, elle n'affecte que l'isoforme long (*ETV6-M1*).

1.5 Rationnelle et hypothèses de recherche

La B-LAL est le cancer le plus fréquemment diagnostiqué chez les enfants de moins de 14 ans. L'anomalie génétique la plus fréquente est la translocation $t(12;21)(p13;q22)$ qui implique les facteurs de transcription *ETV6* et *RUNX1*. Plusieurs études se sont penchées sur le rôle de la chimère *ETV6-RUNX1* dans l'initiation et le développement de cette leucémie. Or, il a été démontré que la présence de cette fusion était plutôt associée à un état préleucémique qui devait être complété par d'autres événements délétères pour mener au phénotype leucémique (modèle en deux étapes).

Nous postulons qu'un événement clé dans la transition préleucémique à leucémique serait l'inactivation du facteur de transcription *ETV6*. Malgré l'abondance d'indices convergents vers son implication dans la LAL, peu d'études se sont intéressées à *ETV6* au niveau moléculaire. Une meilleure compréhension du réseau de régulation d'*ETV6*, autant en amont qu'en aval, permettrait de mieux comprendre l'impact fonctionnel de sa délétion dans la B-LAL.

1.6 But principal

L'inactivation d'un facteur de transcription perturbe l'expression d'un bon nombre de gènes pouvant être impliqués dans une variété de processus biologiques. Pour *ETV6*, les facteurs régulant son activité transcriptionnelle, l'identité de ses gènes cibles et leurs implications sont encore méconnus. Pour être en mesure de mieux comprendre l'impact de la perte d'*ETV6* dans la B-LAL, nous devons améliorer notre compréhension de son réseau de régulation.

Plus spécifiquement, les objectifs de cette étude visaient à:

1. Identifier de manière exhaustive les cibles transcriptionnelles d'*ETV6*.
2. Évaluer l'impact fonctionnel de ses gènes cibles dans la leucémogénèse.
3. Déterminer son profil de liaison à l'ADN.
4. Identifier les modulateurs de sa fonction transcriptionnelle.

Chapitre 2: Premier article

CLIC5 dans la B-LAL

2.1 Préface

Nous proposons que la dérégulation des cibles transcriptionnelles d'*ETV6* a un impact important dans la leucémogénèse. Par contre, seulement quelques cibles d'*ETV6* ont été identifiées par des approches de gènes candidats dans des modèles souvent peu représentatifs du système hématopoïétique. Les répercussions d'*ETV6* sur le profil d'expression de cellules leucémiques n'ont donc jamais été interrogées dans son intégralité.

Afin d'identifier les cibles transcriptionnelles d'*ETV6*, nous avons fait l'analyse différentielle de transcriptomes de cellules leucémiques où *ETV6* avait été réintroduit. Cette analyse a permis l'identification de gènes dont l'expression corrélait négativement avec celle d'*ETV6*, suggérant des cibles transcriptionnelles. Parmi ces gènes, plusieurs étaient communs à la signature particulière des patients B-LAL t(12;21), démontrant l'effet de l'inactivation d'*ETV6* sur le profil d'expression de ces patients. *CLIC5* apparaissait comme la meilleure cible d'*ETV6* et a été étudiée plus en détail au niveau de sa fonction moléculaire et du rôle que pouvait avoir sa surexpression dans la leucémogénèse. Des études fonctionnelles ont permis de lier *CLIC5* à une résistance au stress oxydatif. Ce nouveau rôle de *CLIC5* permet de mieux définir le phénotype associé à la perte de fonction d'*ETV6* dans la B-LAL.

CLIC5: a novel *ETV6* target gene in childhood acute lymphoblastic leukemia

Benjamin Neveu^{1,2}, Jean-François Spinella^{1,3}, Chantal Richer¹, Karine Lagacé^{1,2}, Pauline Cassart¹, Mathieu Lajoie¹, Silvana Jananji¹, Simon Drouin¹, Jasmine Healy¹, Gilles R.X. Hickson^{1,4} and Daniel Sinnett^{1,2,5}

¹ CHU Sainte-Justine Research Center, Montreal, Canada

² Department of Biochemistry and Molecular Medicine, Faculty of Medicine, University of Montreal, Montreal, Canada

³ Molecular biology program, Faculty of Medicine, University of Montreal, Montreal, Canada

⁴ Department of Pathology and Cellular Biology, Faculty of Medicine, University of Montreal, Montreal, Canada

⁵ Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Canada

Corresponding author: Daniel Sinnett, Division of Hematology-Oncology, Sainte-Justine UHC Research Center, 3175 Côte Ste-Catherine, Montréal (Québec) Canada, H3T 1C5; Tel: (514) 345-4931/ext. 2990; Email: daniel.sinnett@umontreal.ca

Cet article est publié dans Haematologica. 2016 Dec;101(12):1534-1543.

2.2 Author contributions

DS is the principal investigator and takes primary responsibility for the paper. BN and DS conceived and designed the study and BN performed most of the experiments. BN, JFS, CR and DS contributed to the interpretation of the data and BN wrote the paper. JFS performed transcriptome analyses. CR contributed to the study design and performed sequencing library preparation. KL contributed to the chromatin immunoprecipitation experiments and PC contributed to the apoptosis assays. ML was involved in cell line RNA-seq analyses. SJ and GH performed immunofluorescence experiments. JFS, SD, JH, GH and DS were involved in critical manuscript revision. All authors approved the final version of the manuscript.

2.3 Abstract

The most common rearrangement in childhood precursor B-cell acute lymphoblastic leukemia is the t(12;21)(p13;q22) translocation resulting in the ETV6-AML1 fusion gene. A frequent concomitant event is the loss of the residual *ETV6* allele suggesting a critical role for the *ETV6* transcriptional repressor in the etiology of this cancer. However, the precise mechanism through which loss of functional *ETV6* contributes to disease pathogenesis is still unclear. To investigate the impact of *ETV6* loss on the transcriptional network and to identify new transcriptional targets of *ETV6*, we used whole transcriptome analysis of both pre-B leukemic cell lines and patients combined with chromatin immunoprecipitation. Using this integrative approach, we identified 4 novel direct *ETV6* target genes: *CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L*. To further evaluate the role of chloride intracellular channel protein *CLIC5* in leukemogenesis, we generated cell lines overexpressing *CLIC5* and demonstrated an increased resistance to hydrogen peroxide-induced apoptosis. We further described the implications of *CLIC5*'s ion channel activity in lysosomal-mediated cell death, possibly by modulating the function of transferrin receptor with which it co-localizes intracellularly. For the first time, we showed that loss of *ETV6* leads to significant overexpression of *CLIC5*, which in turn leads to decreased lysosome-mediated apoptosis. Our data suggest that heightened *CLIC5* activity could promote a permissive environment for oxidative-stress induced DNA damage accumulation and thereby contribute to leukemogenesis.

2.4 Introduction

ETV6 is a known transcriptional repressor [1] involved in hematopoiesis [2]. *ETV6* rearrangements are frequently observed in multiple hematological diseases including precursor B-cell acute lymphoblastic leukemia (pre-B ALL), the most common pediatric cancer [3]. In fact, the t(12;21)(p13;q22) translocation which generates an in-frame *ETV6*-*AML1* fusion product [4], is the most frequent chromosomal abnormality in childhood pre-B ALL present in 20% of cases [5]. The expression of *ETV6*-*AML1* is systematically observed in t(12;21)-positive pre-B ALL [6, 7], indicating a possible function for this chimeric protein in pre-B ALL etiology. However, it was shown that the frequency of the t(12;21) translocation is 100 times greater than that of pre-B ALL [8], suggesting that its presence alone is insufficient to induce leukemia. It has been demonstrated that the second non-rearranged *ETV6* allele is frequently deleted or inactivated in t(12;21)-positive pre-B ALL [7, 9-11] and recent studies have reported germline *ETV6* loss-of-function mutations that were shown to be associated with familial hematological disorders including ALL [12]. Although these data suggest that *ETV6* plays a key tumor suppressor role and that its complete inactivation may be required for leukemogenesis [13], little is known about the function of *ETV6* in normal hematopoiesis and leukemic transformation.

Given its role in transcriptional repression, we postulated that loss of *ETV6* could result in deregulated expression of downstream target genes and perturb key cellular processes and pathways leading to oncogenesis. Only two transcriptional targets of *ETV6* have been identified to date: the *MMP3* matrix metalloproteinase

[14] and the anti-apoptotic protein *Bcl-X_L* [15]. To comprehensively identify novel *ETV6* target genes, we combined whole transcriptome analysis and chromatin immunoprecipitation (ChIP) assays. Using an in vitro cell-based system combined with data from childhood pre-B ALL patient tumors, we identified 4 genes (*CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L*) whose expression was directly regulated by *ETV6*. Functional interrogation of *CLIC5* revealed its implication in lysosome-mediated cell death, possibly by regulating iron homeostasis through the transferrin receptor with which it co-localizes intracellularly. In this study, we provide the first evidence of a role for *CLIC5*-mediated resistance to oxidative stress that may contribute to leukemogenesis.

2.5 Methods

Complete methods can be found in the supplemental methods section

2.5.1. Expression profiling by RNA-sequencing

Total RNA from two different Reh clones (generated in methylcellulose media) each stably expressing ETV6-His and ETV6 Δ ETS_NLS-His (and pLENTI control) was processed through the TruSeq Stranded Total RNA protocol and sequenced on the HiSeq 2500 system (Illumina). Reads for each samples were mapped to the hg19 reference genome using STAR with default settings [16] and read counts per genes were determined using HTSeq-count [17]. To identify differentially expressed genes (DEGs), we used the R bioconductor package edgeR [18] with Benjamini Hochberg p-value adjustment. The two clones were considered as biological replicates.

The patient cohort used for RNA-sequencing was composed of 9 hyperdiploid and 9 t(12;21) patients. Total RNA was extracted from leukemic bone marrow samples of all patients and from control pre-B cells (CD19⁺CD10⁺) isolated from healthy cord blood samples. cDNA libraries were prepared using the SOLiD Total RNA-seq kit and sequenced on the SOLiD 4/5500 System (Life Technologies). Reads were aligned to the hg19 reference genome and read counts per gene obtained using LifeScope Genomic Analysis Software with default parameters.

2.5.2. Quantitative real-time PCR

350ng of total RNA were retro-transcribed with M-MLV reverse transcriptase (Life Technologies). cDNA was then subjected to quantitative real-time PCR using the primer sets listed in **Table S1**. Relative expression was determined by the $2^{-(\Delta\Delta Ct)}$ comparative method [19] using *GAPDH* as the reference gene.

2.5.3. Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed on 10×10^6 transduced Reh cells cross-linked directly in cell medium for 10min with 1% methanol-free formaldehyde (Polysciences). Immunoprecipitation of sheared chromatin was carried out using anti-HA magnetic beads (Thermo Fisher Scientific). Beads were eluted twice with HA peptides (Thermo Fisher Scientific) before reverse-crosslinking. DNA was purified twice by standard phenol/chloroform/isoamyl alcohol (Sigma) extraction prior to qRT-PCR analysis (primers are listed in **Table S2**).

2.5.4. Apoptosis assays

Apoptosis was induced by treating cells for 20h with hydrogen peroxide (PRXD; Sigma), camptothecin (CPT; Tocris Bioscience) or doxorubicin (DOXO; Sigma) and assayed by Alexa Fluor 488-coupled Annexin V and propidium iodide (PI) double staining. 1×10^4 stained cells were analysed by flow cytometry. Total apoptosis includes Annexin V+/PI - (early apoptotic), Annexin V+/PI + (late apoptotic) and Annexin V-/PI+ (necrotic) cells.

2.5.5. Immunofluorescence microscopy

Reh cells were seeded at 2×10^6 cells/mL in a 96-well glass plate (Whatman) and fixed in a 3.7% formaldehyde solution. Immunostaining was performed overnight with CLIC5A antibody (ab191102 dil. 1:1000; Abcam) and transferrin receptor antibody (ab84036 dil. 1:200; Abcam). Goat anti-mouse Alexa Fluor 488 (dil. 1:500; Thermo Fisher Scientific) was used to detect anti-CLIC5A and Goat anti-rabbit Alexa Fluor 546 (dil. 1:500; Thermo Fisher Scientific) was used to detect anti-transferrin receptor. Hoechst 33258 DNA stain (dil. 1:500; Thermo Fisher Scientific) was included to stain nuclei.

2.5.6. Statistical tests

Significance of observations was assessed using one or two-tailed Fisher's exact test or Mann-Whitney-U test when appropriate.

2.5.7. Ethics statement

The CHU Sainte-Justine Research Ethics Board approved the protocol. Informed consent was obtained from the parents of the patients to participate in this study and for publication of this report and any accompanying images.

2.6 Results

2.6.1. *ETV6* represses the expression of 6 genes in pre-B ALL cell lines and patient samples

To identify novel direct *ETV6*-regulated genes, we carried out a transcriptome analysis in both cell lines and patient tumor samples (**Figure 1**). Based on expression profiles of transduced t(12;21)-positive pre-B ALL Reh clones (**Figure S1**), we found 331 genes repressed in His-tagged *ETV6* (*ETV6*-His) cells compared to control cells (pLENTI empty vector; p-value ≤ 0.05), of which 88 remained significant after multiple testing corrections (FDR ≤ 0.1 ; **Table S3**). 18 genes were significantly repressed by *ETV6*-His compared to its DNA-binding deficient mutant *ETV6* Δ ETS_NLS-His (FDR ≤ 0.1 ; **Table S4**), of which 11 were both confidently expressed (logCPM ≥ 1) and also present in the above-mentioned list of 88 genes. These genes are thus more likely to be direct targets of *ETV6* since their repression depends on *ETV6*'s DNA-binding domain. However only 7 of these genes absolutely required the DNA-binding domain for repression (*ETV6* Δ ETS_NLS-His vs. pLENTI; p-value ≥ 0.05 or logFC ≥ -0.5), further confirming their direct regulation by *ETV6*: *CLIC5*, *BIRC7*, *DDIT4L*, *ANGPTL2*, *WBP1L*, *LRRC4*, and *SLC51A*.

We then assessed whether the *ETV6*-dependent transcriptional repression observed *in vitro* translated to childhood pre-B ALL patient tumor samples. Expression of the 331 genes repressed by *ETV6* *in vitro* was evaluated in transcriptome data from 9 t(12;21)-positive samples (*ETV6* negative) and compared to 9 hyperdiploid cases as well as to 3 normal pre-B cell

(CD19⁺/CD10⁺) samples (*ETV6* positive). We identified 45 genes that were downregulated *in vitro* (13.6%) and that were also specifically overexpressed in t(12;21)-positive patients (**Figure 2**). Interestingly, these include 6 of the 7 genes identified as putative direct *ETV6* targets *in vitro* (*CLIC5*, *BIRC7*, *ANGPTL2*, *WBP1L*, *LRRC4* and *SLC51A*, but not *DDIT4L*), further supporting a role for *ETV6* in their regulation.

2.6.2. *CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L* are direct targets of *ETV6*

To validate *ETV6*-dependent expression of these 6 genes (**Table 1**), we used quantitative real-time PCR (qRT-PCR) in both Reh clones and the original Reh population overexpressing *ETV6* WT, *ETV6*-HA or GFP as a control (**Figure S2**). Both *ETV6* WT and *ETV6*-HA efficiently repressed expression of these genes except for *LRRC4* (**Figure 3A**). To assess the physical interaction between *ETV6* and the proximal promoters of these 5 putative *ETV6* targets we performed ChIP experiments in Reh cells overexpressing *ETV6*-HA or *ETV6* as a negative control. Importantly, *ETV6*-HA behaves similarly to *ETV6* WT in our qRT-PCR experiments (**Figure 3A**), indicating that the epitope tag does not negatively interfere with normal *ETV6* repressor function. As shown in **Figure 3B**, we successfully enriched the proximal promoters of *CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L*, but not *SLC51A*, further confirming that these 4 genes are indeed direct targets of *ETV6*.

2.6.3. *CLIC5A* reduces hydrogen peroxide-induced apoptosis

We pursued functional interrogation of our strongest candidate, the chloride intracellular channel *CLIC5*, to investigate its cellular function and potential contribution to childhood pre-B ALL. The *CLIC5* locus encodes two major isoforms, *CLIC5A* and *CLIC5B* [20] (**Figure S3A**), transcribed by two alternative promoters and differing only from their first exon. Using ChIP experiments (as above), we showed that the *CLIC5A* promoter was specifically enriched whereas the *CLIC5B* promoter showed no significant enrichment compared to the negative control region (**Figure S3B**). *ETV6* overexpression was also shown to lead to a marked decrease of the *CLIC5A* protein whereas *CLIC5B* levels remained constant (**Figure S3C**). Together, these results confirm specific *ETV6*-mediated repression of *CLIC5A*.

In light of these results, we overexpressed the *CLIC5A* isoform in Reh cells (**Figure 4A**) to investigate its role on B-lymphoblast function. Importantly, the overexpression of *CLIC5A* in our Reh cells is highly similar to what is observed in a validation cohort of t(12;21) ALL patients (**Figure S4**). Given that changes in migration were observed upon silencing of *CLIC5A* [21], we first evaluated this phenotype. We found no particular difference in migration toward stromal cell-derived factor 1 (SDF-1, also known as *CXCL12*) in a classic trans-well experiment, between control and *CLIC5A* overexpressing cells (**Figure S5**). Although *CLIC5A* had never been shown to be associated with apoptosis, suppression of its closely related family member *CLIC4* had previously been shown to enhance hydrogen peroxide-induced apoptosis [22]. To test this

hypothesis in our cell model, we treated *CLIC5A* overexpressing cells with hydrogen peroxide, camptothecin, or doxorubicin, and evaluated apoptosis. Of note, peroxide induces an apoptotic cell death in these conditions (**Figure 4B**) rather than necrosis. We observed a modest yet consistent reduction in hydrogen peroxide-induced apoptosis compared to control cells (**Figure 4C**), suggesting a potential role in the intracellular response to free radicals. A similar reduction in apoptosis following peroxide treatment was observed with *CLIC5A* overexpression in the IM9 B-lymphoblastoid cell line (**Figure 4 D-F**) endogenously expressing wild-type *ETV6*, further confirming that *CLIC5A* overexpression specifically reduces hydrogen peroxide-induced apoptosis across cellular backgrounds.

2.6.4. *CLIC5A* is an endosomal ionic channel involved in lysosome-mediated apoptosis

Given that hydrogen peroxide is known to trigger lysosomal membrane permeabilization (LMP) and initiate the lysosomal-mediated apoptosis pathway (**Figure 5A**) [23, 24], we hypothesized that *CLIC5A*'s role in protecting cells against apoptosis may function through the modulation of LMP. Since LMP has a direct impact on mitochondrial outer membrane permeabilization (MOMP) that can be assessed through mitochondrial membrane potential (MMP), we evaluated MMP in our Reh cellular model treated with hydrogen peroxide. We observed a significant reduction of MMP loss correlated with *CLIC5A* overexpression compared to control indicating that substantially more mitochondria remained intact following peroxide exposure when *CLIC5A* was

overexpressed (**Figure S6**). This result supports a role for *CLIC5A* in protecting cells against peroxide-induced apoptosis and suggests that it functions upstream of MOMP, which indeed corroborates a possible implication of *CLIC5A* in LMP regulation.

Deleterious effects of hydrogen peroxide on lysosome membranes are strongly dependent of lysosomal Fe^{2+} availability since it dictates its conversion to the highly reactive hydroxyl radicals [25]. We thus investigated *CLIC5A*'s impact on lysosome-mediated cell death by modulating lysosomal Fe^{2+} availability prior to hydrogen peroxide exposure by pre-treating cells with the iron chelator deferoxamine mesylate salt (DFO) [26, 27]. As shown in **Figure 5B**, cells treated with DFO prior to peroxide showed a drastic reduction in apoptosis. The residual apoptotic activity observed in DFO-treated cells could be driven by DNA damage [28] which appears to be *CLIC5A*-independent given the results obtained with the two DNA damaging agents camptothecin and doxorubicin (**Figure 4C**).

Inversely, we used ferric ammonium citrate (FAC) to positively modulate Fe^{2+} concentration in lysosomes [29]. Increased lysosomal Fe^{2+} concentration favors hydroxyl radical production from hydrogen peroxide and hence is expected to increase LMP and apoptosis. Accordingly, cells pre-treated with FAC showed increased peroxide-induced apoptosis and the protective effect of *CLIC5A* overexpression was completely lost (**Figure 5C**), indicating that *CLIC5A* plays a role upstream of LMP. This specific role further corroborates *CLIC5A*'s inability to protect cells against DNA damage.

To evaluate *CLIC5A*'s ion channel activity [30] in this context we pre-treated *CLIC5A* overexpressing Reh cells with indanyloxyacetic acid 94 (IAA-94), a known CLIC-specific ion channel inhibitor [30, 31]. IAA-94 treatment increased peroxide-induced apoptosis and completely prevented *CLIC5A*-mediated protection (**Figure 5D**) similarly to what was observed in the presence of increased Fe^{2+} availability following FAC treatment (**Figure 5C**). Endogenous *CLIC5A* inhibition contributes to this phenotype and explains the increased apoptosis level of control cells. Together, these data confirm that *CLIC5A*'s ion channel activity is required to protect cells against peroxide-induced apoptosis, perhaps by limiting Fe^{2+} availability in the lysosomal pathway.

To further investigate the functional implications of *CLIC5A* in lysosomal apoptosis, we examined the intracellular localization of *CLIC5A* using immunofluorescence. Though we did not observed co-localization of *CLIC5A* with lysosomes (**Figure S7**), we did show positive co-localization with transferrin receptor (**Figure 6**) which is in line with *CLIC5A*'s postulated role in modulating lysosomal Fe^{2+} availability. Transferrin receptors (*TFRC* gene) are responsible for cellular iron intake through binding and internalisation of its iron-bound ligand transferrin (*TF* gene) [32, 33]. *CLIC5A* could perturb this process and thereby impact cellular iron concentrations and modulate lysosome sensitivity to hydrogen peroxide. Interestingly, double-positive staining appears to be particularly present in distinct vesicle-like structures that are likely transferrin receptor-containing recycling endosomes, further suggesting that *CLIC5A* may

interfere with normal iron homeostasis thus reducing lysosome sensitivity to oxidative stress.

Taken together, our data strongly support a role for the newly identified *ETV6* transcriptional target *CLIC5A* in modulating lysosome-mediated apoptosis. We propose that *CLIC5A* overexpression in t(12;21)-positive, *ETV6* depleted pre-B cells could contribute to increased resistance to oxidative stress and therefore promote cell survival.

2.7 Discussion

Approximately 20% of childhood pre-B ALL patients harbor the t(12;21) translocation, yielding an ETV6-AML1 fusion protein that is, however, insufficient to initiate leukemia [8, 34, 35]. This process often requires further loss of the remaining wild type *ETV6* allele [7, 9-11, 13], suggesting that deregulation of the *ETV6* transcriptional machinery could play an important role in leukemogenesis. Unfortunately, very few *ETV6*-regulated transcriptional targets are known and the mechanisms through which they are involved in leukemogenesis remain elusive. Here, we combined both *in vitro* (cell lines) and *ex vivo* (pre-B ALL patients) transcriptome data to identify candidate *ETV6* target genes and through additional cellular assays identified 4 novel direct *ETV6* target genes: *CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L*.

The *CLIC5* gene was previously associated with the t(12;21)-positive ALL molecular signature [36] and our pediatric pre-B ALL expression data corroborated this result with strong *CLIC5* overexpression shown to be specific to the t(12;21)-positive subgroup within our cohort. Thus *CLIC5* overexpression in these patients is likely due to *ETV6* loss. Unfortunately, very little is known on *CLIC5* potential contribution to leukemogenesis. Using engineered cell lines, we demonstrated an increased resistance to hydrogen peroxide-induced apoptosis following overexpression of *CLIC5A*. Notably, Reh cells already express the endogenous *CLIC5A* isoform which could explain the modest effect of the overexpression of *CLIC5A* (mean = 7.46%, p-value = 8.32×10^{-11} , merged n = 36). Given that normal B-cells show no expression of *CLIC5* (mean FPKM < 0.1 over

our 3 normal CD19⁺/CD10⁺ pre-B cell samples isolated from human cord blood), an eventual stronger impact of *CLIC5A* re-expression in a pre-B cell upon *ETV6* depletion can be expected.

We further linked this phenotype to *CLIC5A*'s ion channel activity in lysosomal-mediated cell death. The presence of *CLIC5A* on transferrin receptor-containing endosomes potentially negatively impacts lysosomal iron availability thus reducing lysosome sensitivity to oxidative stress. However, the exact mechanism by which *CLIC5A* modulates lysosomal iron to prevent peroxide-induced apoptosis remains unclear. It has been demonstrated that changes in concentration of several ions can modulate endosomal pH through ion dependent H⁺ pumps [37]. With *CLIC5A* being a poorly selective ion channel [38], we can hypothesize a somewhat similar function: slight differences in endosome acidification could modulate transferrin iron release or trafficking and therefore impact lysosomal iron concentration leading to differential peroxide sensitivity. Additional experiments should be performed to further dissect *CLIC5A*'s role in lysosome-mediated apoptosis.

Nonetheless, the observed effects of *CLIC5A* in peroxide resistance could play a key role in ALL initiation. A recent study highlighted the high oxidative stress levels of leukemic blasts in the bone marrow niche induced by bone marrow stromal cell signaling [39]. Furthermore, it has been shown that increased levels of ROS in t(12;21)-positive cells was associated with DNA damage accumulation [40]. High oxidative stress levels are thought to trigger apoptotic cell death through the lysosomal pathway thus preventing DNA damage accumulation.

However in t(12;21)-positive ALL, we have shown that loss of *ETV6* expression leads to significant overexpression of *CLIC5A*. This overexpression leads to decreased lysosome-mediated apoptosis, which in turn can promote a more permissive environment to higher ROS levels (**Figure 7**). Cells evading apoptosis can thus accumulate ROS-induced mutations at a greater rate. Moreover, this mutational process of t(12;21)-positive pre-B cells could be facilitated not only by *CLIC5A* overexpression but together with several *ETV6* deregulated targets such as the inhibitor of apoptosis *BIRC7*. Although the overall mutational burden of ALL is low compared to other cancers [41], the difference in ROS-induced DNA damage accumulation can contribute, over time, to promote leukemic transformation when impacting key cancer genes.

It remains challenging however to evaluate the contribution of this pathway to the total amount of DNA damage found in pre-B ALL patients. Although ROS-mediated alterations of nucleotides are well characterized, their signature in sequencing data remains unclear [42, 43]. Interrogation of our patient sequencing data did not reveal any particular mutational profile that could undoubtedly be attributed to ROS. However, increased ROS-mediated DNA double strand breaks have been observed in t(12;21)-positive B-cells when assessed by comet assays [40]. These alterations do often lead to deletions or rearrangements, due to aberrant repair, that are frequent events in leukemia, thus supporting a role for ROS-induced DNA damage in ALL.

With a similar *CLIC5A*-mediated protective effect on peroxide-induced apoptosis obtained in both Reh and IM9 cell lines, we demonstrated a phenotype that is

independent of a particular genetic background. Interestingly, *CLIC5* expression has been associated with poor prognosis in breast cancer [44] and expression arrays of a broad range of normal and tumoral tissues obtained through the GENT database [45] showed an overexpression of *CLIC5* in ovarian cancers (**Figure S8**). Based on our results, and given the importance of oxidative stress resistance in these solid tumors that require angiogenesis to promote growth, it suggests that *CLIC5A* protection against oxidative stress may not be limited to pre-B cell leukemia. The unfavorable prognosis associated with *CLIC5* deregulation may be due to increased oxidative stress resistance thus reducing the necessity of angiogenesis and fostering an environment prone to ROS-induced DNA damage. Although *ETV6* alterations have not been reported in these solid tumors, other mechanisms could drive *CLIC5* overexpression. Interestingly, ETS factors binding sites (EBS) were found to be highly enriched for small non-coding mutations across a wide variety of cancers [46]. These mutations disrupt the consensus binding site of ETS factors such as *ETV6* and thus may prevent its binding and repression of its targets. Despite the presence of wild-type *ETV6* in these cases, some of its key targets (*CLIC5*, *BIRC7* or *ANGPTL2*) may be overexpressed thus leading to a selective advantage.

While we focused on *CLIC5*, the other identified targets were also of interest. Notably, the caspases inhibitor *BIRC7* has previously been shown to be part of a t(12;21)-positive pre-B ALL molecular signature [36]. Overexpression of *BIRC7* was also observed in a variety of tumors and contributes to oncogenesis through inhibition of apoptosis [47], suggesting a similar involvement for *BIRC7* in

t(12;21)-positive childhood pre-B ALL. The *ANGPTL2* gene encodes a secreted protein with, among others, pro-angiogenic and anti-apoptotic properties. Although its expression has been linked to a broad range of diseases, including cancers, it has been shown to increase survival and expansion of hematopoietic stem cells [48]. *ANGPTL2*'s contribution to leukemogenesis remains unknown. Lastly, *WBP1L*, also known as *OPAL1* (*Outcome Predictor in Acute Leukemia 1*), is associated with the t(12;21) favorable outcome in ALL [49]. It is still unclear though whether *WBP1L* plays a role in leukemia since its molecular function is yet to be characterized.

In conclusion, for the first time, we describe a role for *CLIC5*-mediated resistance to oxidative stress that could promote cell survival and contribute to leukemogenesis. We propose a mechanism in which complete loss of wild-type *ETV6* expression in a pre-leukemic blast leads to *CLIC5A* overexpression, thus creating a permissive environment for the accumulation of mutations driven by high oxidative stress, eventually giving rise to full leukemic transformation.

2.8 Declarations

2.8.1. Availability of data and material

The datasets generated and/or analysed during the current study are available in the GEO repository,

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79373>

2.8.2. Competing interests

The authors declare that they have no competing interests

2.8.3. Funding

This study was supported by research funds provided by the Terry Fox Research Institute and the Canadian Institutes for Health Research. BN is the recipient of a Cole Foundation scholarship. JFS is the recipient of a Réseau de médecine génétique appliquée (RMGA) scholarship. DS holds the François-Karl-Viau Research Chair in Pediatric Oncogenomics. Whole transcriptome sequencing was performed at the Integrated Clinical Genomics Centre in Pediatrics at the CHU Sainte-Justine Research Center. Computations were performed on the Briarée supercomputer at the Université de Montréal, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). GH thanks the CFI, the Cole Foundation and the Fonds de recherche du Québec-Santé for infrastructure, transition award and salary support, respectively.

2.8.4. Acknowledgments

The authors would like to thank the patients and their parents for participating in this study. We are grateful to Dr. Christian Beauséjour who kindly provided lentiviral-related material.

2.9 Figures

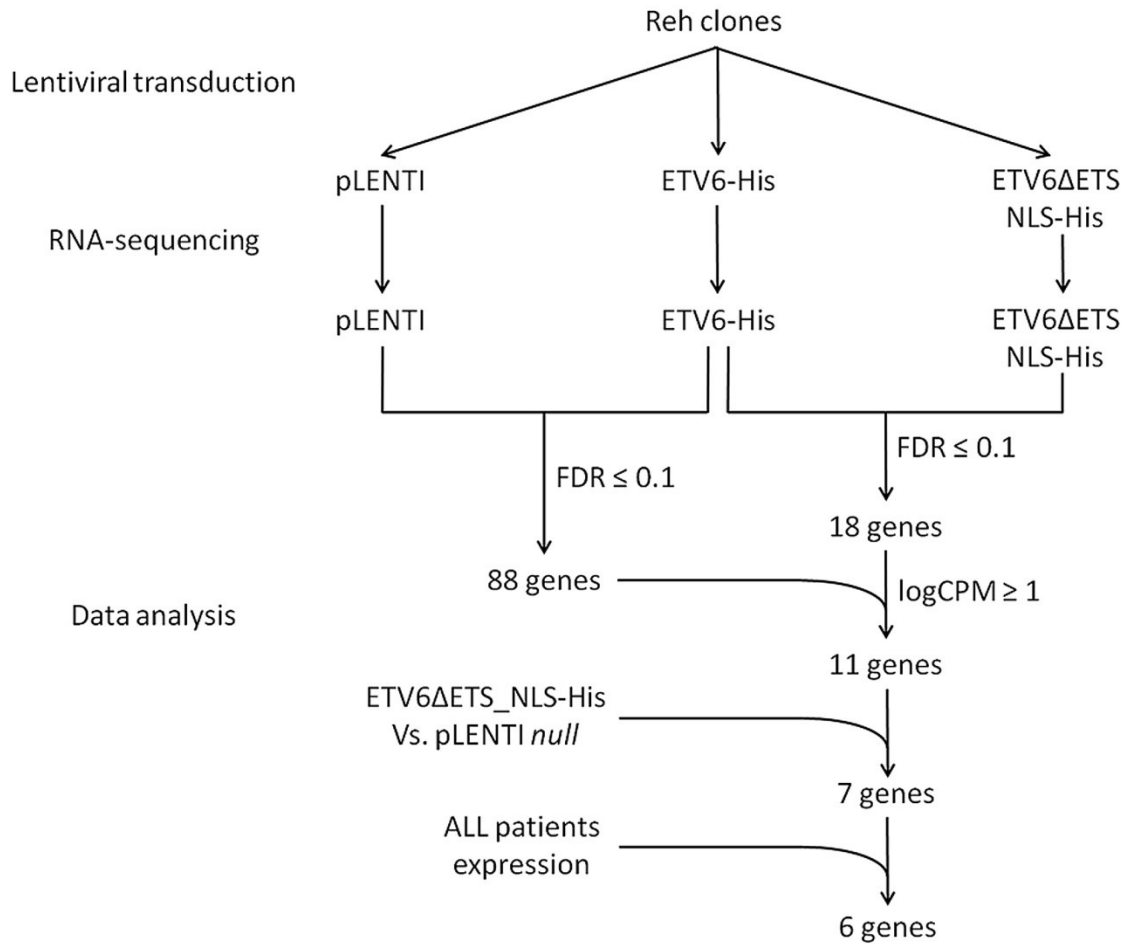


Figure 1: Schematic representation of the transcriptome-based design to detect putative direct *ETV6* target genes. To identify direct targets of *ETV6*, we first designed an *in vitro* RNA-seq experiment using *ETV6*^{-/-} Reh-derived clones. Cells were transduced with lentiviral constructs to express ETV6-His and ETV6ΔETS_NLS-His. Total RNA was extracted from stable cell populations and RNA-seq libraries were sequenced. Expression profiles were analyzed using EdgeR. Gene expression profiles in ETV6-His cells were first compared with ETV6ΔETS_NLS-His and pLENTI cells to identify repressed genes (FDR ≤ 0.1). We then included data from the ETV6ΔETS_NLS-His vs. pLENTI comparison and further considered genes whose expression remains constant (p-value ≥ 0.05 or logFC ≥ -0.5) which are more likely to be direct *ETV6* targets. Finally, only genes that showed a specific overexpression in t(12;21)-positive childhood pre-B ALL patients were considered.

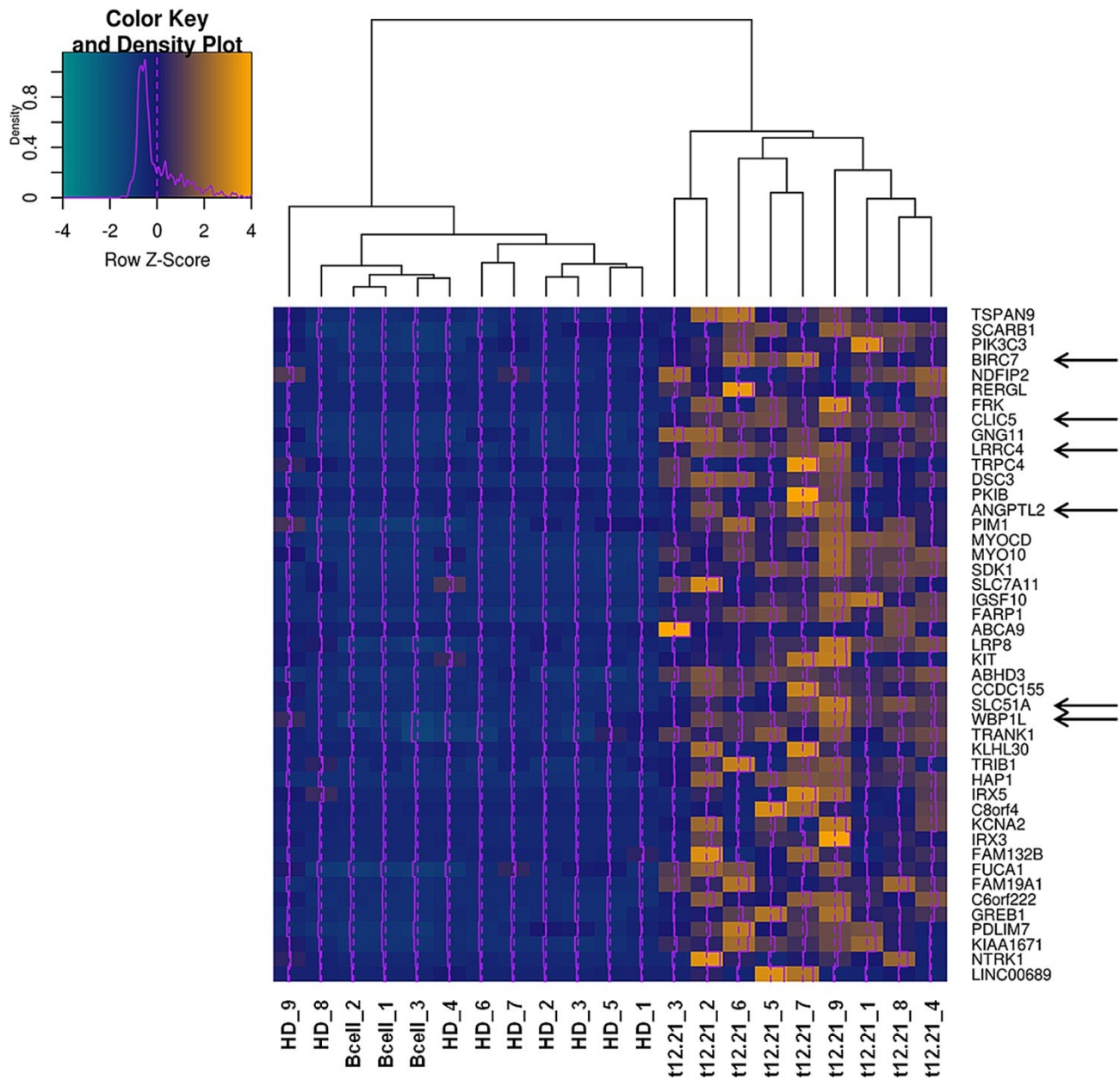


Figure 2: Expression of putative *ETV6* targets in pediatric pre-B ALL patients. An unsupervised clustering heatmap was generated with the 45 genes specifically overexpressed (yellow) in t(12;21)-positive ALL patients (t12.21, n=9) compared to hyperdiploid ALL patients (HD, n=9) and normal B-cells controls (Bcell, n=3). Among the 7 genes identified as putative direct *ETV6* targets in our cell line model, *CLIC5*, *BIRC7*, *ANGPTL2*, *WBP1L*, *LRRC4* and *SLC51A* were specifically overexpressed in t(12;21)-positive pediatric pre-B ALL patient tumor samples (Arrows).

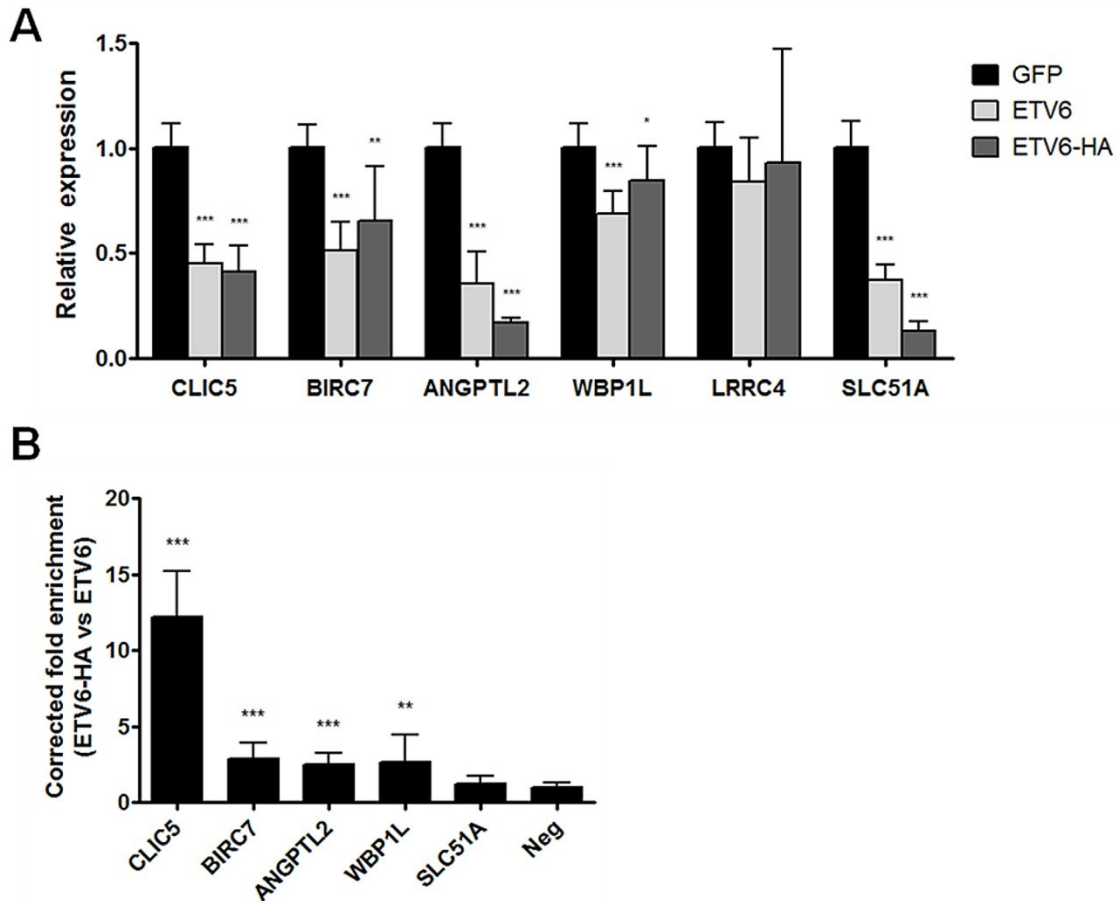


Figure 3: Quantitative real-time PCR and chromatin immunoprecipitation validation of putative *ETV6* target genes. Reh cells together with the two Reh derived clones were infected with pCCL GFP, *ETV6* WT or *ETV6*-HA. **A.** Total RNA was extracted from these cells and complementary DNA (cDNA) was generated. This cDNA was submitted to qRT-PCR analysis to quantify relative expression of putative *ETV6* targets. Expression of all genes but *LRRC4* is repressed by *ETV6* WT and *ETV6*-HA. **B.** CHIP experiments were performed in *ETV6* WT and *ETV6*-HA cells. Putative *ETV6* target gene proximal promoter enrichment was determined by qRT-PCR using promoter-specific primers. Results are presented as the ratio of the input percentage obtained in *ETV6*-HA cells compared to *ETV6* WT cells, corrected by the background enrichment obtained with an unbound region (Neg). *CLIC5*, *BIRC7*, *ANGPTL2* and *WBP1L* promoters are enriched. For both qRT-PCR and CHIP experiments, 4 values were calculated for each of the 3 cell lines and were merged for a total of 12 values (n=12). Error bars represent the standard deviation. Statistical significance is calculated by two-tailed and one-tailed Mann-Whitney-U test for qRT-PCR and CHIP analysis, respectively.

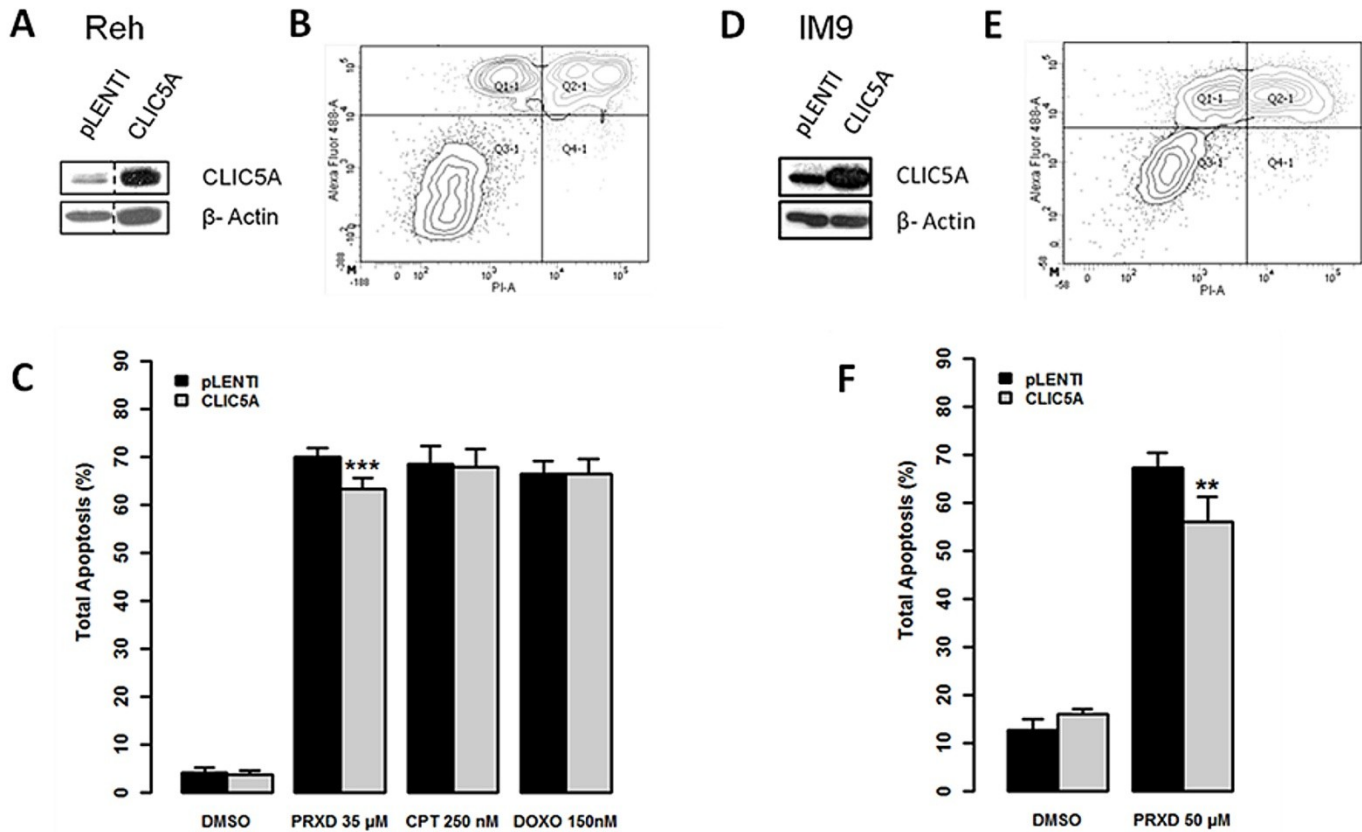


Figure 4: *CLIC5A* protects cells from hydrogen peroxide-induced apoptosis. **A.** pLENTI control and *CLIC5A* stably infected Reh cells were challenged for 20h with 35μM hydrogen peroxide (PRXD), 250nM camptothecin (CPT), or 150nM doxorubicin (DOXO) and analysed by flow cytometry with Alexa Fluor 488-coupled Annexin V and propidium iodide (PI) staining. **B.** A representative example of staining following PRXD treatment is presented with PI on the x axis and Alexa Fluor 488 on the y axis. **C.** Total apoptosis is calculated for each sample. *CLIC5A* overexpressing cells displayed reduced apoptosis compared to control only when treated with PRXD. Each experiments were performed 3 times in triplicates (n=9). **D.** IM9 cells transduced with *CLIC5A* or pLENTI empty vector control were challenged with 50μM PRXD and processed similarly to assess apoptosis. **E.** A representative example of staining and **F.** total apoptosis is shown for the IM9 cell line. Again, *CLIC5A* overexpression leads to reduced apoptosis. Two independent experiments carried out in triplicates and an additional single test were performed (n=7). Error bars in C. and F. represent the standard deviation. Statistical significance is calculated by two-tailed Mann-Whitney-U test. For A. and D. adjustments of brightness and contrast were applied to the whole image.

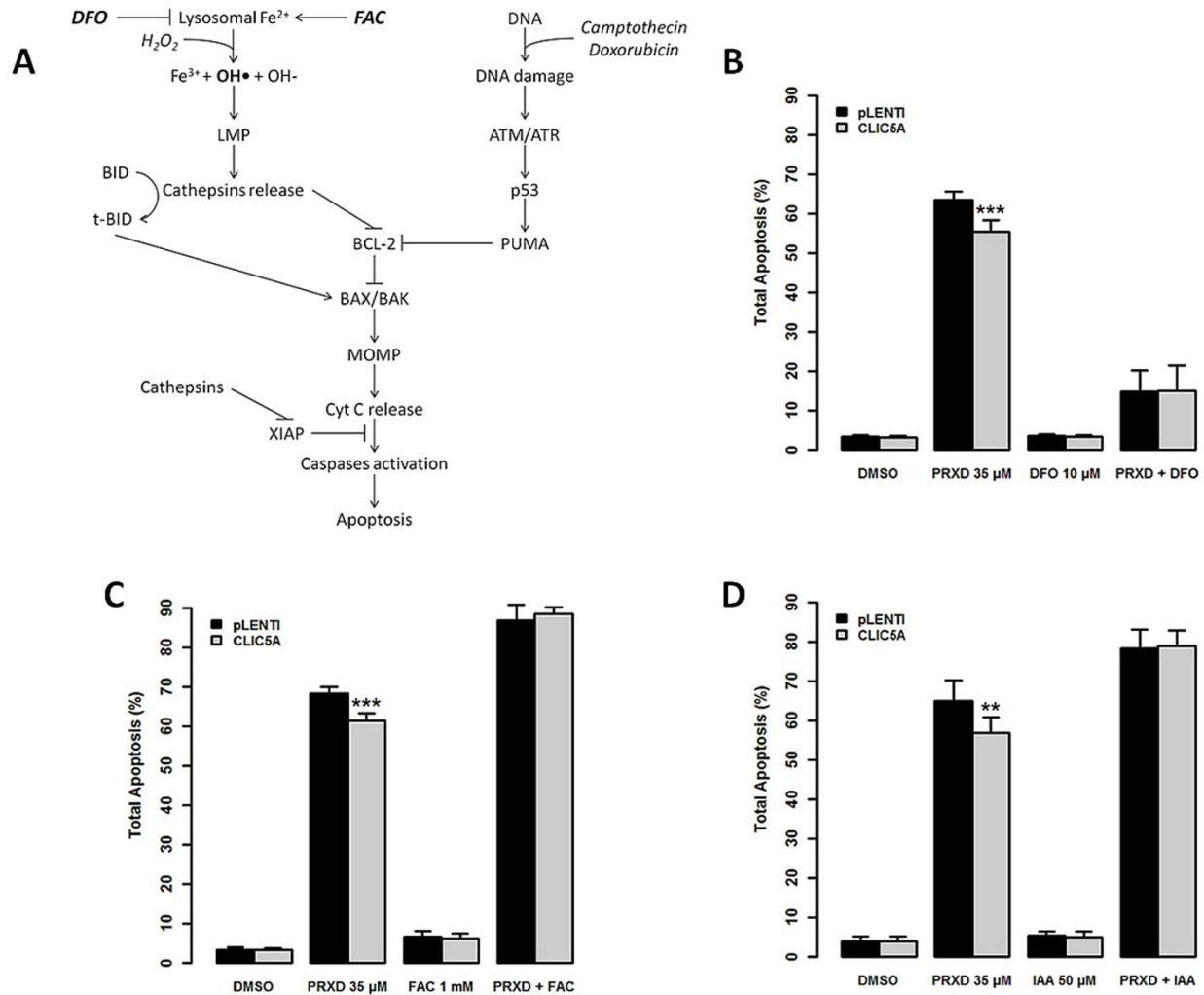


Figure 5: *CLIC5A* is implicated in lysosome-mediated cell death. **A.** Schematic illustration of the lysosomal apoptosis pathway and DNA damage pathway. DFO: deferoxamine mesylate salt; FAC: ferric ammonium citrate; LMP: Lysosomal membrane permeabilization; MOMP: Mitochondrial outer membrane permeabilization. **B.** pLENTI control and *CLIC5A* overexpressing Reh cells were pre-treated with 10 μ M DFO to chelate lysosomal ferrous iron and apoptosis was induced using 35 μ M PRXD for 20h followed by flow cytometry quantification. PRXD-induced apoptosis was greatly reduced with DFO treatment. **C.** Similarly, 1mM FAC was used to increase ferrous iron concentration which led to higher PRXD-induced apoptosis with no protective effect of *CLIC5A* overexpression. **D.** Cells were pre-treated with 50 μ M IAA-94 CLIC-specific ion channel inhibitor and apoptosis was assayed after a subsequent PRXD treatment. In these conditions, *CLIC5A* overexpression did not reduce apoptosis. Each experiment was carried out 3 times in triplicate (n=9). Error bars represent the standard deviation. Statistical significance is calculated by two-tailed Mann-Whitney-U test.

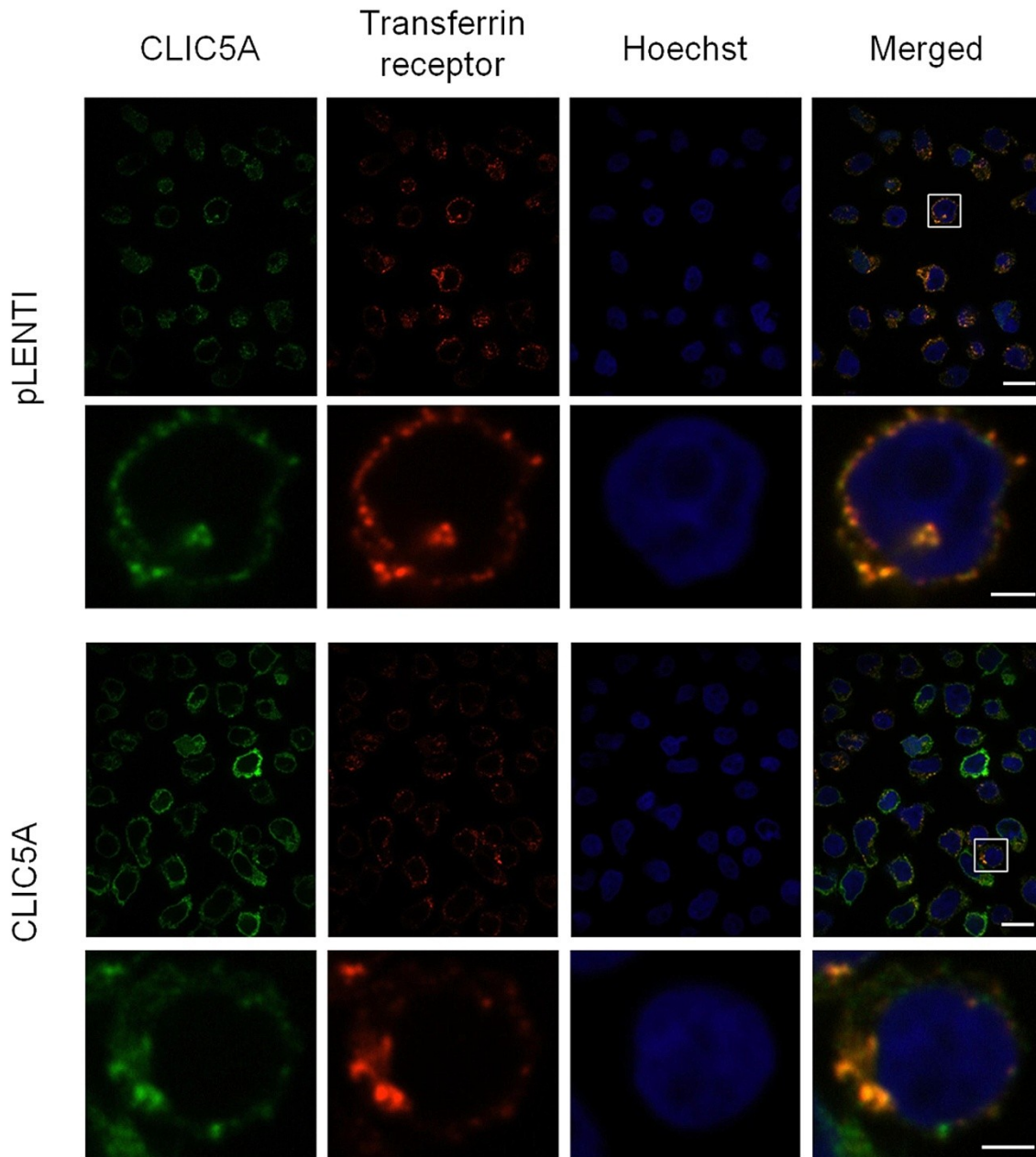


Figure 6: Co-localization of *CLIC5A* and transferrin receptors. pLENTI control and *CLIC5A* overexpressing Reh cells were used for co-localization studies. Immunostaining of both *CLIC5A* and transferrin receptor were performed in the same cells simultaneously with Hoechst DNA staining. Results were obtained at 100X magnification (upper panels; scale bar = 10 μ m). A strong co-localization was observed between *CLIC5A* and transferrin receptors. Additional enlargement for the marked region of the initial image is presented in lower panels (scale bar = 2 μ m). A merged image was generated (right panels).

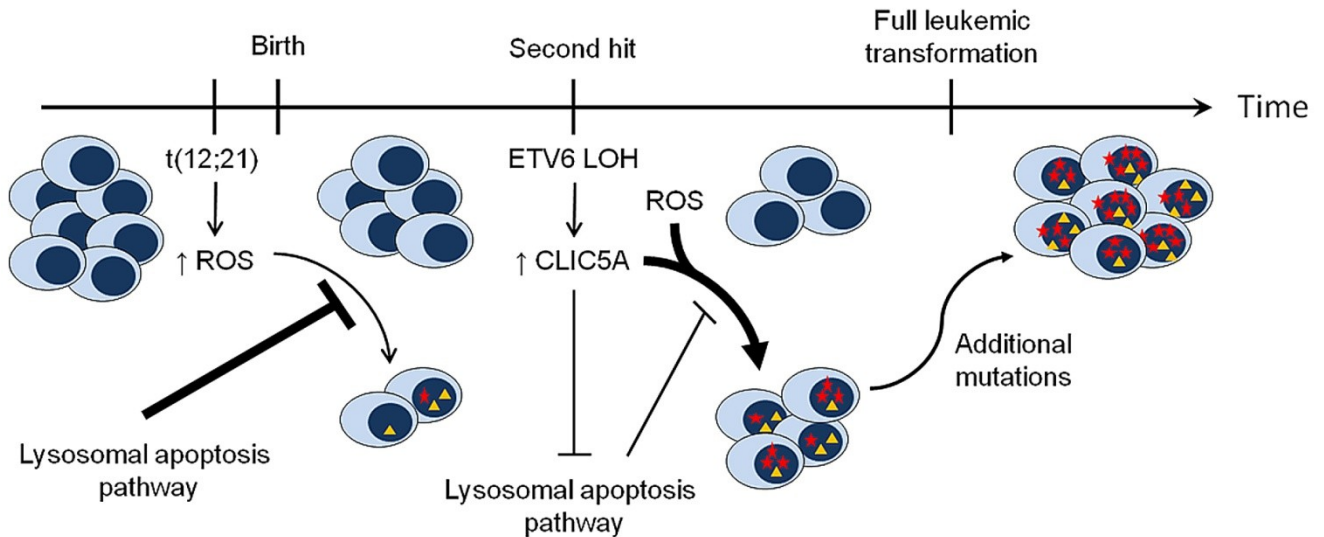


Figure 7: Proposed mechanism of *CLIC5A* involvement in *ETV6*-associated childhood pre-B ALL. The *t(12;21)* translocation occurs early and contributes to an increased level of reactive oxygen species (ROS). Lysosome-mediated apoptosis is sensitive to this excess of ROS and therefore prevents the accumulation of ROS-mediated DNA damage. With the subsequent deletion of the residual *ETV6* allele (LOH), *CLIC5A* expression is drastically upregulated. This overexpression of *CLIC5A* has a negative impact on the lysosomal apoptosis pathway and thus creates a permissive environment for the accumulation of mutations driven by a high oxidative stress. Over time, some of these mutations may impact key cellular biological processes and pathways which will ultimately lead to full leukemic transformation and development of childhood pre-B ALL.

2.10 Tables

Table 1: Expression status of the 6 putative direct *ETV6* target genes

		Gene symbol						
		<i>CLIC5</i>	<i>BIRC7</i>	<i>ANGPTL2</i>	<i>WBP1L</i>	<i>LRRC4</i>	<i>SLC51A</i>	
<i>In vitro</i>	ETV6-His vs ETV6ΔETS_NLS-His	logFC	-3.23	-1.93	-1.31	-0.79	-1.02	-1.31
		logCPM	4.66	4.02	5.00	5.78	3.90	3.11
		PValue	7.00E-52	3.44E-11	4.24E-11	8.04E-06	9.18E-05	1.14E-04
		FDR	9.59E-48	1.45E-07	1.45E-07	9.18E-03	6.29E-02	6.79E-02
	ETV6-His vs pLENTI	logFC	-3.20	-1.39	-1.68	-1.10	-1.16	-1.50
		logCPM	3.75	3.02	4.37	5.18	3.23	2.43
		PValue	6.40E-39	2.30E-05	3.34E-14	1.06E-09	3.99E-06	1.45E-05
		FDR	8.09E-35	4.76E-03	6.41E-11	6.71E-07	1.20E-03	3.28E-03
	ETV6ΔETS_NLS-His vs pLENTI	logFC	0.03	0.53	-0.38	-0.32	-0.15	-0.20
		logCPM	4.62	3.85	4.81	5.49	3.63	2.90
		PValue	0.88	0.03	0.02	0.06	0.55	0.50
		FDR	1	1	1	1	1	1
ALL Patients t(12;21) vs B-cells	logFC	6.36	6.79	5.82	1.92	4.25	4.27	
	logCPM	10.55	9.74	9.42	10.68	9.30	8.26	
	PValue	1.27E-30	1.57E-13	6.39E-08	4.87E-07	7.67E-12	1.91E-06	
	FDR	6.91E-29	1.55E-12	3.11E-07	2.17E-06	6.25E-11	7.59E-06	

ALL: acute lymphoblastic leukemia; logFC: log fold change; logCPM: log counts per million reads; FDR: false discovery rate.

2.11 References

1. Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem.* 1999; 274(42):30132-30138.
2. Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F et al. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* 1998; 12(15):2392-2402.
3. Bohlander SK. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol.* 2005; 15(3):162-174.
4. Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P et al. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A.* 1995; 92(11):4917-4921.
5. Tasian SK, Loh ML, Hunger SP. Childhood acute lymphoblastic leukemia: Integrating genomics into therapy. *Cancer.* 2015; 121(20):3577-3590.
6. Agape P, Gerard B, Cave H, Devaux I, Vilmer E, Lecomte MC et al. Analysis of ETV6 and ETV6-AML1 proteins in acute lymphoblastic leukaemia. *Br J Haematol.* 1997; 98(1):234-239.
7. Poirel H, Lacronique V, Mauchauffe M, Le Coniat M, Raffoux E, Daniel MT et al. Analysis of TEL proteins in human leukemias. *Oncogene.* 1998; 16(22):2895-2903.
8. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A.* 2002; 99(12):8242-8247.
9. Patel N, Goff LK, Clark T, Ford AM, Foot N, Lillington D et al. Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br J Haematol.* 2003; 122(1):94-98.
10. Lilljebjorn H, Sonesson C, Andersson A, Heldrup J, Behrendtz M, Kawamata N et al. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet.* 2010; 19(16):3150-3158.
11. Montpetit A, Larose J, Boily G, Langlois S, Trudel N, Sinnett D. Mutational and expression analysis of the chromosome 12p candidate tumor suppressor genes in pre-B acute lymphoblastic leukemia. *Leukemia.* 2004; 18(9):1499-1504.

12. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol.* 2015.
13. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature.* 2011; 469(7330):356-361.
14. Fenrick R, Wang L, Nip J, Amann JM, Rooney RJ, Walker-Daniels J et al. TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol.* 2000; 20(16):5828-5839.
15. Irvin BJ, Wood LD, Wang L, Fenrick R, Sansam CG, Packham G et al. TEL, a putative tumor suppressor, induces apoptosis and represses transcription of Bcl-XL. *J Biol Chem.* 2003; 278(47):46378-46386.
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1):15-21.
17. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31(2):166-169.
18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26(1):139-140.
19. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-(-Delta Delta C(T))} Method. *Methods.* 2001; 25(4):402-408.
20. Shanks RA, Larocca MC, Berryman M, Edwards JC, Urushidani T, Navarre J et al. AKAP350 at the Golgi apparatus. II. Association of AKAP350 with a novel chloride intracellular channel (CLIC) family member. *J Biol Chem.* 2002; 277(43):40973-40980.
21. Flores-Tellez TN, Lopez TV, Vasquez Garzon VR, Villa-Trevino S. Co-Expression of Ezrin-CLIC5-Podocalyxin Is Associated with Migration and Invasiveness in Hepatocellular Carcinoma. *PLoS One.* 2015; 10(7):e0131605.
22. Xu Y, Kang J, Yuan Z, Li H, Su J, Li Y et al. Suppression of CLIC4/mtCLIC enhances hydrogen peroxide-induced apoptosis in C6 glioma cells. *Oncol Rep.* 2013; 29(4):1483-1491.

23. Turk B, Turk V. Lysosomes as "suicide bags" in cell death: myth or reality? *J Biol Chem.* 2009; 284(33):21783-21787.
24. Aits S, Jaattela M. Lysosomal cell death at a glance. *J Cell Sci.* 2013; 126(Pt 9):1905-1912.
25. Kruszewski M. Labile iron pool: the main determinant of cellular response to oxidative stress. *Mutat Res.* 2003; 531(1-2):81-92.
26. Kurz T, Leake A, Von Zglinicki T, Brunk UT. Relocalized redox-active lysosomal iron is an important mediator of oxidative-stress-induced DNA damage. *Biochem J.* 2004; 378(Pt 3):1039-1045.
27. Boya P, Kroemer G. Lysosomal membrane permeabilization in cell death. *Oncogene.* 2008; 27(50):6434-6451.
28. Oller AR, Thilly WG. Mutational spectra in human B-cells. Spontaneous, oxygen and hydrogen peroxide-induced mutations at the hprt gene. *J Mol Biol.* 1992; 228(3):813-826.
29. Repnik U, Cesen MH, Turk B. The endolysosomal system in cell death and survival. *Cold Spring Harb Perspect Biol.* 2013; 5(1):a008755.
30. Berryman M, Bruno J, Price J, Edwards JC. CLIC-5A functions as a chloride channel in vitro and associates with the cortical actin cytoskeleton in vitro and in vivo. *J Biol Chem.* 2004; 279(33):34794-34801.
31. Landry DW, Akabas MH, Redhead C, Edelman A, Cragoe EJ, Jr., Al-Awqati Q. Purification and reconstitution of chloride channels from kidney and trachea. *Science.* 1989; 244(4911):1469-1472.
32. Bresgen N, Eckl PM. Oxidative stress and the homeodynamics of iron metabolism. *Biomolecules.* 2015; 5(2):808-847.
33. Maxfield FR, McGraw TE. Endocytic recycling. *Nat Rev Mol Cell Biol.* 2004; 5(2):121-132.
34. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet.* 2001; 130(2):93-104.
35. van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J et al. Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood.* 2011; 118(4):1041-1051.

36. Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*. 2003; 102(8):2951-2959.
37. Scott CC, Gruenberg J. Ion flux and the function of endosomes and lysosomes: pH is just the start: the flux of ions across endosomal membranes influences endosome function not only through regulation of the luminal pH. *Bioessays*. 2011; 33(2):103-110.
38. Singh H, Cousin MA, Ashley RH. Functional reconstitution of mammalian 'chloride intracellular channels' CLIC1, CLIC4 and CLIC5 reveals differential regulation by cytoskeletal actin. *FEBS J*. 2007; 274(24):6306-6316.
39. Liu J, Masarekar A, Johnson S, Chakraborty S, Griffiths J, Smith D et al. Stromal cell-mediated mitochondrial redox adaptation regulates drug resistance in childhood acute lymphoblastic leukemia. *Oncotarget*. 2015.
40. Kantner HP, Warsch W, Delogu A, Bauer E, Esterbauer H, Casanova E et al. ETV6/RUNX1 induces reactive oxygen species and drives the accumulation of DNA damage in B cells. *Neoplasia*. 2013; 15(11):1292-1300.
41. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415-421.
42. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J*. 2003; 17(10):1195-1214.
43. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*. 2014; 15(9):585-598.
44. Yau C, Sninsky J, Kwok S, Wang A, Degnim A, Ingle JN et al. An optimized five-gene multi-platform predictor of hormone receptor negative and triple negative breast cancer metastatic risk. *Breast Cancer Res*. 2013; 15(5):R103.
45. Shin G, Kang TW, Yang S, Baek SJ, Jeong YS, Kim SY. GENT: gene expression database of normal and tumor tissues. *Cancer Inform*. 2011; 10:149-157.
46. Araya CL, Cenik C, Reuter JA, Kiss G, Pande VS, Snyder MP et al. Identification of significantly mutated regions across cancer types highlights a rich landscape of functional molecular alterations. *Nat Genet*. 2015.

47. Wang L, Zhang Q, Liu B, Han M, Shan B. Challenge and promise: roles for Livin in progression and therapy of cancer. *Mol Cancer Ther.* 2008; 7(12):3661-3669.
48. Thorin-Trescases N, Thorin E. Angiopoietin-like-2: a multifaceted protein with physiological and pathophysiological properties. *Expert Rev Mol Med.* 2014; 16:e17.
49. Holleman A, den Boer ML, Cheok MH, Kazemier KM, Pei D, Downing JR et al. Expression of the outcome predictor in acute leukemia 1 (OPAL1) gene is not an independent prognostic factor in patients treated according to COALL or St Jude protocols. *Blood.* 2006; 108(6):1984-1990.

2.12 Supplementary material

2.12.1. Supplemental methods

Constructs

pcDNA3.1 ETV6 contains the complete wild-type sequence of *ETV6*. pcDNA3.1 ETV6 Δ ETS was generated by creating KpnI restriction sites flanking the ETS domain by site-directed mutagenesis (QuickChange Multi-Site/XL Directed Mutagenesis kit; Stratagene) followed by enzymatic digestion and re-ligation. The SV40 Large T antigen Nuclear Localization Signal (NLS; PKKKRKV) was added to this mutant to assure its nuclear localization. ETV6 Δ ETS was amplified by PCR using the forward primer 5'-CCGAATTCGGCACGAGGAACTTCTTA -3' and NLS containing (underlined) reverse primer 5'-CCGAATTCTCACACCTTCCTTCTTCTTCGGTTGTTTCATCCAGCTCCTGGGACTC-3'. EcoRI digestion was used to clone the ETV6 Δ ETS_NLS PCR product into pcDNA3.1 vector. His tag was added in C-terminal to both ETV6 and ETV6 Δ ETS_NLS by a similar approach using His-tag containing reverse primers (**Table S5**) and EcoRI/NotI or BamHI/NotI, respectively. ETV6-His and ETV6 Δ ETS_NLS-His constructs were subcloned into Gateway compatible vector pENTR3C and transferred into pLENTI CMV puro DEST (w118-1) lentiviral vector (kindly provided by Dr. Christian Beauséjour) using LR clonase II reactions (Thermo Fisher Scientific). A C-terminal HA-tagged *ETV6* was also generated. An oligomer containing 3 tandem HA tag repeats flanked by BSMI and NotI restriction sites (**Table S5**) was synthesized (Invitrogen), annealed and

subcloned directly into pcDNA3.1 ETV6 (partial BSMI digestion). Both ETV6 and ETV6-HA were subcloned into pCCL lentiviral vector (kindly provided by Dr. Christian Beauséjour). Briefly, GFP was excised from pCCL hPGK-GFP with AgeI/Sall digestion. ETV6 and ETV6-HA cDNA were obtained by enzymatic digestion. Extremities were filled-in for a subsequent blunt-end ligation. All *ETV6* constructs are shown in **Figure S9**. pDONR221 CLIC5A was purchased from Harvard PlasmID Repository (clone # HsCD00044695) and transferred into pLENTI lentiviral vector as described previously.

Cell culture

Reh (ATCC® CRL-8286™) and IM9 (ATCC® CCL159™) cell lines were maintained in RPMI 1640 (Wisent) 10% Fetal Bovine Serum (FBS; Wisent) in a 5% CO₂ incubator at 37°C.

Lentiviral production

15X10⁶ HEK293T cells (kindly provided by Dr. Christian Beauséjour) were seeded into 15cm petri dishes in DMEM (Wisent) 10% FBS. The next day, cells were transfected with 9µg lentiviral expression plasmid (pLENTI or pCCL backbone) together with 6µg pRSV-Rev, 7.8µg pMD2.VSVG and 15µg pMDL third generation encapsidation plasmids (kindly provided by Dr. Christian Beauséjour) in fresh RPMI 1640 10% FBS medium using Polyethylenimine (Polysciences). Media was removed 16h post-transfection and replaced by fresh DMEM 10% FBS. After 30h, viral particles were either retrieved from media by ultracentrifugation (50 000g) and quantified by p24 antigen ELISA (Advanced

Bioscience Laboratories) or media was used directly for transduction (see lentiviral infection section).

Lentiviral infection

1×10^6 Reh or IM9 cells were seeded in 1mL of RPMI 1640 10% FBS medium. Either concentrated virus (100-1000ng p24) or 1mL of viral supernatant were added to cells. Polybrene was added to a final concentration of 5-8 μ g/mL. 24h post-infection, medium was changed with fresh RPMI 1640 10% FBS. The next day (48h post-infection), infected cells were selected with 1 μ g/mL puromycin (for pLENTI infected cells). These stably infected cells were maintained at least 2 weeks in culture before validating expression by western blotting (see related section) and carrying out further experiments.

Expression profiling by RNA-sequencing

Total RNA from two different Reh clones (generated in methylcellulose media) each stably expressing ETV6-His and ETV6 Δ ETS_NLS-His (and pLENTI control) was extracted (RNeasyTM mini kit, QIAGEN) and quantified by Nanodrop (Thermo Fisher scientific). RNA quality was assessed using the Agilent 2100 Bioanalyzer. Libraries were generated with 1 μ g total RNA (RIN \geq 9.9) treated with DNaseI (Turbo DNA-free; Ambion) and processed through the TruSeq Stranded Total RNA protocol with ribosomal ribonucleic acid (rRNA) removal mix (Illumina). Two libraries per high throughput lane of HiSeq 2500 (Illumina) were sequenced for an average coverage of approximately 100 million reads per sample. Reads for each samples were mapped on hg19 reference genome using STAR with default

settings and read counts per genes were determined using HTSeq-count. For the differentially expressed genes (DEG) analysis, the R bioconductor package edgeR was used with Benjamini Hochberg p-value adjustment. The two clones were considered as biological replicates.

The patient cohort used for RNA-sequencing was composed of 9 hyperdiploid and 9 t(12;21) patients. Using the mirVana Isolation kit (Ambion) according to the manufacturer's protocol, total RNA was extracted from leukemic bone marrow samples of all patients and from control pre-B cells (CD19⁺CD10⁺) isolated from healthy cord blood samples by Fluorescence Activated Cell Sorting (FACS). Following a DNaseI treatment, total or mature RNA samples were quantified by NanoDrop ND1000 (Thermo-Fisher Scientific, Waltham, MA) and RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). rRNA were depleted using the Invitrogen RiboMinus Eukaryote kit (Life Technologies). cDNA libraries were prepared using the SOLiD Total RNA-seq kit based on manufacturer's protocol and sequenced on the Life Technologies SOLiD 4/5500 System (paired-end: 50bp x 35bp). Reads were aligned to the hg19 reference genome and read counts per gene obtained using LifeScope Genomic Analysis Software (Whole Transcriptome Analysis pipeline, default parameters). Remaining ribosomal sequences were filtered out. Based on results obtained from the cell lines, further analysis were limited to a selection of the top 331 candidate genes (p-value < 0.05). DEG analysis was done as described above. Reads Per Kilobase per Million mapped reads (RPKM) were calculated for the identified genes using edgeR and a heatmap depicting the unsupervised

clustering of samples based on RPKM values of the overexpressed candidate genes was generated using the heatmap.2 library of the gplots R package.

Quantitative real-time PCR

Total RNA was extracted and quantified as described above. 350ng of total RNA were retro-transcribed with M-MLV reverse transcriptase (Life Technologies). cDNA was then subjected to quantitative real-time PCR using the primer sets listed in **Table S1**. Reactions were performed in duplicate in a total volume of 25 μ L composed of 5 μ L cDNA (diluted 1:10), 5 μ L primers (0.2 μ M final concentration), 12.5 μ L SYBR Green PCR Master Mix (1X final concentration, Life technologies) and completed with H₂O. Amplification and detection were carried out on the ABI PRISM 7000 sequence detection system (Life Technologies) on the course of 40 cycles [95°C 15 sec; 60°C 1min]. Relative expression was determined by the $2^{-(\Delta\Delta Ct)}$ comparative method using *GAPDH* as the reference gene.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) have been performed on 10×10^6 transduced Reh cells cross-linked directly in cell medium for 10min with 1% methanol-free formaldehyde (Polysciences). Cells were resuspended in TpA buffer (0.25% Triton, 10mM Tris pH 8.0, 10mM EDTA, 0.5mM EGTA, 1X protease inhibitor cocktail [Roche]) and incubated on ice for 5min. Pelleted cells were then resuspended in TpB buffer (200mM NaCl, 10mM Tris pH 8.0, 1mM EDTA, 0.5mM EGTA, 1X protease inhibitor cocktail) and incubated on ice for

30min. Fragilized nuclei were pelleted and resuspended in TpS buffer (0.5% SDS, 0.5% Triton, 10mM Tris pH 8.0, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 1X protease inhibitor cocktail) for sonication. Chromatin was sheared using a S2 ultrasonicator (Covaris) and cleared by centrifugation. Supernatant (cleared sheared chromatin) was combined with 1.4 volume TpS SDS-free and 2.5 volume of pre-washed anti-HA magnetic beads (Thermo Fisher Scientific) in ChIP dilution buffer (1% Triton, 10mM Tris pH 8.0, 150mM NaCl, 2mM EDTA) followed by a 4h incubation with rotation (4°C). A fraction (1:10) of supernatant was conserved for Inputs. Magnetic beads were washed twice with wash buffer (0.5% NP40, 150mM KCl, 10mM Tris pH 8.0, 1mM EDTA). To elute DNA-protein complexes from beads, we used a competitive approach using a solution of HA peptides (Thermo Fisher Scientific). Beads were eluted twice with HA peptides (2mg/mL in TBS) for 30min at 37°C. Reverse-crosslinking was performed for both eluates and Input DNA in TpE buffer (0.3% SDS, 50mM Tris pH 8.0, 10mM EDTA, 400mM NaCl) at 65°C overnight. 1.3 volume of RNase A mix (154µg/mL glycogen [Roche], 115µg/mL RNase A [Roche] in TE buffer) was added to all samples followed by a 2h incubation at 37°C. SDS and proteinase K (Thermo Fisher Scientific) were added to a final concentration of 0.3% and 8.4µg/mL, respectively, for an additional 2h incubation at 37°C. DNA was finally purified twice by standard phenol/chloroform/isoamyl alcohol (Sigma) extraction using MaXtract High Density gel columns (QIAGEN) and ethanol precipitation. DNA pellets were washed with 70% ethanol, dried and resuspended in EB buffer (10mM Tris pH 8.0). This purified ChIP and Input DNA were quantified by qRT-

PCR using primers designed in the first 500bp upstream of genes' transcription start site (listed in **Table S2**).

Protein extraction

Whole cell extracts were prepared in RIPA buffer (50mM Tris pH 8.0, 150mM NaCl, Nonidet P-40 1%, 0.5% NaDOC, 0.1% SDS, 1X protease inhibitor cocktail) and mechanical disruption in 21G gauge needles (Becton Dickinson). For nuclear extracts, cells pellets were resuspended in A Buffer (10mM HEPES pH 7.9, 10mM KCl, 0.1mM EDTA, 0.1 mM EGTA, 1mM DTT, 0.5mM AEBSF, 1X protease inhibitor cocktail) for 15min followed by addition of Nonidet P-40. Nuclei were pelleted, resuspended in C buffer (20mM HEPES pH 7.9, 400mM NaCl, 1mM EDTA, 1mM EGTA, 1mM DTT, 1mM AEBSF, 1X protease inhibitor cocktail) and incubated with rotation to released nuclear proteins. After centrifugation, supernatant was collected. Both total and nuclear extracts were quantified on SpectraMax 190 (Molecular Devices) using the DC protein assay (BioRad) according to the manufacturer protocol.

Western blotting

20µg of nuclear extracts or 100µg of whole cell extracts were diluted in Laemmli buffer (final concentration of 63 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 100 mM DTT, 0.0025% bromophenol blue) prior to heat denaturation at 95°C for 3min. Samples were loaded and migrated on SDS-denaturing 10% polyacrylamide gels (37.5:1) and transferred at 4°C overnight on polyvinylidene difluoride membranes in transfer buffer (25mM Tris, 192mM glycine, 10%

methanol). Membranes were blocked in Blotto A solution (1X TBS, 5% milk and 0.05% Tween-20) prior to immunodetection. Antibodies (see **Table S6**) were diluted in Blotto A and membranes were blotted for 1h at room temperature. 3 washes of 5min in TBS-0.05% Tween-20 were done before the incubation with HRP-coupled secondary antibodies (see **Table S6**). Membranes were washed 3 times for 10-15min and assayed by enhanced chemiluminescence (ECL) detection with Western Lightning Plus-ECL (PerkinElmer) according to the manufacturer protocol.

Apoptosis assays

Cells were seeded at 2×10^5 cells/mL in RPMI 1640 10% FBS. Apoptosis was induced by treating cells for 20h with either hydrogen peroxide (PRXD; Sigma), camptothecin (CPT; Tocris Bioscience) or doxorubicin (DOXO; Sigma) with a concentration set to achieve approximately 70% cell death. Apoptosis was also modulated by pre-treating cells with different molecules prior to the addition of PRXD. With ferric ammonium citrate (FAC; Sigma), cells were treated for 24h and FAC-containing media was changed with fresh medium before apoptosis induction. For deferoxamine mesylate salt (DFO; Sigma), cells were treated for 2h before adding PRXD. Finally, CLIC-specific channel inhibitor IAA-94 (Sigma) was added 10min before PRXD. Apoptosis was assayed by Alexa Fluor 488-coupled Annexin V and propidium iodide (PI) double staining with Dead cell Apoptosis Kit (Thermo Fisher Scientific) according to manufacturer's protocol. 1×10^4 stained cells were analysed by flow cytometry on a LSRFortessa FACS (Becton Dickinson) using blue laser (488nm) with 505LP and 530/30 filters for

Alexa Fluor 488 detection and with 600LP and 610/20 filters for PI detection. Data were acquired with the FACSDiva software (Becton Dickinson). Total apoptosis includes Annexin V⁺/PI⁻ (early apoptotic), Annexin V⁺/PI⁺ (late apoptotic) and Annexin V⁻/PI⁺ (necrotic) cells.

Immunofluorescence microscopy

Reh cells were seeded at 2×10^6 cells/mL in a poly-D-Lysine Hydrobromide (Sigma) treated glass 96-wells plate (Whatman) with or without LysoTracker Red DND-99 staining (Thermo Fisher Scientific) at 50nM final concentration. The plate was centrifuged and incubated for 2h at 37°C, 5% CO₂ to let cells attach to the glass surface. Media was removed and cells were fixed for 10min with freshly made 3.7% formaldehyde in phosphate-buffered saline (PBS). Fixing solution was removed and cells were washed with PTX (0.01% Triton in PBS). Blocking was performed for 1h at room temperature with a 5% Normal Goat Serum (NGS; Jackson ImmunoResearch Laboratories) solution in PTX. Blocking solution was removed and immunostaining was performed overnight with CLIC5A antibody (ab191102 dil. 1:1000; Abcam) and transferrin receptor antibody (ab84036 dil. 1:200; Abcam) in 5% NGS solution. Cells were washed 4 times with PTX before adding secondary antibodies solution. Goat anti-mouse Alexa Fluor 488 (dil. 1:500; Thermo Fisher Scientific) was used to detect anti-CLIC5A and Goat anti-rabbit Alexa Fluor 546 (dil. 1:500; Thermo Fisher Scientific) was used to detect anti-transferrin receptor. Hoechst 33258 DNA stain (dil. 1:500; Thermo Fisher Scientific) was added to the mix to stain nuclei. After 1h incubation with secondary antibodies, cells were washed 4 times with PTX and mounted in

Fluoromount-G (SouthernBiotech). Cells were imaged using a 1.4NA plan-apochromat 100X oil immersion objective on an inverted microscope (DMI6000 B; Leica) fitted with an Ultraview Vox spinning disc confocal system (Perkin Elmer). Images were acquired using an Orca R2 charge-coupled device camera (Hamamatsu) controlled by Volocity 6 software (PerkinElmer/Improvision).

Statistical tests

Significance of observations were assessed using one or two-tailed Fisher's exact test or Mann-Whitney-U test when appropriate.

Ethics statement

The CHU Sainte-Justine Research Ethics Board approved the protocol. Informed consent was obtained from the parents of the patients to participate in this study and for publication of this report and any accompanying images.

2.12.2. Supplemental figures

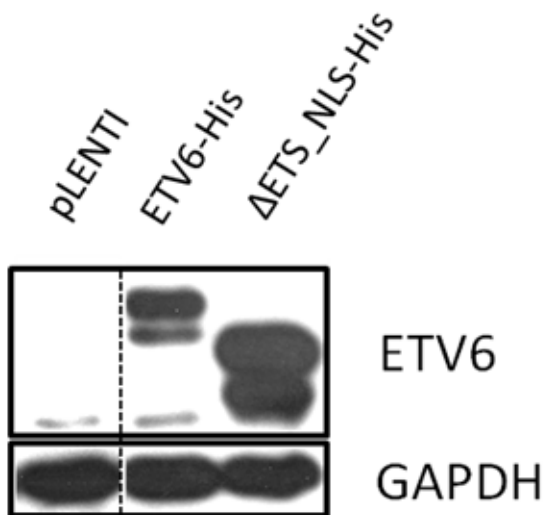


Figure S1: Protein expression levels of ETV6-His and ETV6 Δ ETS_NLS-His in Reh cells. Western blot of nuclear extracts from Reh cells infected with pLENTI empty vector, ETV6-His or ETV6 Δ ETS_NLS-His. Anti-ETV6 detects both ETV6 and ETV6 Δ ETS_NLS forms. *GAPDH* is used as a loading control. The result is shown for one given Reh clone. *Adjustments of brightness and contrast were applied to the whole image.*

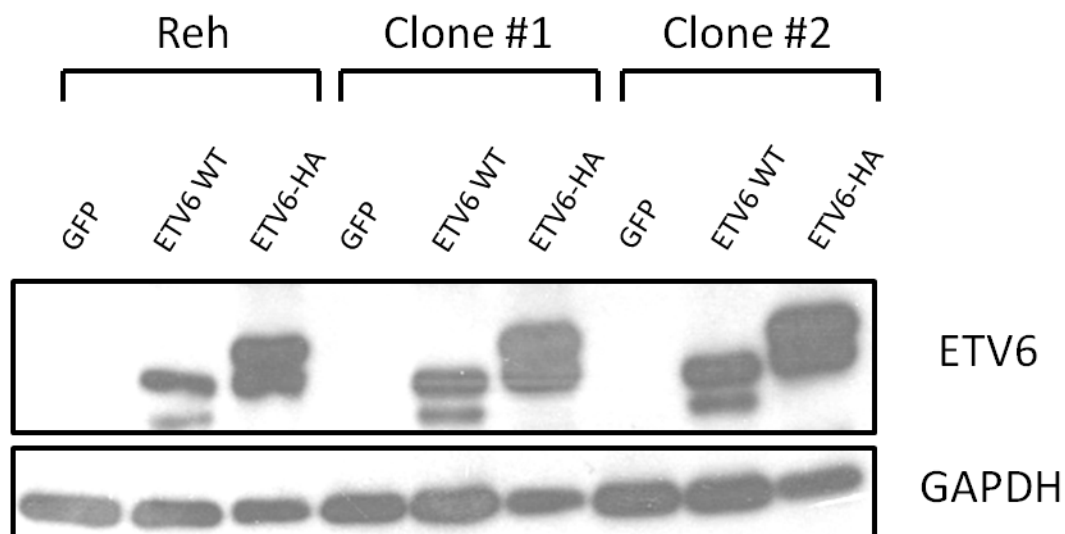


Figure S2: Expression of *ETV6* and *ETV6*-HA proteins in Reh cells. Western blot analysis of nuclear extracts from Reh cells and both Reh clones infected with pCCL GFP control vector, Wild-type *ETV6* (*ETV6* WT) and HA-tagged *ETV6* (*ETV6*-HA). *ETV6* WT and *ETV6*-HA were detected with Anti-*ETV6*. *GAPDH* is used as a loading control. All clones used had similar expression patterns. *Adjustments of brightness and contrast were applied to the whole image.*

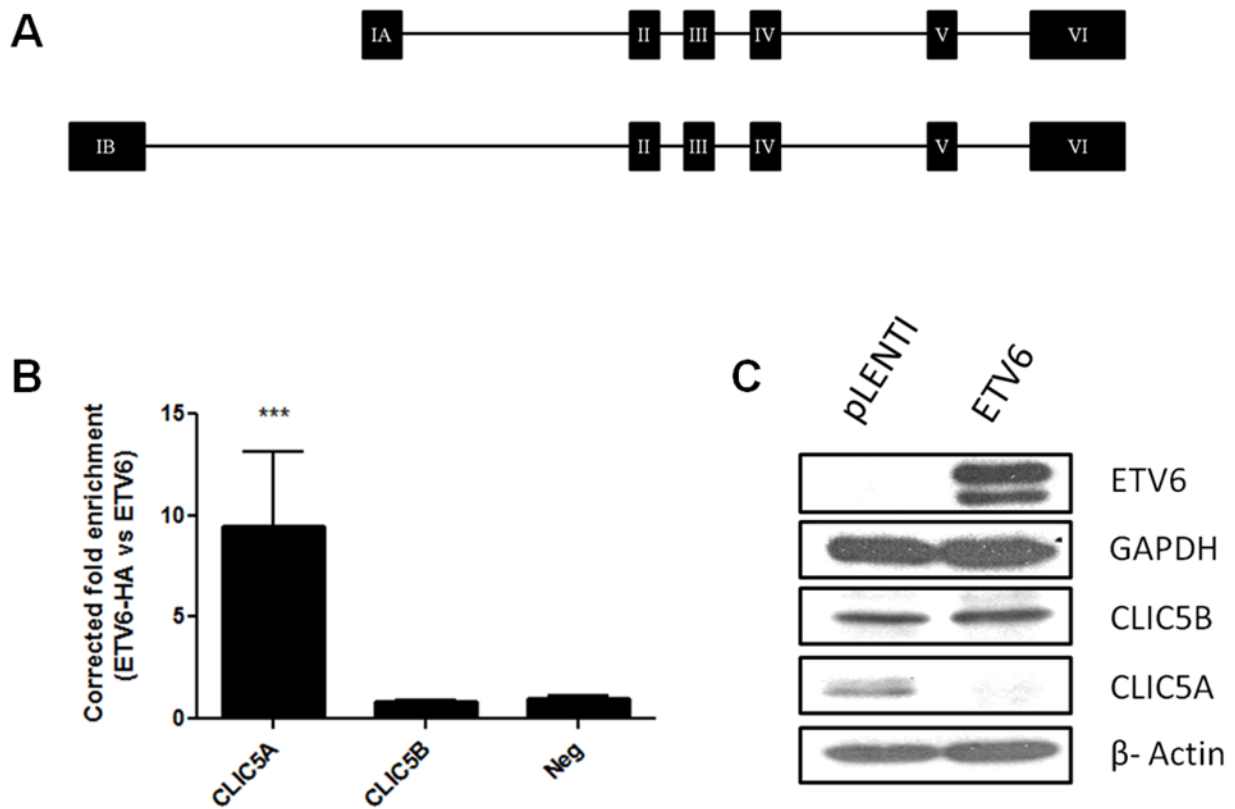


Figure S3: *ETV6* specifically binds and represses the *CLIC5A* isoform. **A.** Representation of *CLIC5A* and *CLIC5B* isoforms at genomic location chr6:45866190-46048085. Protein products encoded by *CLIC5A* and *CLIC5B* contain 251(30 kDa) and 410 (55 kDa) amino acids, respectively. The last 234 amino acids encoded by exons 2 to 6 are identical. **B.** Chromatin immunoprecipitation was performed as previously described and enrichment for *CLIC5A* and *CLIC5B* promoters were calculated. *CLIC5A* proximal promoter is specifically enriched. **C.** Nuclear extracts (for *ETV6* and *GAPDH*) or total extracts (*CLIC5A/B* and β -Actin) from Reh pLenti control cells and Reh *ETV6* cells were analyzed by western blotting. *CLIC5A* but not *CLIC5B* protein level is reduced with *ETV6* overexpression. *Adjustments of brightness and contrast were applied to the whole image.*

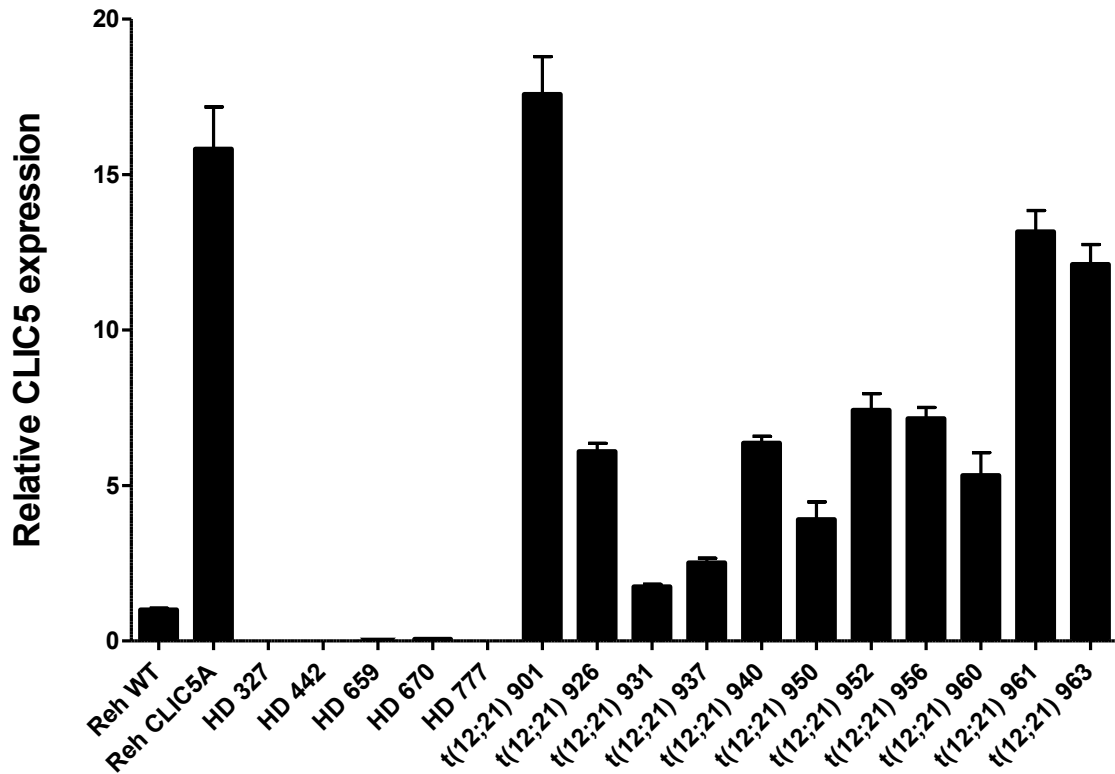


Figure S4: *CLIC5* expression level in *CLIC5A* overexpressing Reh cells and leukemic patients samples. Total RNA was extracted from Reh WT and Reh *CLIC5A* cell lines and additional ALL patients cells for complementary DNA (cDNA) generation. This cDNA was submitted to qRT-PCR analysis to quantify relative expression of *CLIC5*. No or very low expression of *CLIC5* was detected in hyperdiploid (HD) patients samples whereas a clear re-expression is observed in all t(12;21) samples. Additionally, *CLIC5* expression in our Reh *CLIC5A* cells is very similar to what is seen in these patients. Error bars represent the standard deviation (n=3).

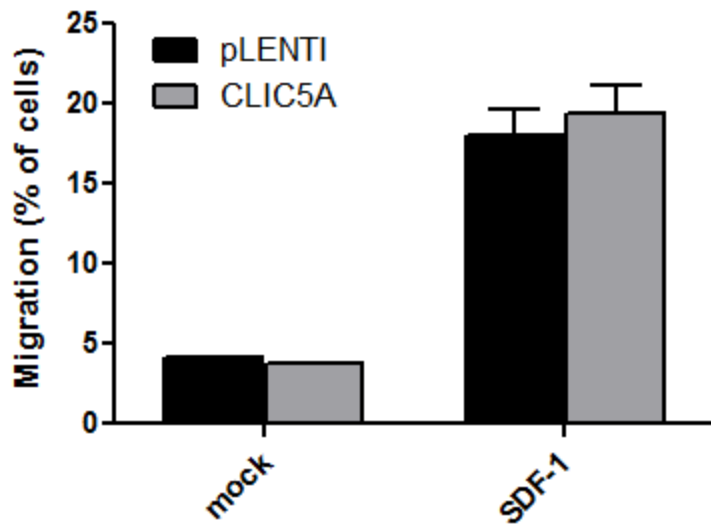


Figure S5: Migration of Reh cells to SDF-1. pLenti control and *CLIC5A* overexpressing Reh cells were used in trans-well migration experiments with the SDF-1 chemokine. 200 000 cells were loaded on a 5 μ m-pore filter plate containing media with or without 10ng/mL SDF-1. Cells that migrated after 90min are present in the bottom chamber and were counted to calculate a migration percentage. SDF-1 induced migration of Reh cells but no difference was observed between *CLIC5A* overexpressing cells and control. This experiment was performed in duplicate for the SDF-1 condition. Error bars represent the standard deviation.

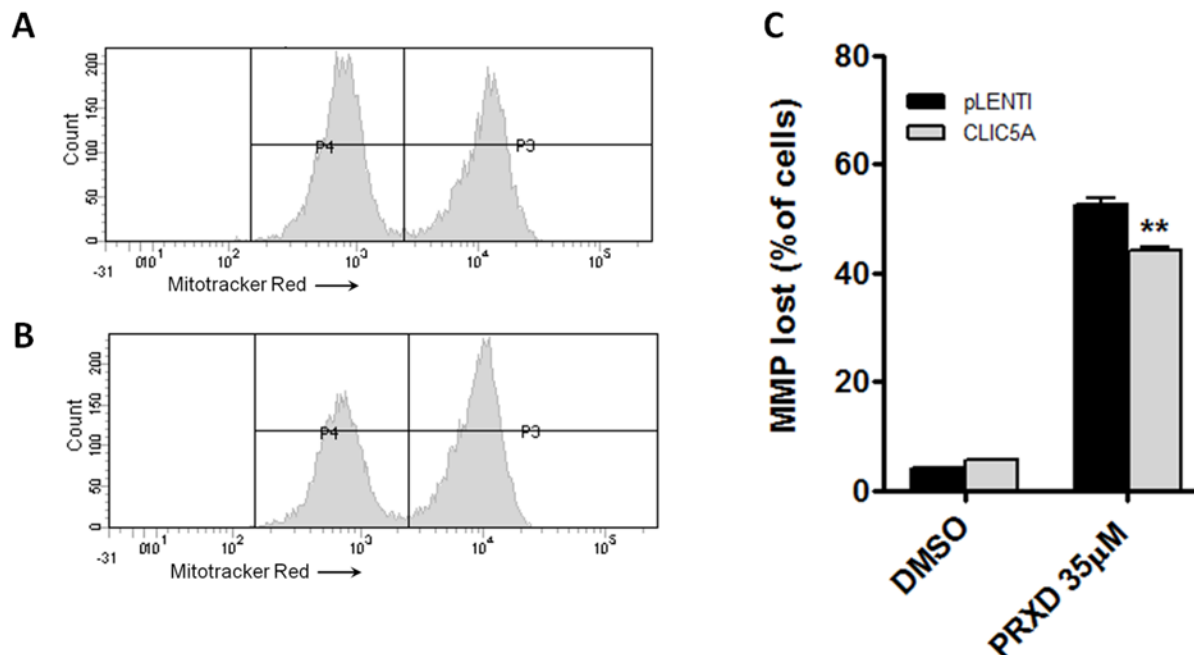


Figure S6: Mitochondrial membrane potential (MMP) in hydrogen peroxide-treated cells. pLenti control and *CLIC5A* overexpressing Reh cells were treated for 20h with 35µM hydrogen peroxide (PRXD). Cells were then stained with Mitotracker Red CMXRos and analysed by flow cytometry using the PI detector. Mitotracker Red CMXRos only stains cells with positive MMP and thus is an interesting marker of mitochondria integrity. **A.** pLenti control cells displayed a higher number of negatively stained cells **B.** compared to *CLIC5A* overexpressing cells, indicating that more mitochondria remain intact to a peroxide challenge with *CLIC5A* overexpression. **C.** This experiment was performed in triplicate. Error bars represent the standard deviation. Statistical significance is calculated by two-tailed Mann-Whitney-U test.

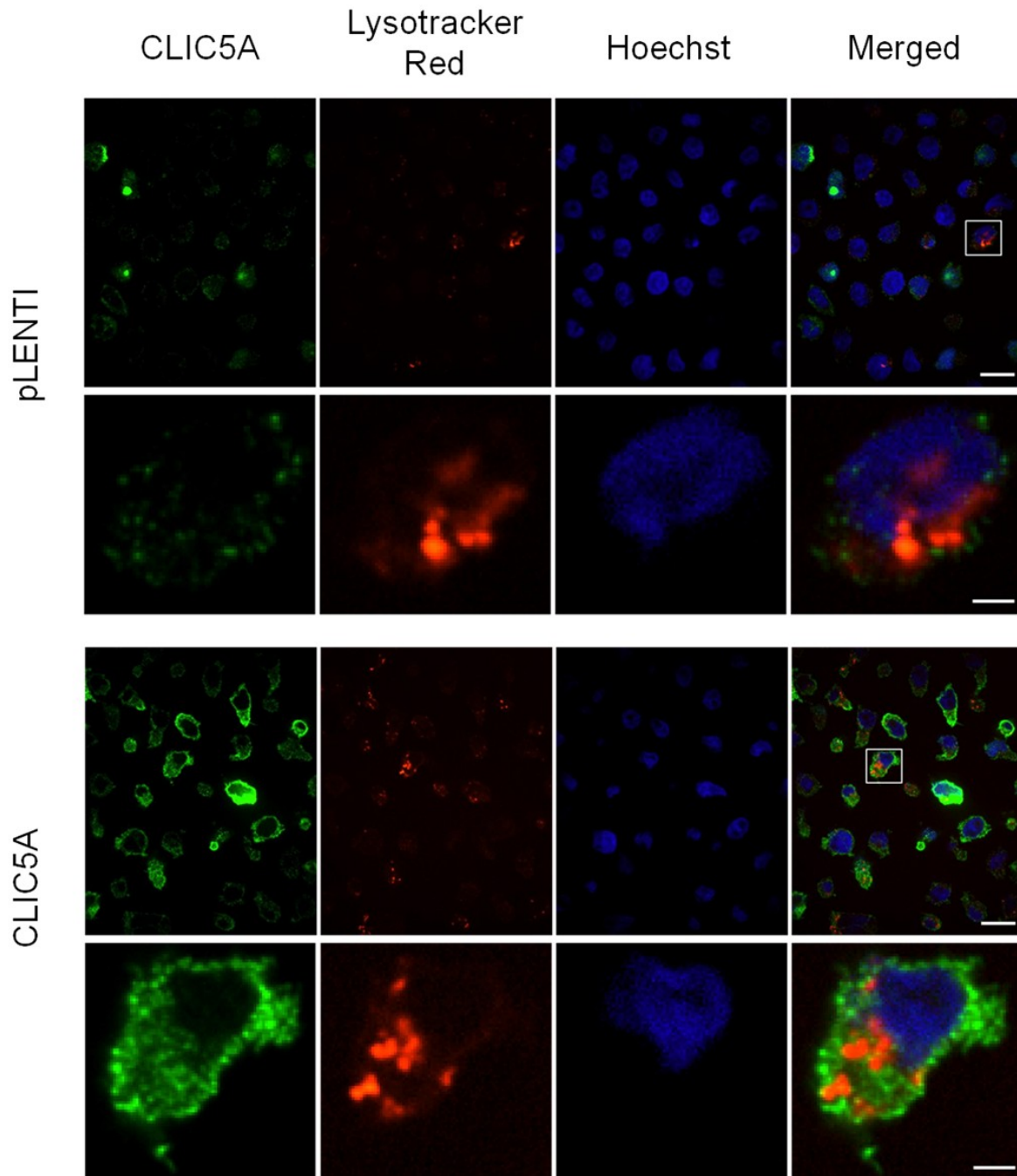


Figure S7: Localization of *CLIC5A* and lysosomes. pLENTI control and *CLIC5A* overexpressing Reh cells were stained with Lysotracker Red DND-99 dye prior to fixation and *CLIC5A* immunostaining. Hoechst DNA staining was also done. Results were obtained at 100X magnification (upper panels; scale bar = 10 μ m). No co-localization was observed between *CLIC5A* and lysosomes. Additional enlargement for the marked region of the initial image is presented in lower panels (scale bar = 2 μ m). A merged image was generated (right panels).

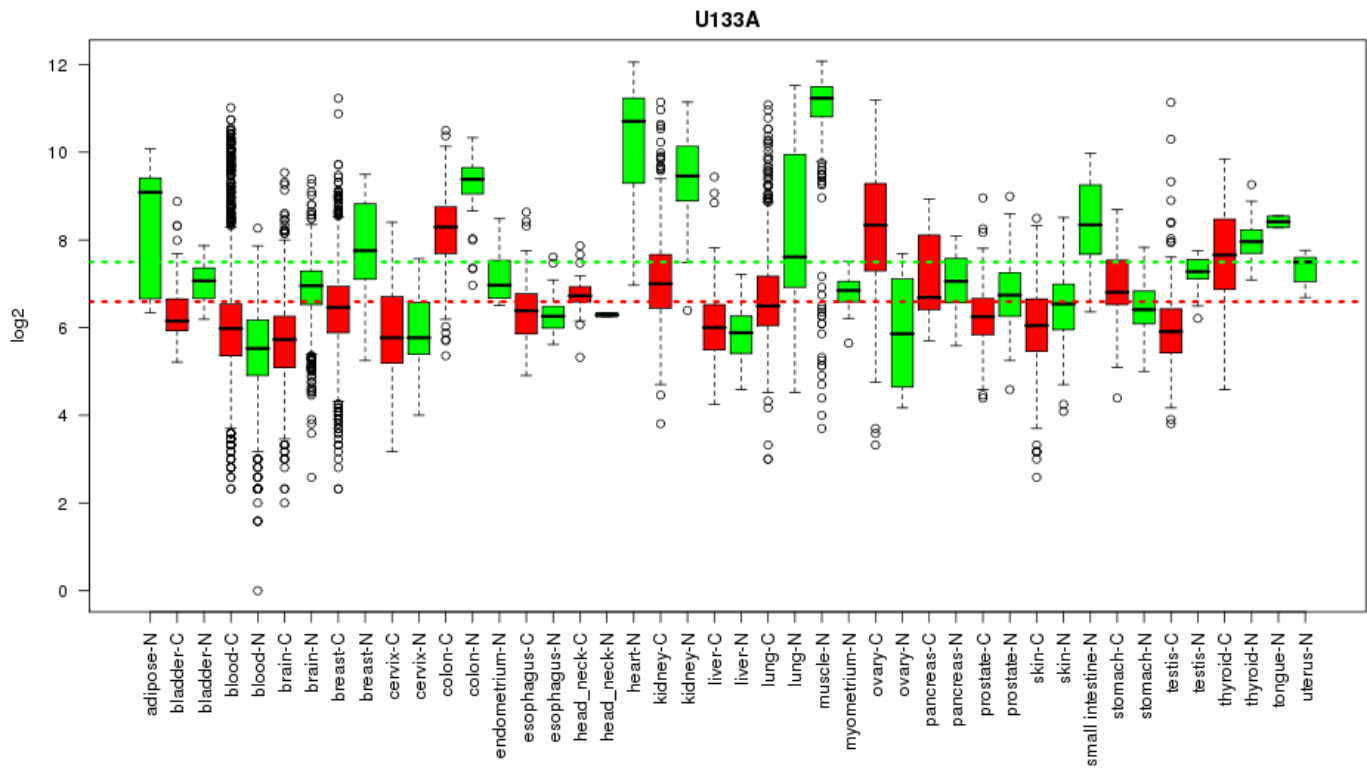


Figure S8: Expression profile of *CLIC5* in normal and matched tumor tissue pairs. Affymetrix expression data (U133A array) for *CLIC5* was obtained from the GENT database. Box plots of expression levels for normal (N; in green) and tumor (C; in red) tissues are shown. Notably, ovarian cancers overexpress *CLIC5* compared their normal counterpart.

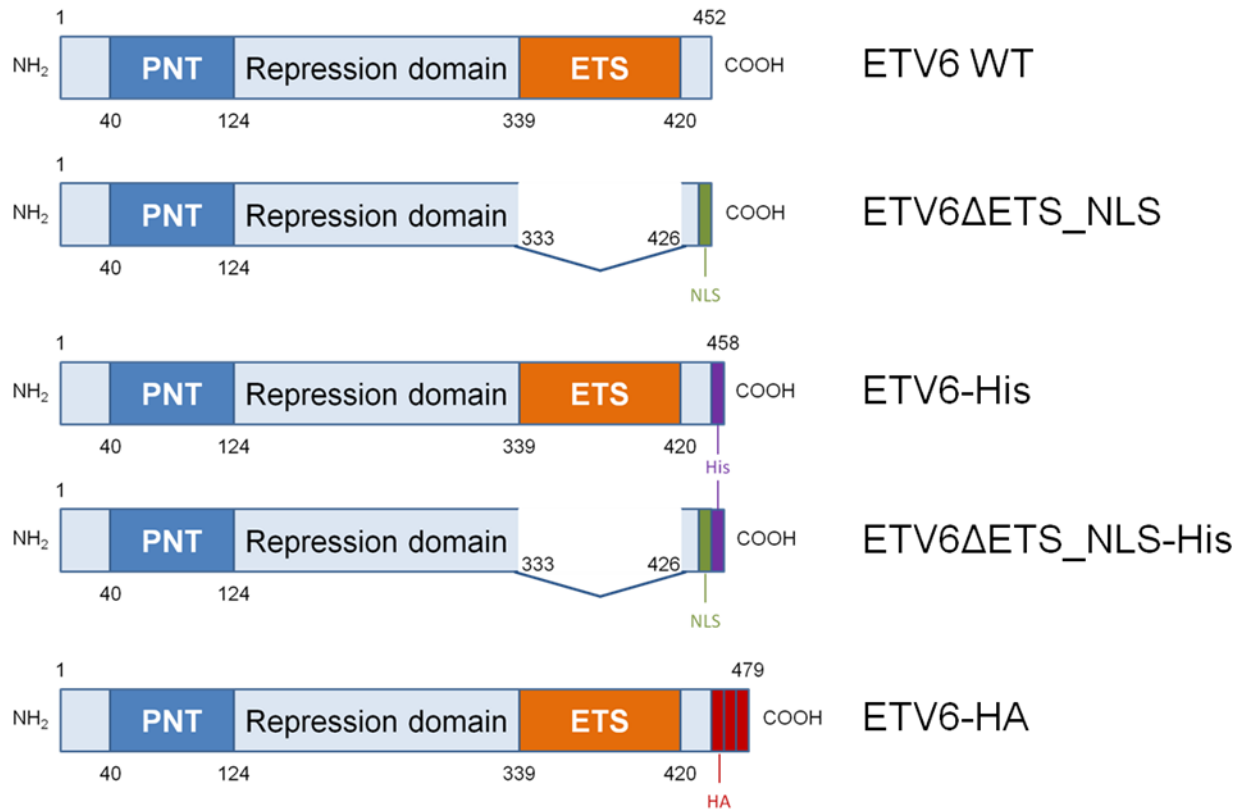


Figure S9: *ETV6* constructs. Positions of *ETV6* domains and deletions are indicated. Tags that were added at the C-terminus of the proteins are also indicated. Nuclear localization signal (NLS) in green, Poly-Histidine tag (His) in purple and hemagglutinin tag (HA) in red.

2.12.3. Supplemental tables

Table S1: Oligonucleotide sequences used for quantitative real-time PCR validation

Oligo Name	Sequence (5' to 3')
CLIC5_F	TGGCTGACTGCAATCTGTTG
CLIC5_R	CATCTCAGCCGGGATATCATAG
BIRC7_F	CTTCTTCCACACAGGCCATCA
BIRC7_R	GGAAGTACAGCTGGGGAA
ANGPTL2_F	GGTACTCCTTTACATGGCAC
ANGPTL2_R	GTTATACCACCAGCCTCCCTT
WBP1L_F	CAATTACTCAGCGCTGCCAT
WBP1L_R	GAGGAGTTGGAGGTCGGTTC
LRR4_F	ACGAAAAGGGTACAGGCTCA
LRR4_R	TGCATGCGTGTCACCTTGTA
SLC51A_F	GGGCATCATTTCCCGTCAAG
SLC51A_R	CGTTGGCCAAGACTGAGAAG
GAPDH_F	GATCATCAGCAATGCCTC
GAPDH_R	TGTGGTCATGAGTCCTTCCA

Table S2: Oligonucleotide sequences used for chromatin immunoprecipitation qPCR

Oligo Name	Sequence (5' to 3')
ChIP_CLIC5_1_F	AGTTGGAGGAGGGAGCTCTA
ChIP_CLIC5_1_R	CCTCATTGGCAGTAACTCCA
ChIP_CLIC5_2_F	CCAGCACCCCTACTCTCCTTT
ChIP_CLIC5_2_R	CTCCACTACCAGCCTTACCC
ChIP_BIRC7_F	GAGCCATCATCTCCAGCTCC
ChIP_BIRC7_R	TCTCTCTGAACAAGACGGGAA
ChIP_ANGPTL2_F	CTGCACGCACAGCTGGAG
ChIP_ANGPTL2_R	CCCTCTCCCCAAAGCTCAGG
ChIP_WBP1L_F	CTTCCGCCACACTCAAGATG
ChIP_WBP1L_R	TAAAATTCGTCCCAACCCGC
ChIP_SLC51A_F	CAAGACCTGTCAAATGCCCA
ChIP_SLC51A_R	TTCCCAGACCCCAGTTCATC
Neg_F	ATGGTTGCCACTGGGGATCT
Neg_R	TGCCAAAGCCTAGGGGAAGA

Table S3: Downregulated transcripts in ETV6-His expressing Reh cells vs pLENTI cells

Ensembl ID	Gene symbol	logFC	logCPM	PValue	FDR
ENSG00000112782	CLIC5	-3,20	3,75	6,40E-39	8,09E-35
ENSG00000111816	FRK	-3,68	3,31	2,88E-30	1,82E-26
ENSG00000171403	KRT9	-5,94	-0,10	2,76E-16	1,16E-12
ENSG00000153993	SEMA3D	-2,54	2,70	3,18E-15	1,01E-11
ENSG00000218574	HNRNPA1P37	-5,11	0,15	6,22E-15	1,57E-11
ENSG00000136859	ANGPTL2	-1,68	4,37	3,34E-14	6,41E-11
ENSG00000168427	KLHL30	-3,34	1,35	2,79E-13	4,41E-10
ENSG00000030419	IKZF2	-1,78	6,64	6,03E-13	8,47E-10
ENSG00000178425	NT5DC1	-1,23	4,98	3,68E-12	4,23E-09
ENSG00000164920	OSR2	-2,79	2,25	4,94E-12	5,21E-09
ENSG00000145358	DDIT4L	-1,54	4,35	5,71E-12	5,55E-09
ENSG00000151012	SLC7A11	-1,62	5,51	1,01E-09	6,71E-07
ENSG00000166272	WBP1L	-1,10	5,18	1,06E-09	6,71E-07
ENSG00000123700	KCNJ2	-2,14	2,16	2,66E-09	1,60E-06
ENSG00000102471	NDFIP2	-3,60	-0,01	6,40E-09	3,68E-06
ENSG00000229164	TRAC	-2,12	2,03	1,05E-08	5,79E-06
ENSG00000173805	HAP1	-1,22	4,19	1,79E-08	9,45E-06
ENSG00000230138		-2,14	1,87	2,97E-08	1,50E-05
ENSG00000089902	RCOR1	-0,88	6,82	3,61E-08	1,75E-05
ENSG00000024422	EHD2	-2,48	1,69	6,89E-08	3,11E-05
ENSG00000177508	IRX3	-1,88	2,79	8,10E-08	3,53E-05
ENSG00000198400	NTRK1	-2,29	1,89	1,25E-07	5,29E-05
ENSG00000104888	SLC17A7	-2,20	1,16	3,76E-07	1,49E-04
ENSG00000154258	ABCA9	-1,37	4,95	5,89E-07	2,26E-04
ENSG00000165029	ABCA1	-1,50	4,44	8,21E-07	3,05E-04
ENSG00000224957		-5,90	-1,28	2,41E-06	8,48E-04
ENSG00000075213	SEMA3A	-2,49	1,94	2,85E-06	9,47E-04
ENSG00000102445	KIAA0226L	-1,53	2,44	3,27E-06	1,03E-03
ENSG00000157657	ZNF618	-0,96	4,87	3,54E-06	1,09E-03
ENSG00000128594	LRRC4	-1,16	3,23	3,99E-06	1,20E-03
ENSG00000211821	TRDV2	-2,38	0,68	4,51E-06	1,33E-03
ENSG00000173898	SPTBN2	-1,43	2,88	5,57E-06	1,60E-03
ENSG00000145147	SLIT2	-1,27	3,22	6,08E-06	1,67E-03
ENSG00000090104	RGS1	-3,18	1,03	6,67E-06	1,76E-03
ENSG00000050405	LIMA1	-0,79	4,95	8,28E-06	2,14E-03
ENSG00000100439	ABHD4	-1,24	3,08	8,82E-06	2,23E-03
ENSG00000176659	C20orf197	-2,71	0,00	1,16E-05	2,77E-03

ENSG00000185433	LINC00158	-2,30	1,20	1,21E-05	2,84E-03
ENSG00000163959	SLC51A	-1,50	2,43	1,45E-05	3,28E-03
ENSG00000081237	PTPRC	-0,73	7,67	1,51E-05	3,35E-03
ENSG00000163738	MTHFD2L	-0,78	5,87	1,69E-05	3,69E-03
ENSG00000130477	UNC13A	-1,99	2,30	2,08E-05	4,39E-03
ENSG00000101197	BIRC7	-1,39	3,02	2,30E-05	4,76E-03
ENSG00000169855	ROBO1	-3,94	2,65	2,51E-05	5,12E-03
ENSG00000141052	MYOCD	-0,83	4,51	3,71E-05	7,33E-03
ENSG00000228277		-3,29	-1,36	3,81E-05	7,40E-03
ENSG00000115232	ITGA4	-0,64	7,74	3,98E-05	7,62E-03
ENSG00000105851	PIK3CG	-0,72	5,31	5,02E-05	9,33E-03
ENSG00000169429	IL8	-2,08	0,77	5,60E-05	1,01E-02
ENSG00000112972	HMGCS1	-0,69	5,50	5,81E-05	1,02E-02
ENSG00000152580	IGSF10	-1,15	7,37	6,50E-05	1,11E-02
ENSG00000176887	SOX11	-0,97	4,93	6,91E-05	1,14E-02
ENSG00000132549	VPS13B	-0,83	5,95	6,79E-05	1,14E-02
ENSG00000076864	RAP1GAP	-1,98	0,65	7,28E-05	1,18E-02
ENSG00000136158	SPRY2	-0,99	3,40	8,28E-05	1,33E-02
ENSG00000120669	SOHLH2	-4,59	0,16	1,00E-04	1,59E-02
ENSG00000137193	PIM1	-0,75	4,87	1,08E-04	1,68E-02
ENSG00000136167	LCP1	-0,59	8,50	1,15E-04	1,75E-02
ENSG00000189325	C6orf222	-1,91	1,22	1,23E-04	1,81E-02
ENSG00000101752	MIB1	-0,59	8,81	1,25E-04	1,81E-02
ENSG00000177301	KCNA2	-2,00	0,37	1,52E-04	2,17E-02
ENSG00000027697	IFNGR1	-0,62	5,47	1,69E-04	2,35E-02
ENSG00000137441	FGFBP2	-2,47	-1,13	1,89E-04	2,58E-02
ENSG00000152413	HOMER1	-0,79	4,40	1,97E-04	2,65E-02
ENSG00000118308	LRMP	-0,70	6,27	2,06E-04	2,74E-02
ENSG00000058091	CDK14	-0,73	5,64	2,27E-04	2,98E-02
ENSG00000196954	CASP4	-2,00	0,35	2,84E-04	3,59E-02
ENSG00000235492		-1,92	1,08	2,87E-04	3,59E-02
ENSG00000007237	GAS7	-0,66	4,95	2,98E-04	3,63E-02
ENSG00000198053	SIRPA	-2,46	0,40	3,15E-04	3,76E-02
ENSG00000112245	PTP4A1	-0,76	5,70	3,48E-04	4,04E-02
ENSG00000169252	ADRB2	-1,96	-0,27	3,71E-04	4,23E-02
ENSG00000105409	ATP1A3	-1,11	4,19	3,97E-04	4,48E-02
ENSG00000078142	PIK3C3	-0,70	6,99	4,58E-04	5,08E-02
ENSG00000151276	MAGI1	-3,79	-0,62	4,75E-04	5,23E-02
ENSG00000183496	MEX3B	-1,23	2,04	5,00E-04	5,43E-02
ENSG00000153822	KCNJ16	-0,86	3,29	5,15E-04	5,52E-02
ENSG00000178752	FAM132B	-1,24	2,31	5,41E-04	5,75E-02

ENSG00000011105	TSPAN9	-0,89	3,30	5,91E-04	6,08E-02
ENSG00000080503	SMARCA2	-0,54	6,71	6,47E-04	6,54E-02
ENSG00000064666	CNN2	-1,00	4,18	6,71E-04	6,68E-02
ENSG00000173559	NABP1	-0,71	4,53	6,82E-04	6,70E-02
ENSG00000158201	ABHD3	-0,55	6,40	6,83E-04	6,70E-02
ENSG00000167094	TTC16	-1,29	1,98	7,62E-04	7,33E-02
ENSG00000188452	CERKL	-0,54	5,49	7,65E-04	7,33E-02
ENSG00000242265	PEG10	-0,62	6,06	7,89E-04	7,47E-02
ENSG00000160255	ITGB2	-1,43	2,02	8,43E-04	7,84E-02
ENSG00000241151		-2,22	0,00	9,10E-04	8,28E-02
ENSG00000234883	MIR155HG	-0,99	2,83	9,10E-04	8,28E-02
ENSG00000109861	CTSC	-0,61	5,47	9,36E-04	8,46E-02
ENSG00000152104	PTPN14	-1,77	0,44	1,03E-03	9,05E-02
ENSG00000100097	LGALS1	-0,92	2,88	1,06E-03	9,20E-02
ENSG00000134762	DSC3	-0,89	6,98	1,14E-03	9,80E-02

logFC: log fold change; logCPM: log counts per million reads; FDR: false discovery rate.

Table S4: Downregulated transcripts in ETV6-His expressing Reh cells vs ETV6ΔETS_NLS-His cells

Ensembl ID	Gene symbol	logFC	logCPM	PValue	FDR
ENSG00000112782	CLIC5	-3,23	4,66	7,00E-52	9,59E-48
ENSG00000101197	BIRC7	-1,93	4,02	3,44E-11	1,45E-07
ENSG00000145358	DDIT4L	-1,35	5,04	3,84E-11	1,45E-07
ENSG00000136859	ANGPTL2	-1,31	5,00	4,24E-11	1,45E-07
ENSG00000218574	HNRNPA1P37	-3,99	0,66	6,67E-11	1,83E-07
ENSG00000139083	ETV6	-1,66	4,49	4,27E-09	8,36E-06
ENSG00000111816	FRK	-1,78	3,60	1,20E-08	2,05E-05
ENSG00000102471	NDFIP2	-3,05	0,66	1,08E-07	1,65E-04
ENSG00000230138		-1,66	2,49	7,65E-06	9,18E-03
ENSG00000166272	WBP1L	-0,79	5,78	8,04E-06	9,18E-03
ENSG00000153993	SEMA3D	-1,35	3,12	9,81E-06	1,03E-02
ENSG00000101198	NKAIN4	-1,85	1,68	5,42E-05	4,95E-02
ENSG00000099365	STX1B	-2,08	-0,11	5,82E-05	4,99E-02
ENSG00000250366	LINC00617	-2,10	-0,69	7,42E-05	5,64E-02
ENSG00000128594	LRRC4	-1,02	3,90	9,18E-05	6,29E-02
ENSG00000178425	NT5DC1	-0,62	5,50	8,74E-05	6,29E-02
ENSG00000024422	EHD2	-1,66	2,22	1,06E-04	6,60E-02
ENSG00000163959	SLC51A	-1,31	3,11	1,14E-04	6,79E-02
ENSG00000099250	NRP1	-1,48	1,77	1,58E-04	9,04E-02

logFC: log fold change; logCPM: log counts per million reads; FDR: false discovery rate.

Table S5: Oligonucleotide sequences used for cloning

Oligo Name	Sequence (5' to 3')
ETV6_F	GGGAATTCGGCACGAGGAAACT
ETV6His_R	CCGCGGCCGCTCAATGATGATGATGATGATGGCATTATCTTCT TGGTATATTTGTT
ETV6deltaETS_NLS_F	CCGGATCCA ATTCGGCACGAGGAAACT
ETS_NLSHis_R	CCGCGGCCGCTCAATGATGATGATGATGATGCACCTTCCTCTTC TTCTTCGGTTG
Oligo_HA ₃	GAATGCTTAT ACCCATACGATGTTCCAGATTACGCTTACCCATAC GATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGC TTGAGCGGCCG

Table S6: Antibodies used in western blotting

<i>Primary antibodies</i>	Host	Antibody # (Provider)	Working dilution
ETV6	mouse	ab54705 (Abcam)	1:1000
GAPDH	goat	sc-31915 (Santa Cruz)	1:1000
CLIC5	rabbit	ab75948 (Abcam)	1:100
Beta-Actin	mouse	sc-81178 (Santa Cruz)	1:1000
<i>Secondary antibodies (HRP-coupled)</i>			
Anti-mouse	rabbit	sc-358914 (Santa Cruz)	1:5000
Anti-goat	chicken	sc-2961 (Santa Cruz)	1:5000
Anti-rabbit	donkey	sc-2077 (Santa Cruz)	1:5000

2.13 Reprint permission

Rights & Permissions

All articles published in Haematologica are protected by copyright. The Ferrata Storti Foundation (owner of Haematologica) makes available all contents on the following terms:

- You are entitled to access and use the Haematologica Journal website for your personal research purposes.
- You can download or print individual articles from Haematologica.
- Any use of this material in whole or in part should include bibliographic citation, including author attribution, date, article title, and a statement as follows: Obtained from the Haematologica Journal website <http://www.haematologica.org>

Tiré de: <http://www.haematologica.org/terms-conditions>

©2016 Ferrata Storti Foundation

Material published in Haematologica is covered by copyright. All rights reserved to the Ferrata Storti Foundation. Copies of articles are allowed for personal or internal use. Permission in writing from the publisher is required for any other use.

OPEN  ACCESS

Chapitre 3: Deuxième article

La liaison d'ETV6 à l'ADN

3.1 Préface

Le recrutement des facteurs de transcription à l'ADN est une facette importante de leur régulation et de leur fonction. Les caractéristiques des régions préférentiellement liées par ces facteurs, comme la localisation génomique et les séquences consensus, peuvent aider à élaborer les mécanismes sous-jacents à leur activité transcriptionnelle. Bien que notre première étude (Chapitre 2: Article 1) ait démontré la présence d'*ETV6* aux promoteurs de certains gènes par une approche ciblée, le recrutement d'*ETV6* à l'ADN n'avait pas été examiné plus en détail. Pour être en mesure de mieux comprendre l'impact d'*ETV6* sur la transcription, il était important de déterminer son profil de liaison.

En utilisant une approche ChIP-seq, nous avons identifié les régions du génome où *ETV6* était localisé. Avec les données de transcriptomes, nous avons corrélié la présence d'*ETV6* à des changements au niveau de l'expression des gènes, corroborant son impact transcriptionnel. L'étude en profondeur des régions liées par *ETV6* a confirmé sa dépendance pour les séquences EBS en plus d'avoir mis en évidence son rôle important de cofacteur, notamment sur la fonction de la fusion *ETV6-RUNX1* et des facteurs IRF. L'intégration de ces données a ainsi permis de peaufiner le réseau de régulation transcriptionnel d'*ETV6* dans des cellules leucémiques.

Genome wide mapping of *ETV6* binding sites in pre-B leukemic cells

Benjamin Neveu^{1,2}, Maxime Caron¹, Karine Lagacé^{1,2}, Chantal Richer¹ and Daniel Sinnett^{1,2,3,*}

- 1) Sainte-Justine UHC Research Center, Montreal, Qc, Canada;
- 2) Department of Biochemistry and Molecular Medicine, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada.
- 3) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada.

*Corresponding author: Daniel Sinnett, Division of Hematology-Oncology, Sainte-Justine UHC Research Center, 3175 Côte Ste-Catherine, Montréal (Québec) Canada, H3T 1C5; Tel: (514) 345-4931/ext. 2990; Email: daniel.sinnett@umontreal.ca

Cet article a été soumis à Scientific Reports le 14 février 2018.

3.2 Author contributions

DS is the principal investigator and takes primary responsibility for the paper. BN, MC and DS contributed to the conception and design of the study. KL performed ChIP experiments. CR was involved in sample and library preparation. MC performed bioinformatics analyses and data processing. BN contributed to data integration and analyses. BN drafted the paper and DS contributed to interpretation of the data and was involved in critical revision of the manuscript. All authors approved the final version.

3.3 Abstract

Genetic alterations in the transcriptional repressor *ETV6* are associated with hematological malignancies. Notably, the t(12;21) translocation leading to an *ETV6*-*AML1* fusion gene is the most common genetic alteration found in childhood acute lymphoblastic leukemia. Moreover, most of these patients also lack *ETV6* expression, suggesting a tumor suppressor function. To gain insights on *ETV6* DNA-binding specificity and genome wide transcriptional regulation capacities, we performed chromatin immunoprecipitation experiments coupled to deep sequencing in a t(12;21)-positive pre-B leukemic cell line. This strategy led to the identification of *ETV6*-bound regions that were further associated to gene expression. *ETV6* binding is mostly cell type-specific as only few regions are shared with other blood cell subtypes. Peaks localization and motif enrichment analyses revealed that this unique binding profile could be associated with the *ETV6*-*AML1* fusion protein specific to the t(12;21) background. This study underscores the highly regulated mechanisms of *ETV6* binding and uncovers *ETV6* transcriptional network in pre-B leukemia cells bearing the recurrent t(12;21) translocation.

3.4 Introduction

ETV6 is a member of the ETS superfamily of transcription factors that are critical modulators of cellular homeostasis in several tissues. The normal function of ETS factors is mandatory for appropriate cell fate as dysregulation or deleterious events affecting these factors are frequently observed in a variety of cancers [1]. *ETV6* is essential in the establishment and maintenance of hematopoiesis within the bone marrow compartment [2, 3]. *ETV6* translocations are frequently observed in various hematological disorders [4] and germline mutations have been associated to predispositions for such diseases [5-10].

The most common *ETV6* aberration is the t(12;21)(p13;q22) translocation which fuses *ETV6* to the *AML1* gene (or *RUNX1*) and generates an in-frame *ETV6*-*AML1* chimeric protein [11]. This is the most frequent chromosomal abnormality in childhood pre-B cell acute lymphoblastic leukemia (pre-B ALL), occurring in 20% of cases [12]. However, the *ETV6*-*AML1* fusion protein seems insufficient to induce leukemia by itself [13-15], suggesting that additional events are required to fully develop pre-B ALL [16]. Interestingly, the complete inactivation of *ETV6* in t(12;21)-positive pre-B ALL cases was underscored by several studies [17-21] and indicates that *ETV6* depletion could lead to pre-B ALL initiation.

Unlike the majority of ETS members, *ETV6* acts as a transcriptional repressor [22, 23]. *ETV6* has a N-terminal pointed (PNT) helix-loop-helix domain required for protein-protein interactions and homodimerization [24]. Its central repressive domain is also implicated in protein-protein interactions with members of the

SMRT/N-CoR/mSin3A/HDAC co-repressor complexes [25-27]. The C-terminal part of *ETV6* contains an ETS DNA-binding domain that recognizes a consensus ETS-binding site consisting of a core GGAA/T sequence with adjacent purine-rich sequences [24]. Interestingly, the ETV6-AML1 fusion protein combines both PNT and central repressive domains of *ETV6* with Runt DNA binding and transactivation domains of *AML1*, thus converting *AML1* from a transcriptional activator to a putative repressor [28].

Although the molecular functions of ETV6-AML1 have been studied [29-31], the exact role of *ETV6* remains poorly understood. To gain insights into *ETV6* function in t(12;21)-positive pre-B leukemia cells, we sought to identify *ETV6* binding sites using chromatin immunoprecipitation coupled to high throughput sequencing (ChIP-seq). By including expression data [32], we extensively described *ETV6* binding properties and transcriptional activity in this particular context.

3.5 Methods

3.5.1. Constructs

The complete wild-type coding sequence of *ETV6* was subcloned into pcDNA3.1 (pcDNA3.1 ETV6). The C-terminal HA-tagged ETV6 construct was generated by restriction enzyme digestion as described previously [32] using an oligomer containing 3 tandem HA tag repeats. Both ETV6 and ETV6-HA were subcloned into pCCL lentiviral vector (kindly provided by Dr. Christian Beauséjour) through enzymatic digestion and ligation.

3.5.2. Cell culture

Reh (ATCC ® CRL-8286™), a t(12;21)-positive pre-B ALL cell line, was maintained in RPMI 1640 (Wisent) 10% Fetal Bovine Serum (FBS; Wisent) in a 5% CO₂ incubator at 37°C.

3.5.3. Lentiviral production

1.5x10⁷ HEK293T cells were seeded into 15cm petri dishes in DMEM (Wisent) 10% FBS. The next day, cells were transfected with 9µg pCCL plasmids together with 6µg pRSV-Rev, 7.8µg pMD2.VSVG and 15µg pMDL third generation encapsidation plasmids (kindly provided by Dr. Christian Beauséjour) in fresh RPMI 1640 10% FBS medium using polyethylenimine (Polysciences) at a final concentration of 6.5µg/mL. Media was removed 16h post-transfection and replaced by fresh DMEM 10% FBS. After 30h, viral particles were retrieved from media by ultracentrifugation (50 000g) and quantified by p24 antigen ELISA (Advanced Bioscience Laboratories).

3.5.4. Lentiviral infection

2×10^6 Reh cells and two different Reh clones (generated in methylcellulose media) were seeded in 2mL of RPMI 1640 10% FBS medium. 200ng of concentrated virus were added to cells with polybrene (Sigma) to a final concentration of $8 \mu\text{g}/\text{mL}$. 24h post-infection, medium was changed with fresh RPMI 1640 10% FBS. These cells were maintained 2 weeks in culture before carrying out further experiments.

3.5.5. Western blotting

$20 \mu\text{g}$ of nuclear protein extracts were diluted in Laemmli buffer and migrated on SDS-denaturing 10% polyacrylamide gels. Transfer on polyvinylidene difluoride membranes was performed at 4°C overnight. Membranes were blocked in Blotto A solution (1X TBS, 5% milk and 0.05% Tween-20) prior to immunoblotting using the primary antibodies against ETV6 (1:1000; ab54705; Abcam) or GAPDH (1:1000; sc-31915; Santa Cruz) and HRP-coupled secondary antibodies anti-mouse (1:5000; sc-358914; Santa Cruz) and anti-goat (1:5000; sc-2961; Santa Cruz) IgG, respectively. Membranes were then assayed by enhanced chemiluminescence detection with Western Lightning Plus-ECL (PerkinElmer) according to the manufacturer protocol.

3.5.6. Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) has been performed as previously described [32]. Briefly, cross-linked chromatin isolated from 1.0×10^7 transduced Reh cells was used for immunoprecipitation with anti-HA magnetic beads

(Thermo Fisher Scientific). DNA-protein complexes were eluted from the beads by competition with HA peptides prior to reverse-crosslinking and standard purification using phenol/chloroform/isoamyl alcohol (Sigma). The purified ChIP DNA was processed through TruSeq ChIP Sample Preparation Kit (Illumina) according to the manufacturer protocol. As positive control, a fraction of the amplified ChIP material was used to assess the *ETV6* binding enrichment at the *CLIC5A* promoter [32] by quantitative PCR (qPCR) using primers listed in [Supplementary Table S1](#). Libraries were sequenced on the HiSeq 2500 system (Illumina) in paired-end mode (2 x 100bp).

3.5.7. ChIP-seq data analysis

Raw reads were aligned using bowtie v2.2.3 [33] and filtered using a mapping quality threshold of 20. Reads were merged for both ETV6 and ETV6-HA populations (n=3). Peaks were called in the ETV6-HA condition over the background obtained in the ETV6 condition (negative IP control) using MACS v2.1.1.20160309 [34] with a q-value threshold of 0.1. Peaks overlapping blacklisted regions (ENCODE's EncodeDacMapabilityConsensus [35], svelter's exclude file [36] and canva's filter file [37]) or having a fold enrichment above 50 were discarded. Peaks were associated to ensembl genes using the closest gene transcription start site (TSS) annotations from HOMER and genomic coordinates from bedops' closest-features v2.4.19 [38]. To define the minimal peak fold enrichment threshold (≥ 4.5), a Fisher's exact test p-value was calculated for the overlaps between *ETV6* peaks and *ETV6*-modulated genes [32] against the proportion of *ETV6* peaks in the complete ensembl genes

repertoire. Genomic annotations of peaks and motif enrichment analyses were performed using HOMER v4.8 [39].

ETV6 ChIP-seq in GM12878 and K562 cells are publically available as downloadable bed files as part of the ENCODE project.

GM12878:<https://www.encodeproject.org/files/ENCFF272DJU/@@download/ENCFF272DJU.bed.gz>

K562:<https://www.encodeproject.org/files/ENCFF514SLV/@@download/ENCFF514SLV.bed.gz>

Chromatin state data of GM12878 cells were obtained from the NIH Roadmap Epigenomics Project Portal (E116, 15 coreMarks) [40].

3.6 Results

3.6.1. Identification of genome wide *ETV6*-bound regions

To determine the genomic regions bound by *ETV6*, we expressed *ETV6* and HA-tagged *ETV6* in 3 biological replicates of Reh pre-B leukemic cells (**Fig 1A**). Of note, Reh cells lack endogenous wild-type *ETV6* expression as a result of a t(12;21) translocation and a 12p13 locus deletion. These cells were then used for ChIP experiments using the HA epitope as a bait. We confirmed *ETV6* binding to the *CLIC5A* promoter region, a validated *ETV6* target gene [32], both by qRT-PCR (**Fig 1B**) and high throughput sequencing (**Fig 1C**), indicating that our ChIP-seq data are suitable for the identification of *ETV6*-bound regions (two examples are shown in [Supplementary Fig S1](#)).

The number of *ETV6* peaks called varies markedly with the fold enrichment threshold (**Fig 1D**; [Supplementary Table S2](#)). Effective *ETV6* binding is expected to impact the expression of nearby genes. Accordingly, *ETV6* peaks were associated to genes and compared to a known *ETV6*-modulated genes list obtained by RNA-seq [32]. This strategy was performed through the entire peak fold enrichment range (**Fig 1E**; [Supplementary Table S2](#)). The overlap between *ETV6* peaks and *ETV6*-modulated genes is systematically higher than the random probability given by the proportion of *ETV6* peaks in the complete ensembl genes repertoire. The ratio calculated between these overlaps is directly correlated to the peak fold enrichment (**Fig 1E**, enrichment ratio; [Supplementary Table S2](#)), indicating that stronger *ETV6* peaks are more likely associated with *ETV6*-mediated transcriptional regulation.

A peak fold enrichment ≥ 4.5 shows the strongest peak-to-gene expression association ($p\text{-value} = 4.46\text{E}^{-35}$; **Fig 1F**; [Supplementary Table S2](#)). Using 4.5 as a specified threshold, 1,931 peaks were associated to 2,223 genes ([Supplementary Table S3](#)), representing 3.6% of the total ensembl genes set ($n=60,235$). However, fifty-one of the 147 (34.6%) known *ETV6*-modulated genes in Reh cells [32] have at least one associated peak called in this condition (74 expression-correlated peaks; [Supplementary Table S4](#)). By including expression data, we thus selected 1,931 high priority *ETV6*-bound regions that were used in the subsequent analyses.

3.6.2. *ETV6* distribution across hematopoietic cell lines genomes

We next compared *ETV6*-bound regions identified in pre-B leukemia cells (Reh) to those of normal lymphoblastoid cells (GM12878) and myelogenous leukemia cells (K562) obtained from the ENCODE project (**Fig 2A**). The overlap between Reh and GM12878 cells (368 of the 1,931 Reh peaks; 19.05%) was greater than with K562 cells (92 of the 1,931 Reh peaks; 4.76%), suggesting that the lineage (lymphoid vs. myeloid) is more likely to shape *ETV6* binding compared to the differentiation stage (pre-B vs. B). The strongest overlap was however observed between K562 and GM12878 cells (988 of the 3,236 K562 peaks; 30.53%). Of note, the same antibody and immunoprecipitation protocol were used to generate these ENCODE datasets and might favor this overlap. Only 67 regions were bound by *ETV6* in all 3 cell lines (3.47% of Reh peaks). Peaks shared across Reh and GM12878 cells were however not more associated to *ETV6*-modulated

genes in Reh [32] ([Supplementary Table S5](#)). *ETV6* binding and expression modulation seems correlated but mostly in a cell type-specific manner.

3.6.3. Genomic features of *ETV6*-bound regions

We next interrogated the genomic localization of *ETV6* peaks. **Fig 2B** ("All peaks"; left panel) shows the percentage of peaks located within distinct genomic regions. The peaks are more often found (42%) within genes. This distribution was similar in both GM12878 and K562 cells ([Supplementary Fig S2](#)). When we assessed the localization of expression-correlated *ETV6* peaks, a stronger association (57%) was observed within gene bodies (**Fig 2B**, right panel). *ETV6* binding within genes seems to induce a change in gene expression.

The number of peaks located in the vicinity of the transcription start site (TSS) was assessed (**Fig 2C**). Of the 1,931 *ETV6* peaks, 502 (26%) were in the flanking region (+/- 10 kb) of a TSS. The highest number of peaks within this range is observed in the first 1,000bp upstream of the TSS (**Fig 2C**, left panel). The distribution of *ETV6* peaks observed in Reh cells is less TSS centered compared to GM12878 and K562 cells ([Supplementary Fig S2](#)). Although the relatively small sample size of expression-correlated *ETV6* peaks, they did not display a tighter distribution around the TSS. (**Fig 2C**, right panel), suggesting that *ETV6*-mediated transcription can be induced by distant TSS binding in Reh cells.

Histone modifications in GM12878 cells were extensively investigated [40] compared to Reh. Considering the number of shared *ETV6* peaks between Reh

and GM12878 cell lines (368 peaks, **Fig 2A**), we assigned Reh-derived *ETV6* peaks to specific chromatin states based on histone marks data from the corresponding regions in GM12878 cells. Most of *ETV6* peaks were associated to a quiescent state (**Fig 2D**, left panel). However, when focusing on shared Reh and GM12878 *ETV6* peaks, 45% of these were categorized as flanking active TSS (**Fig 2D**, center panel). Enhancers were also significantly enriched (**Fig 2D**, center panel). In contrast, Reh-specific peaks are depleted for both flanking active TSS and enhancers chromatin states and are instead further associated with quiescent regions in GM12878 cells (**Fig 2D**, right panel), indicating that *ETV6* binding to a given region strongly depends on a prerequisite chromatin environment.

3.6.4. Motifs enrichment of *ETV6*-bound regions

ETS transcription factors such as *ETV6* have well-known binding motifs [24]. As expected, motif enrichment analyses revealed significant over-representation of ETS motifs within *ETV6* peaks in Reh cells as well as in GM12878 and K562 cells (**Fig 3A**; [Supplementary Table S6](#)). Interestingly, RUNX motifs were also significantly over-represented but specifically in Reh cells (**Fig 3A, B**). RUNX motifs are found centered on the peak summits similarly to ETS motifs ([Supplementary Fig S3](#)), indicating that RUNX-containing peaks are also properly bound. Unlike GM12878 and K562 cells, Reh cells express the *ETV6*-AML1 fusion protein that can interact with the bait protein *ETV6*-HA through their PNT domain and bind RUNX motifs [41]. Therefore, *ETV6* ChIP-seq in this particular

t(12;21)-positive background may have retrieved putative ETV6-AML1-bound regions in addition to *ETV6*-bound regions.

Only minor differences were observed when we assessed the global genomic localization pattern of *ETV6* peaks according to the motif groups (refer to **Fig 3B**) in Reh cells (**Fig 3C**). However, *ETV6* binding occurs predominantly in the first 1,000bp upstream of TSS only in ETS-unique peaks (**Fig 3C**). With ETS motifs peaks being directly bound by ETV6-HA, the intensity (or peak fold enrichment) of these peaks is expected to be higher than RUNX motifs peaks which are presumably bound through ETV6-AML1. A striking difference is indeed observed when we compared the percentage of peaks containing only ETS (direct binding) or RUNX (indirect binding) motifs according to different peak fold enrichment (**Fig 3D**). ETS motifs become increasingly predominant among *ETV6*-bound regions with higher peak fold enrichment thresholds. Inversely, the fraction of *ETV6*-bound regions containing RUNX-unique motifs remains constant through the peak fold enrichment range.

ETV6 peaks containing ETS motifs (unique or with RUNX) are therefore significantly stronger (i.e. higher fold enrichment) than RUNX-unique peaks and ETS/RUNX depleted peaks (**Fig 3E**). Interestingly, peaks correlated with gene expression modulation have significantly higher fold enrichments ([Supplementary Table S7](#)) without significant differences between motif groups ([Supplementary Fig S4](#)). We then assessed the relationship between motifs and expression. As shown in **Fig 3F**, expression-correlated peaks are slightly more enriched for ETS motifs (unique or with RUNX) compared to all the peaks. Inversely, peaks lacking

both ETS and RUNX motifs are less likely to induce a change in gene expression. All together, these data indicate that *ETV6* predominantly binds ETS-containing sequences *in vivo*. ETS-driven binding of *ETV6* is generally stronger than other binding events and are even more associated with a change in gene expression.

Although most *ETV6* binding sites contain either ETS or RUNX motifs, a non-negligible fraction does not (38%, **Fig 3B**). Additional motif enrichment analyses revealed that only 10 motifs were significantly over-represented among this subset of ETS/RUNX free peaks (q-value < 0.1; [Supplementary Table S8](#)). IRF consensus sequences were present in 3 of these enriched motifs with one of them in conjunction with an ETS motif ([Supplementary Table S8](#), see motif #4). Interestingly, this ETS-IRF compound motif was also strongly associated to expression regulation ([Supplementary Fig S5](#)). In fact, 10 of the 20 expression-correlated peaks lacking ETS or RUNX motifs instead contain this ETS-IRF compound motif. It is known that *ETV6* can interact with *IRF8* [42], suggesting a potential regulation through this interaction similarly to what was observed with RUNX motifs and *ETV6*-AML1. These results suggest a role for *ETV6* as a cofactor at these loci.

3.7 Discussion

In this study, we provide the first in-depth mapping of *ETV6* binding sites in t(12;21)-positive pre-B leukemia cells. We further associated *ETV6* binding to transcriptional regulation by integrating expression data [32]. Despite the clear correlation between *ETV6* binding and gene expression in the Reh cell line, *ETV6* binding profile in these cells was markedly different of those obtained in other hematopoietic cell types. By including histone modification data, we were able to unveil the importance of the cell type-specific chromatin environment on *ETV6* binding. In this regard, roughly half of Reh-specific *ETV6* peaks maps in quiescent regions of the lymphoblastoid cell line GM12878. The cell type-specific transcriptional programs and related active chromatin states thus modulate *ETV6* binding. Although ubiquitously expressed, *ETV6* impact on expression profiles might be different across tissues and cell types as its binding depends on higher order chromatin conformations. It is noteworthy that *ETV6* silencing or overexpression in purified CD34⁺ cells from cord blood led to differentially expressed genes with however no recurrence with genes obtained in Reh cells (unpublished observations; [32]), further demonstrating the cell-specific role of *ETV6* in transcriptional regulation.

ETV6 binding might as well rely on other proteins [42, 43] whose availability would lead to additional complexity and divergences across cell types. For instance, ETV6-AML1 seems to shape *ETV6* binding profile in Reh cells. Although additional experiments could further strengthen our observations, the over-representation of RUNX motifs is likely a consequence of their interaction

[41] as these sequences are expectedly targeted by the Runt domain of the fusion protein [29]. The contrasting distribution across TSS of *ETV6* peaks in Reh cells compared to other cell types may be attributed to this interaction as the dissection per motif groups revealed a tighter distribution of ETS-unique peaks. Furthermore, not only *ETV6* is recruited to RUNX-containing sequences but also induced changes in gene expression at these loci. RUNX-unique motifs were found in similar proportions within *ETV6* peaks and expression-correlated peaks. More importantly, *ETV6* peaks containing both ETS and RUNX consensus binding sequences were even more associated with differentially expressed genes. These regions may be bound synergistically by the complex formed by ETV6-HA and endogenous ETV6-AML1 through their ETS and Runt domains, respectively. Interestingly, the expression of *ETV6* [32] or the knock-down of ETV6-AML1 [31] in Reh cells lead to common differentially expressed genes (37 of the 147 *ETV6*-modulated genes are also differentially expressed upon ETV6-AML1 knock-down), further suggesting that *ETV6* and ETV6-AML1 cooperate in transcriptional regulation.

In addition to its canonical function to directly bind ETS sequences, *ETV6* seems to be recruited to additional sites, indicating that *ETV6* can act as a cofactor without direct DNA-binding. The most striking example is the high occurrence of RUNX motifs that could be explained by the ETV6-AML1 fusion protein in Reh cells. However this is unique to the specific t(12;21)-positive background of these cells and might not reflect the general impact of *ETV6* in normal cells. Nonetheless, a significant proportion of peaks contained neither ETS nor RUNX

motifs but were still associated to a change in gene expression. These peaks were enriched for IRF binding sites in combination or not with ETS sites. This result is supported by the known interaction between *ETV6* and *IRF8* [42] and further suggests that the interaction between *ETV6* (and potentially other ETS factors) and IRF proteins is a rather common mechanism of transcriptional regulation. The function of *ETV6* as a cofactor is however poorly understood and additional efforts are required to uncover putative partners and characterize their functions.

All together, this report indicates that *ETV6* binding is highly flexible and therefore *ETV6* inactivation is expected to induce unique transcriptional modifications in a given cellular context. The wide spectrum of hematological diseases associated with *ETV6* alterations [5-10] could be explained, at least partially, by the distinct *ETV6* regulatory network of the cell originally affected by the mutation. Accordingly, the complete disruption of *ETV6* observed in most t(12;21)-positive childhood pre-B ALL cases [17-21] may induce specific transcriptional changes required for complete leukemic transformation. Although this study dissected *ETV6* binding and clarified its transcriptional network in t(12;21)-positive pre-B leukemia cells [32], it remains challenging to associate *ETV6* target genes to leukemia-related phenotypes as their functions are, for most of them, still unclear.

3.8 Conclusions

Molecular characterization of *ETV6* function is mandatory to fully understand its role in leukemogenesis. Towards this goal we built the first genome wide map of *ETV6* binding sites in pre-B leukemia cells bearing the recurrent t(12;21) translocation. By including expression data, we obtained the detailed transcriptional network of *ETV6* in these cells. This comprehensive analysis exposed the binding properties of *ETV6* and suggests that *ETV6* could also act as a cofactor to regulate gene expression. With the recent reports connecting germline *ETV6* mutations to a panel of familial hematological diseases and given the complexity of *ETV6*-mediated transcription, further characterization of *ETV6* remains valuable.

3.9 Declarations

3.9.1. Availability of data and material

The datasets generated and/or analysed during the current study are available in the GEO repository,

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102785>

3.9.2. Competing interests

The authors declare that they have no competing interests

3.9.3. Funding

This study was supported by research funds provided by the Terry Fox Research Institute and the Canadian Institutes of Health Research. BN is the recipient of a Cole Foundation scholarship. DS holds the François-Karl Viau Research Chair in Pediatric Oncogenomics. Next generation sequencing was performed at the Integrated Clinical Genomic Center in Pediatrics at the CHU Sainte-Justine Research Center. Computations were performed on the Briarée supercomputer at the Université de Montréal, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the Fonds de recherche du Québec - Nature et technologies(FRQNT).

3.10 Figures

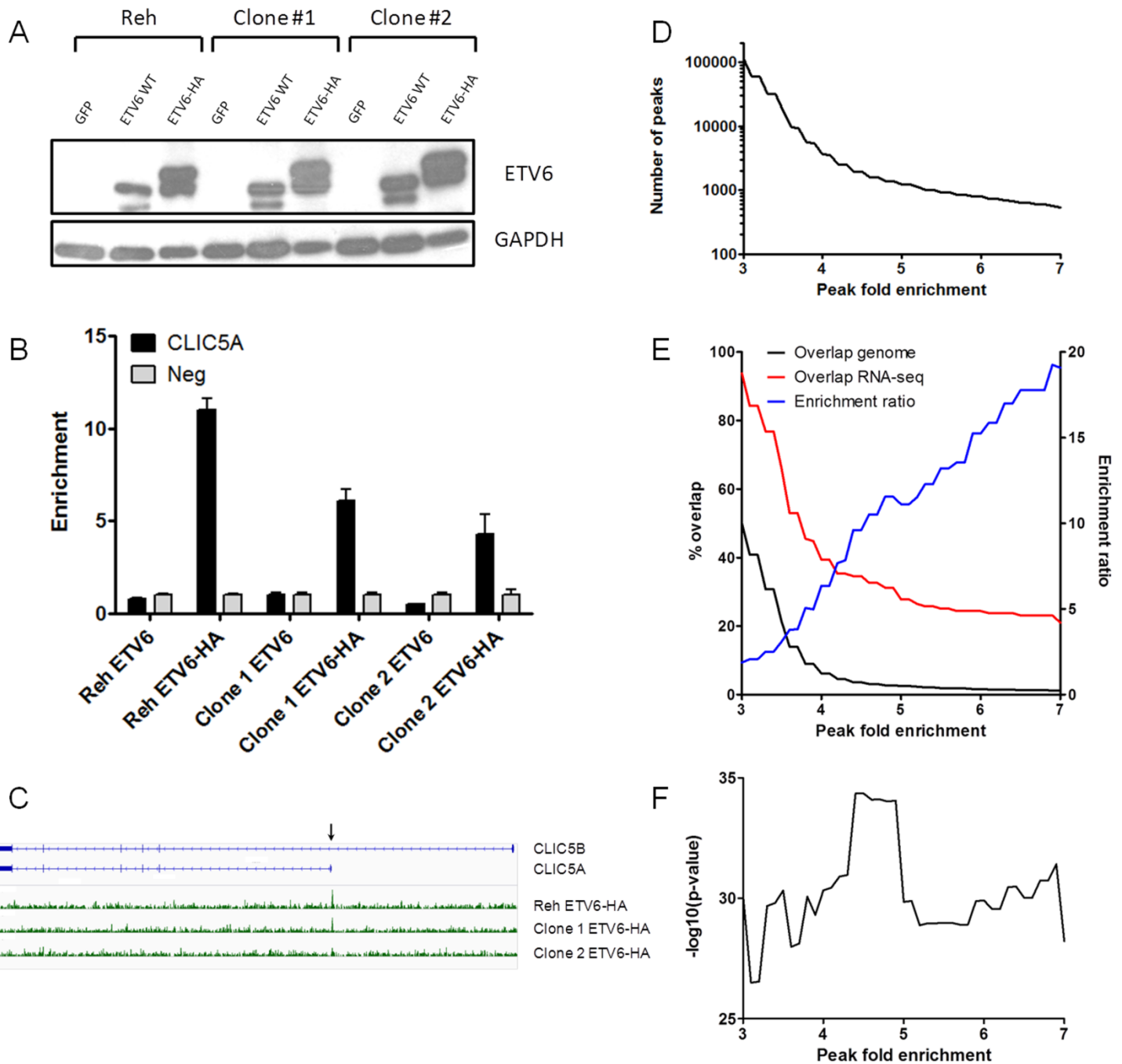


Figure 1: ChIP-seq analysis of *ETV6* binding sites in pre-B leukemia cells.
A. The Reh cell line, a pre-B leukemia cell line which is t(12;21)-positive and *ETV6* negative, was transduced with either pCCL-GFP, *ETV6* or *ETV6*-HA. The expression of *ETV6* and *ETV6*-HA was confirmed by western blot. Results from

two derived clones are shown. **B.** ChIP-qPCR analysis of the *CLIC5* locus from the HA-immunoprecipitated DNA. The *CLIC5A* promoter, a known bound region of *ETV6* [32], was successfully enriched in ETV6-HA populations compared to a negative control region (neg), but not in the untagged *ETV6* populations. Error bars represent the standard deviation (n=4). **C.** Distribution of the reads mapped to the *CLIC5* gene after sequencing of the HA-immunoprecipitated DNA. A ChIP signal at the *CLIC5A* promoter region (arrow) is seen in all three ETV6-HA populations. **D.** The number of *ETV6* ChIP peak signals using variable peak fold enrichment thresholds (0.1 increments in a 3 to 7 range). The number of peaks drastically increases with peak fold enrichment thresholds < 4. **E.** Overlap percentage (left y axis) between the genes associated with putative *ETV6* binding sites and known *ETV6*-modulated genes [32] (red line) or total ensembl genes (black line). The ratio between the percentages of overlap was calculated (blue line; right y axis) and increases with the stringency of the peak fold enrichment thresholds. **F.** A Fisher's exact test *p-value* was calculated for the overlaps between *ETV6* peaks and *ETV6*-modulated genes against the proportion of *ETV6* peaks in the complete ensembl genes repertoire. The most significant enrichment is calculated at a peak fold enrichment of ≥ 4.5 .

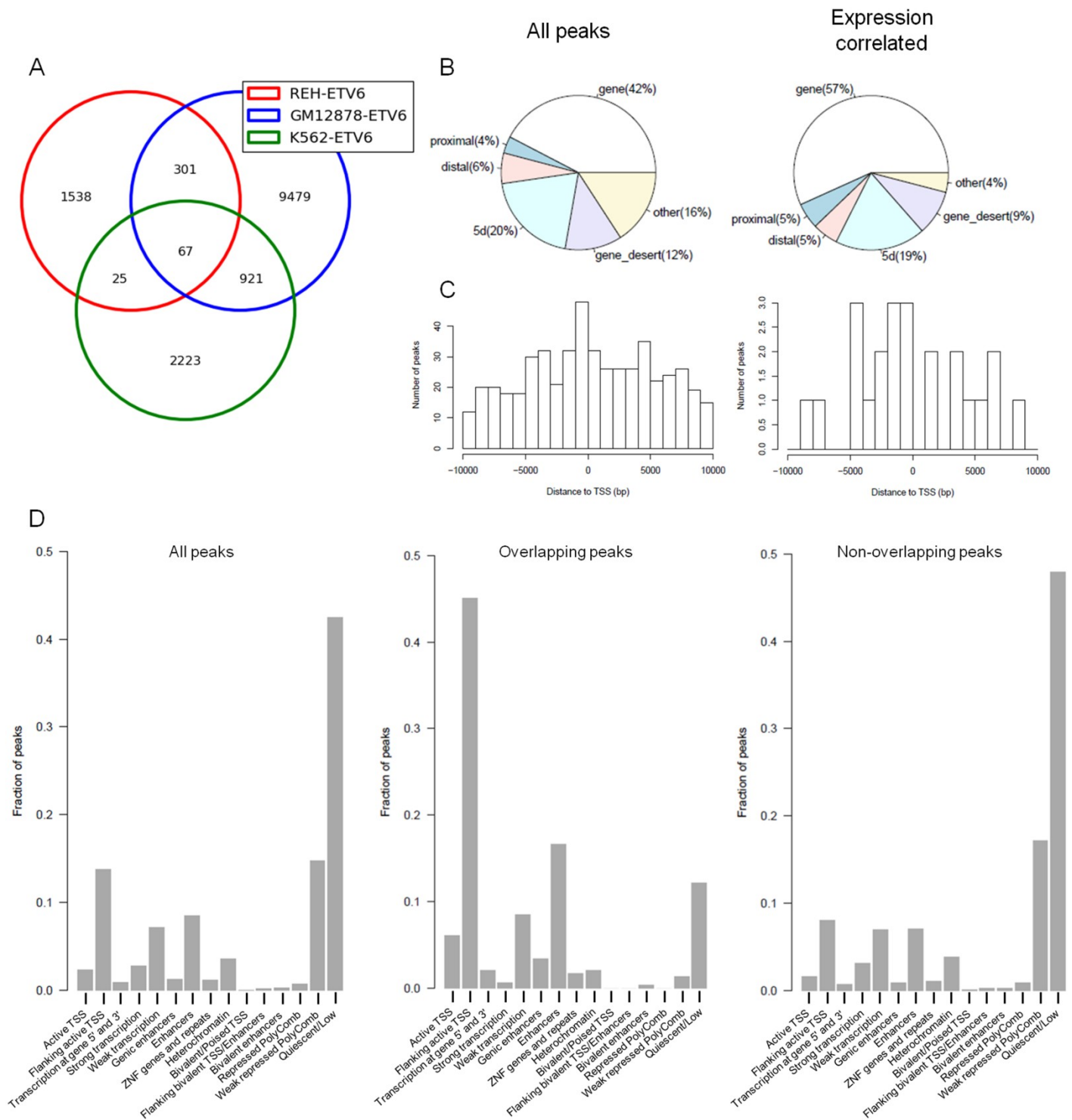


Figure 2: Genomic annotations of *ETV6*-bound regions. **A.** Comparison of *ETV6* binding sites in Reh-*ETV6* cells and two other hematopoietic cell lines, GM12878 and K562. **B.** Genomic distribution of all Reh-derived *ETV6* peaks (left panel; n=1,931) and expression-correlated peaks (right panel; n=74).

Proximal: ≤ 2 kb upstream of TSS; Distal: 2kb to 10kb upstream of TSS; 5d: 10kb to 100kb upstream of TSS. *ETV6* binding occurs mostly in genes. **C.** Distribution of all Reh-derived *ETV6* peaks (left panel) and expression-correlated peaks (right panel) in a 20 kb region across the TSS (± 10 kb). Expression-correlated peaks are not enriched for proximal TSS binding. **D.** Left panel: Reh-derived *ETV6*-bound regions were associated to GM12878 chromatin states (x axis) based on histone marks profiles. Center panel: The same analysis was restricted to Reh and GM12878 overlapping *ETV6* peaks. Those shared *ETV6*-bound regions are mainly flanking active TSS regions and enhancers. Right panel: Reh-specific *ETV6* peaks are mostly associated to quiescent regions in GM12878 cells.

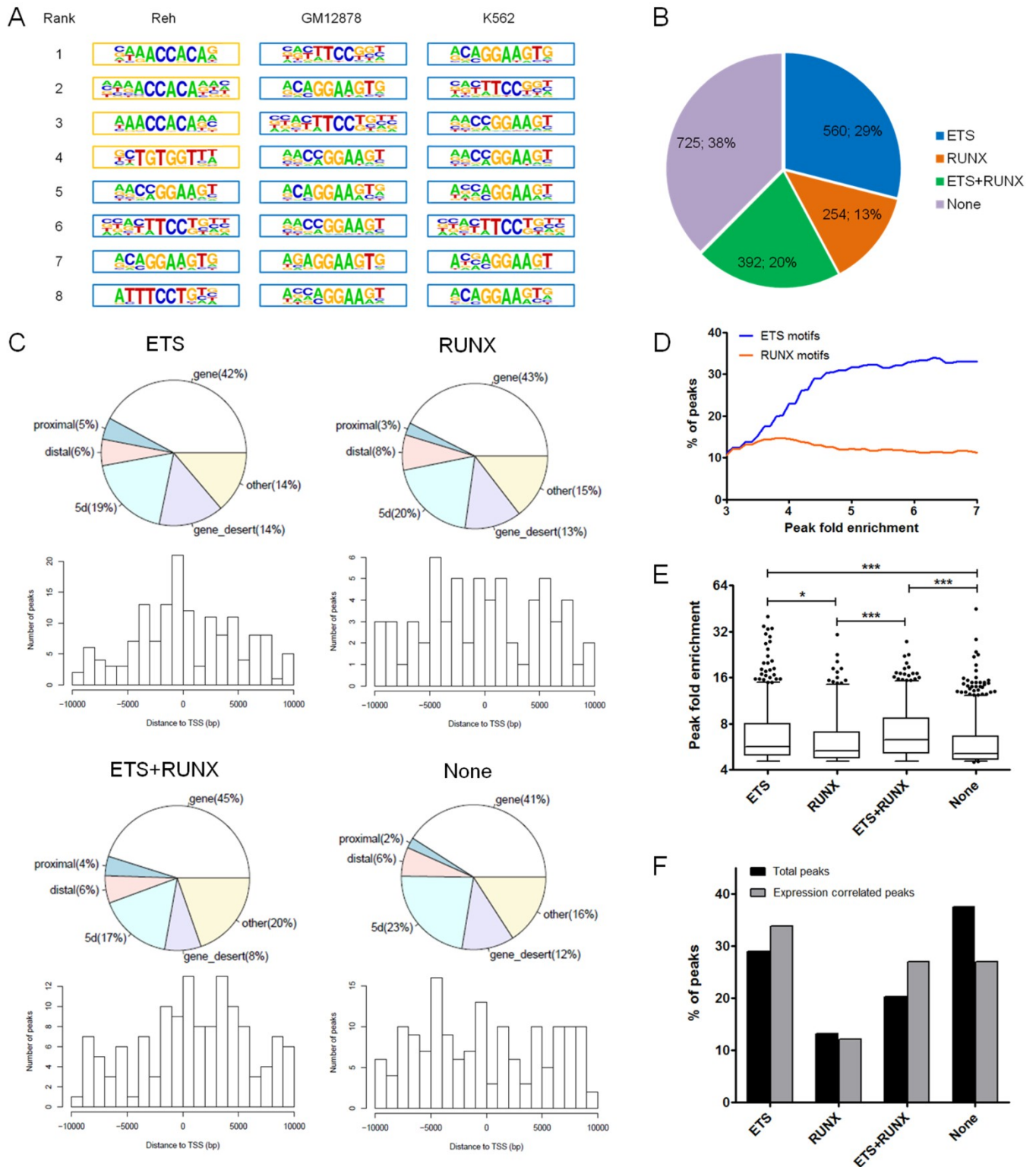


Figure 3: Consensus binding motifs enrichment analyses. A. Consensus binding motifs associated with *ETV6*-bound regions are shown for Reh, GM12878 and K562 cell lines. ETS motifs (blue border) were significantly

enriched in all three cell lines. RUNX motifs (orange border) were also significantly enriched in Reh cells. **B.** Reh-derived *ETV6*-bound regions were classified based on the presence or absence of ETS or RUNX motifs (identified in A). **C.** Genomic annotations of peaks for each motif groups. Proximal: ≤ 2 kb upstream of TSS; Distal: 2kb to 10kb upstream of TSS; 5d: 10kb to 100kb upstream of TSS. Notably, ETS-unique peaks shows a tighter distribution across the TSS region. **D.** Percentages of *ETV6*-bound regions containing either ETS or RUNX using variable peak fold enrichment thresholds. ETS but not RUNX motifs become increasingly more frequent with stronger peak fold enrichments. **E.** Fold enrichment of all *ETV6* peaks in Reh cells according to each motif groups. ETS containing peaks (ETS and the ETS/RUNX compound) have significantly greater fold enrichments. Whiskers represent the 5-95 percentiles. Statistical significance is calculated by two-tailed Student's *t* test. **F.** Proportion of each motif groups in expression-correlated *ETV6* peaks ($n=74$) compared to all *ETV6* peaks ($n=1,931$). ETS containing peaks (ETS and the ETS/RUNX compound) are even more over-represented in expression-correlated peaks.

3.11 References

1. Kar A, Gutierrez-Hartmann A. Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. *Crit Rev Biochem Mol Biol.* 2013; 48(6):522-543.
2. Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F et al. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* 1998; 12(15):2392-2402.
3. Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, Orkin SH. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J.* 1997; 16(14):4374-4383.
4. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: a review. *Leuk Res.* 2012; 36(8):945-961.
5. Zhang MY, Churpek JE, Keel SB, Walsh T, Lee MK, Loeb KR et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet.* 2015; 47(2):180-185.
6. Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet.* 2015; 47(5):535-538.
7. Topka S, Vijai J, Walsh MF, Jacobs L, Maria A, Villano D et al. Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet.* 2015; 11(6):e1005262.
8. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol.* 2015; 16(16):1659-1666.
9. Kirkpatrick G, Noetzli L, Di Paola J, Porter CC. ETV6 mutations define a new cancer predisposition syndrome. *Oncotarget.* 2015; 6(19):16830-16831.
10. Romero D. Haematological cancer: ETV6 germline mutation - a risk for ALL. *Nat Rev Clin Oncol.* 2016; 13(1):4.
11. Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P et al. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A.* 1995; 92(11):4917-4921.

12. Tasian SK, Loh ML, Hunger SP. Childhood acute lymphoblastic leukemia: Integrating genomics into therapy. *Cancer*. 2015; 121(20):3577-3590.
13. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet*. 2001; 130(2):93-104.
14. van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J et al. Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood*. 2011; 118(4):1041-1051.
15. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A*. 2002; 99(12):8242-8247.
16. Greaves M. Childhood leukaemia. *BMJ*. 2002; 324(7332):283-287.
17. Poirel H, Lacronique V, Mauchauffe M, Le Coniat M, Raffoux E, Daniel MT et al. Analysis of TEL proteins in human leukemias. *Oncogene*. 1998; 16(22):2895-2903.
18. Patel N, Goff LK, Clark T, Ford AM, Foot N, Lillington D et al. Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br J Haematol*. 2003; 122(1):94-98.
19. Lilljebjorn H, Sonesson C, Andersson A, Heldrup J, Behrendtz M, Kawamata N et al. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet*. 2010; 19(16):3150-3158.
20. Montpetit A, Boily G, Sinnott D. A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus. *Eur J Hum Genet*. 2002; 10(1):62-71.
21. Montpetit A, Larose J, Boily G, Langlois S, Trudel N, Sinnott D. Mutational and expression analysis of the chromosome 12p candidate tumor suppressor genes in pre-B acute lymphoblastic leukemia. *Leukemia*. 2004; 18(9):1499-1504.
22. Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem*. 1999; 274(42):30132-30138.
23. Mavrothalassitis G, Ghysdael J. Proteins of the ETS family with transcriptional repressor activity. *Oncogene*. 2000; 19(55):6524-6532.

24. Bohlander SK. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol.* 2005; 15(3):162-174.
25. Chakrabarti SR, Nucifora G. The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun.* 1999; 264(3):871-877.
26. Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A et al. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood.* 2000; 96(7):2557-2561.
27. Wang L, Hiebert SW. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene.* 2001; 20(28):3716-3725.
28. Zelent A, Greaves M, Enver T. Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. *Oncogene.* 2004; 23(24):4275-4283.
29. Linka Y, Ginzl S, Kruger M, Novosel A, Gombert M, Kremmer E et al. The impact of TEL-AML1 (ETV6-RUNX1) expression in precursor B cells and implications for leukaemia using three different genome-wide screening methods. *Blood Cancer J.* 2013; 3:e151.
30. Linka Y, Ginzl S, Borkhardt A, Landgraf P. Identification of TEL-AML1 (ETV6-RUNX1) associated DNA and its impact on mRNA and protein output using ChIP, mRNA expression arrays and SILAC. *Genom Data.* 2014; 2:85-88.
31. Ghazavi F, De Moerloose B, Van Loocke W, Wallaert A, Helsmoortel HH, Ferster A et al. Unique long non-coding RNA expression signature in ETV6/RUNX1-driven B-cell precursor acute lymphoblastic leukemia. *Oncotarget.* 2016.
32. Neveu B, Spinella JF, Richer C, Lagace K, Cassart P, Lajoie M et al. CLIC5: a novel ETV6 target gene in childhood acute lymphoblastic leukemia. *Haematologica.* 2016; 101(12):1534-1543.
33. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10(3):R25.
34. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9(9):R137.

35. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57-74.
36. Zhao X, Emery SB, Myers B, Kidd JM, Mills RE. Resolving complex structural genomic rearrangements using a randomized approach. *Genome Biol*. 2016; 17(1):126.
37. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics*. 2016; 32(15):2375-2377.
38. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 2012; 28(14):1919-1920.
39. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38(4):576-589.
40. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518(7539):317-330.
41. Gunji H, Waga K, Nakamura F, Maki K, Sasaki K, Nakamura Y et al. TEL/AML1 shows dominant-negative effects over TEL as well as AML1. *Biochem Biophys Res Commun*. 2004; 322(2):623-630.
42. Kuwata T, Gongora C, Kanno Y, Sakaguchi K, Tamura T, Kanno T et al. Gamma interferon triggers interaction between ICSBP (IRF-8) and TEL, recruiting the histone deacetylase HDAC3 to the interferon-responsive element. *Mol Cell Biol*. 2002; 22(21):7439-7448.
43. Schick N, Oakeley EJ, Hynes NE, Badache A. TEL/ETV6 is a signal transducer and activator of transcription 3 (Stat3)-induced repressor of Stat3 activity. *J Biol Chem*. 2004; 279(37):38787-38796.

3.12 Supplementary material

3.12.1. Supplementary figures

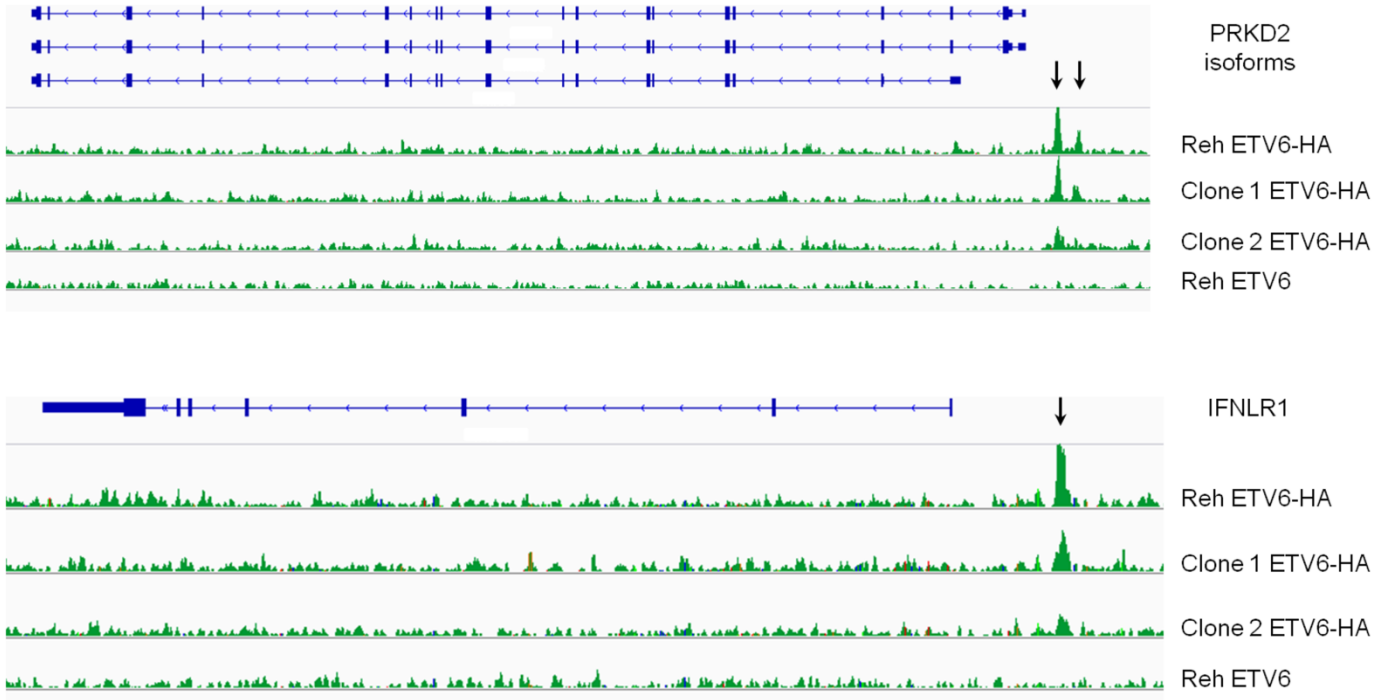


Figure S1: Examples of identified *ETV6* peaks. Distribution of the reads mapped to the *PRKD2* (upper panel) and *IFNLR1* (lower panel) genes after sequencing of the HA-immunoprecipitated DNA. Peaks in their promoter region (arrows) are seen in all three *ETV6*-HA populations but not in Reh *ETV6* cells (negative immunoprecipitation control).

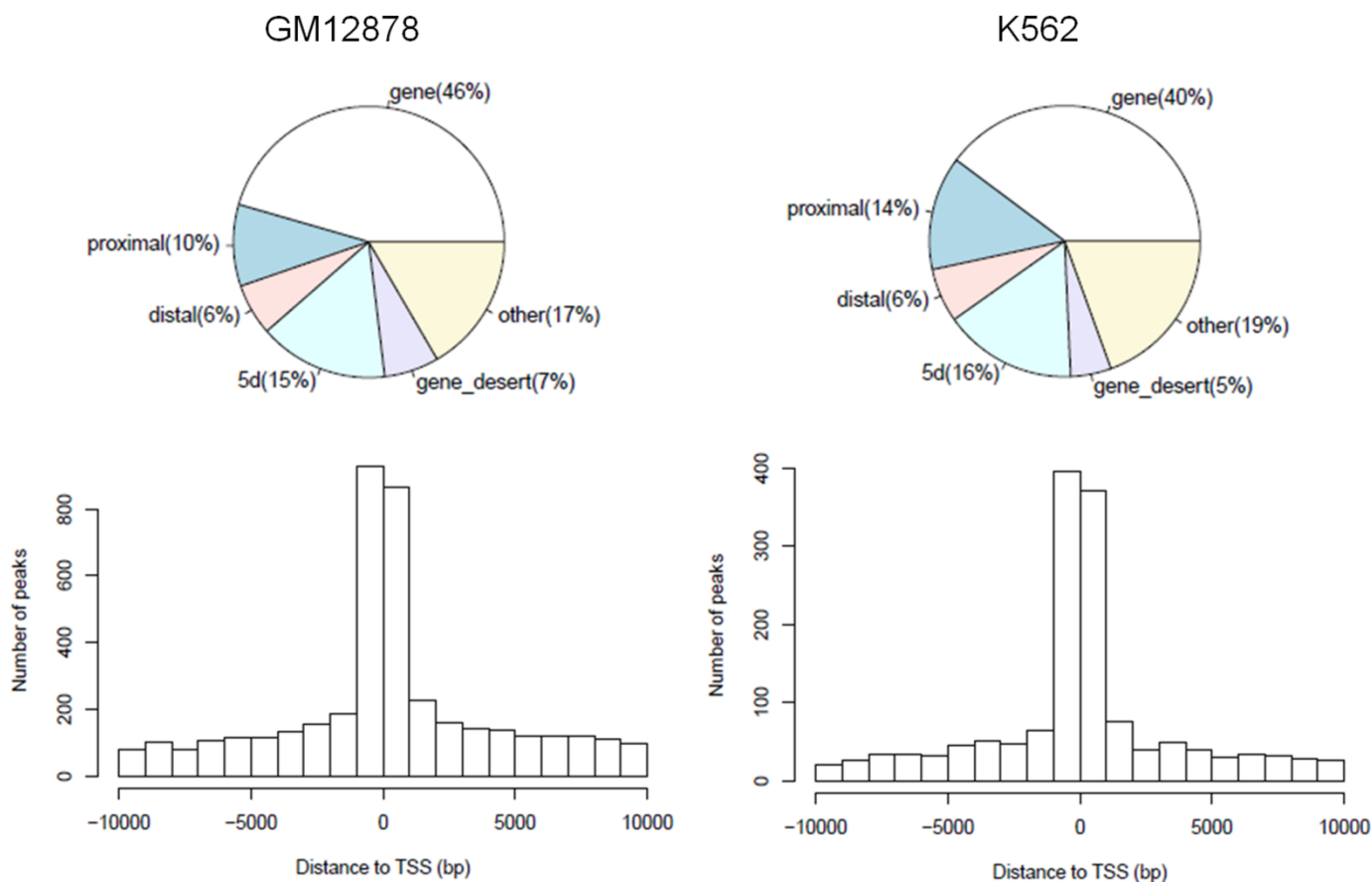


Figure S2: Genomic annotations of *ETV6* peaks in GM12878 and K562 cells. Genomic distribution of *ETV6* binding sites are shown in upper panels. Proximal: ≤ 2 kb upstream of TSS; Distal: 2kb to 10kb upstream of TSS; 5d: 10kb to 100kb upstream of TSS. Distribution of peaks in a 20 kb region spanning the TSS (± 10 kb) is shown in lower panels. *ETV6* peaks in both GM12878 and K562 cells display a clear TSS centered distribution.

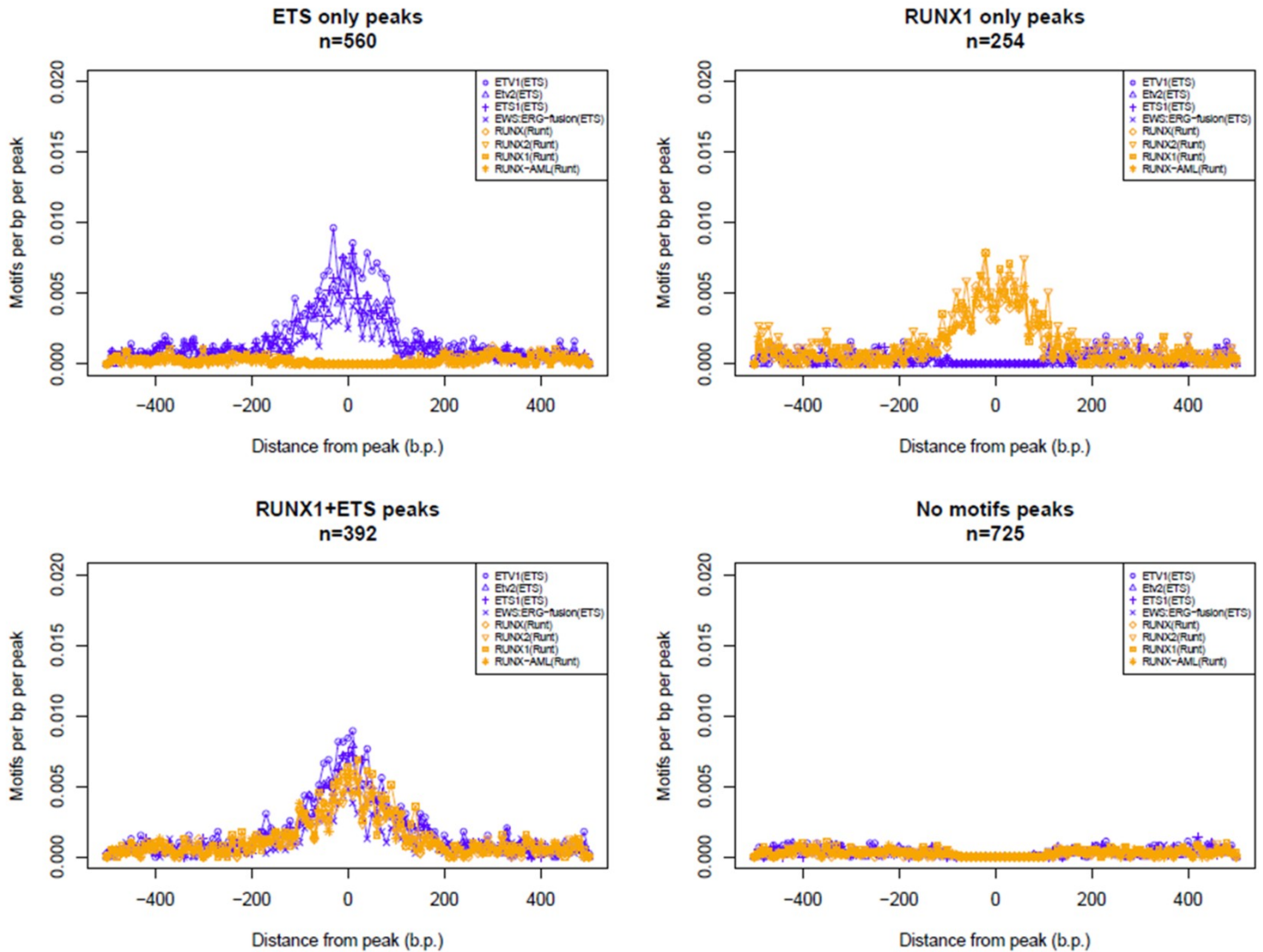


Figure S3: Density of the top 8 over-represented motifs in Reh *ETV6* peaks. The density of the 4 over-represented ETS motifs (blue) and RUNX motifs (yellow) identified in Reh-derived *ETV6*-bound regions is shown in a 1,000 bp window (+/- 500 bp) around the peak summits for each motif group. ETS and RUNX motifs are clustered to the peak summits.

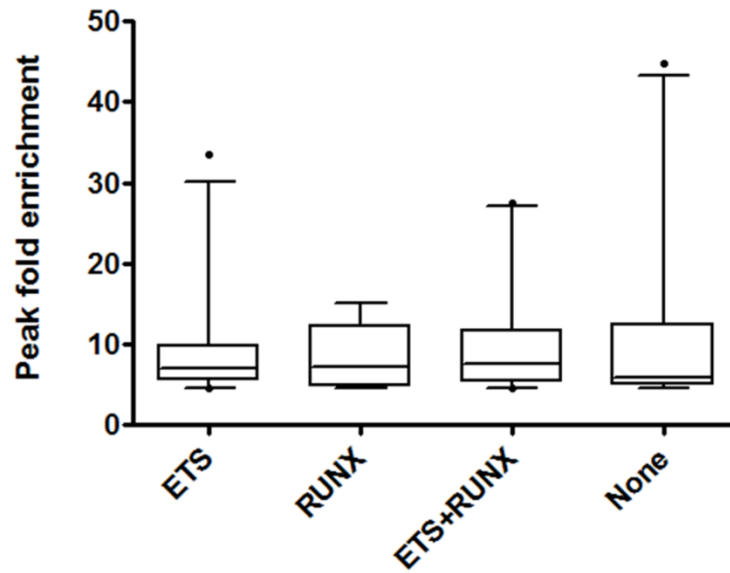


Figure S4: Fold enrichment of expression-correlated *ETV6* peaks in Reh cells according to each motif groups. Peak fold enrichments are not significantly different between motif groups among expression-correlated *ETV6* peaks (n=74). Whiskers represent the 5-95 percentiles. Statistical significance is calculated by two-tailed Student's *t* test.

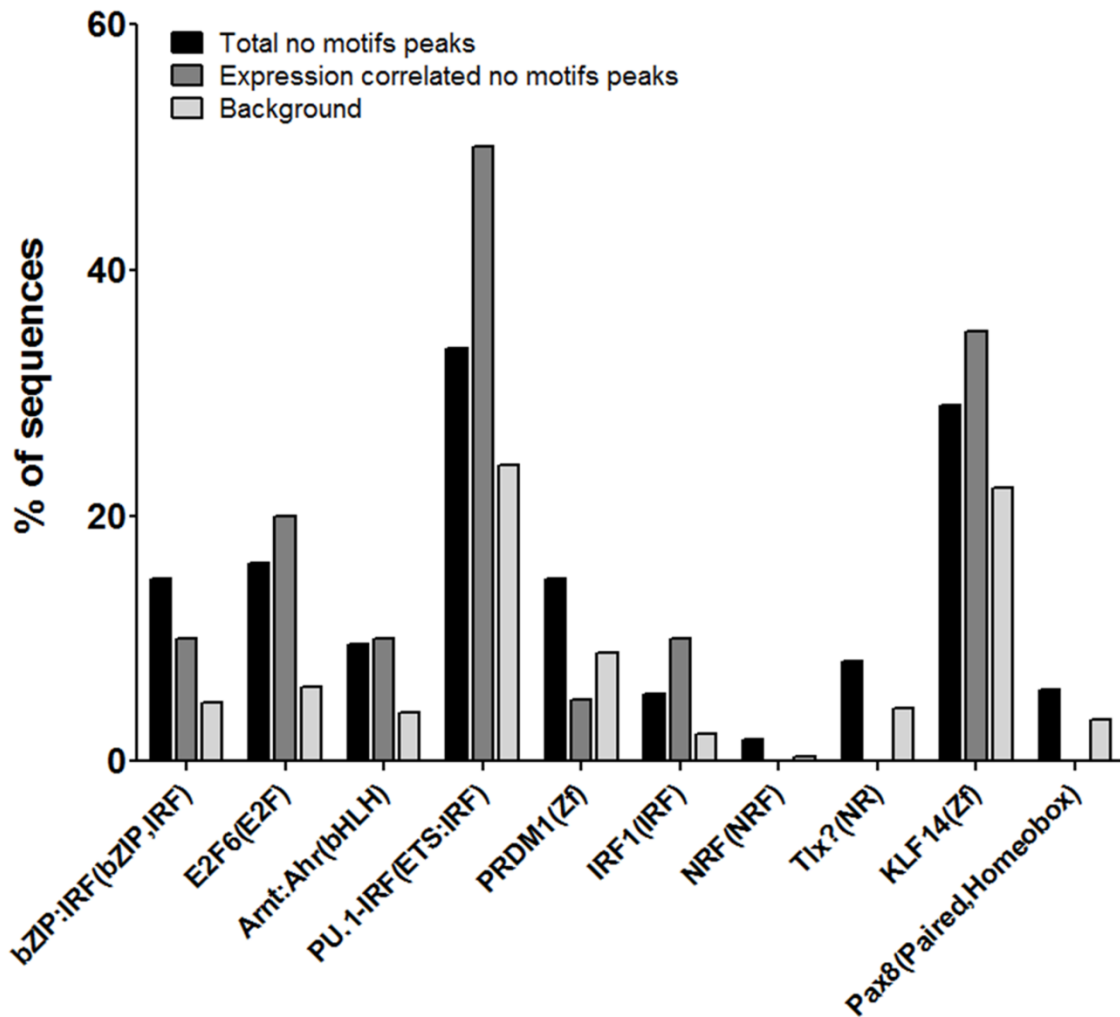


Figure S5: Motif enrichment analysis of ETS/RUNX free *ETV6* peaks in Reh cells. A second round of motif enrichment was performed on ETS/RUNX free *ETV6* peaks (n=725) which identified 10 significantly over-represented motifs (shown on x axis). The percentage of sequences containing each motif is given for the totality of ETS/RUNX free peaks (n=725), expression-correlated ETS/RUNX free peaks (n=20) and for the background (n= 33,996). Notably, an ETS-IRF compound motif (4th motif) was over-represented and strongly associated to expression regulation.

3.12.2. Supplementary tables

Table S1: Oligonucleotide sequences used for chromatin immunoprecipitation qPCR

Oligo Name	Sequence (5' to 3')
ChIP_CLIC5_A_F	CCAGCACCCCTACTCTCCTTT
ChIP_CLIC5_A_R	CTCCACTACCAGCCTTACCC
Neg_F	ATGGTTGCCACTGGGGATCT
Neg_R	TGCCAAAGCCTAGGGGAAGA

Table S2: Summary statistics of Reh-derived *ETV6* peaks using different peak fold enrichment thresholds

Peak fold enrichment	Peaks	Associated genes	Total ensembl genes	Overlap RNA-seq (genes)	Total RNA-seq genes	Overlap ensembl (%)	Overlap RNA-seq (%)	Enrichment ratio	Fisher exact p-value
3	110 435	30133	60235	138	147	50	93.8	1.876	7,61E-31
3,1	60 248	24641	60235	124	147	40.9	84.3	2.061	3,19E-27
3,2	59 771	24584	60235	124	147	40.8	84.3	2.066	2,87E-27
3,3	31 801	18549	60235	113	147	30.7	76.8	2.501	2,09E-30
3,4	31 563	18495	60235	113	147	30.7	76.8	2.501	1,57E-30
3,5	17 036	12864	60235	97	147	21.3	65.9	3.093	4,75E-31
3,6	9 495	8438	60235	78	147	14	53	3.785	1,03E-28
3,7	9 399	8399	60235	78	147	13.9	53	3.812	7,55E-29
3,8	5 527	5478	60235	67	147	9	45.5	5.055	8,45E-31
3,9	5 475	5446	60235	66	147	9	44.8	4.977	4,95E-30
4	3 581	3792	60235	58	147	6.2	39.4	6.354	4,70E-31
4,1	3 549	3773	60235	58	147	6.2	39.4	6.354	3,62E-31
4,2	2 512	2778	60235	52	147	4.6	35.3	7.673	1,26E-31
4,3	2 489	2769	60235	52	147	4.5	35.3	7.844	1,08E-31
4,4	1 934	2223	60235	51	147	3.6	34.6	9.611	4,46E-35
4,5	1 931	2223	60235	51	147	3.6	34.6	9.611	4,46E-35
4,6	1 598	1891	60235	48	147	3.1	32.6	10.516	8,06E-35
4,7	1 595	1890	60235	48	147	3.1	32.6	10.516	7,87E-35
4,8	1 382	1675	60235	46	147	2.7	31.2	11.555	9,60E-35
4,9	1 376	1672	60235	46	147	2.7	31.2	11.555	8,89E-35
5	1 226	1516	60235	41	147	2.5	27.8	11.12	1,42E-30
5,1	1 223	1512	60235	41	147	2.5	27.8	11.12	1,28E-30
5,2	1 108	1391	60235	39	147	2.3	26.5	11.521	1,32E-29
5,3	1 009	1281	60235	38	147	2.1	25.8	12.285	1,09E-29
5,4	1 009	1281	60235	38	147	2.1	25.8	12.285	1,09E-29
5,5	923	1181	60235	37	147	1.9	25.1	13.21	1,04E-29
5,6	923	1181	60235	37	147	1.9	25.1	13.21	1,04E-29
5,7	846	1092	60235	36	147	1.8	24.4	13.555	1,24E-29
5,8	846	1092	60235	36	147	1.8	24.4	13.555	1,24E-29
5,9	790	1021	60235	36	147	1.6	24.4	15.25	1,31E-30
6	787	1019	60235	36	147	1.6	24.4	15.25	1,23E-30
6,1	736	956	60235	35	147	1.5	23.8	15.866	2,78E-30
6,2	736	956	60235	35	147	1.5	23.8	15.866	2,78E-30
6,3	683	897	60235	35	147	1.4	23.8	17	3,47E-31
6,4	681	895	60235	35	147	1.4	23.8	17	3,23E-31
6,5	641	843	60235	34	147	1.3	23.1	17.769	9,59E-31

6,6	641	843	60235	34	147	1.3	23.1	17.769	9,59E-31
6,7	603	800	60235	34	147	1.3	23.1	17.769	1,82E-31
6,8	603	800	60235	34	147	1.3	23.1	17.769	1,82E-31
6,9	568	762	60235	34	147	1.2	23.1	19.25	3,86E-32
7	531	715	60235	31	147	1.1	21	19.09	6,11E-29

Table S3: Characteristics of the 1,931 prioritized Reh-derived ETV6 peaks

Cette table sera téléchargeable avec la version en ligne de l'article.

Table S4: Details of 74 Reh-derived *ETV6* peaks associated with a differentially expressed gene*La table complète sera téléchargeable avec la version en ligne de l'article.*

Peak ID	Fold change	Distance to TSS	Nearest Promoter ID	Gene Name	Closest gene ensembl ID	Closest gene symbol	Differently expressed gene
chr4.16006392.16006581	5,10708	-41628	NM_031950	FGFBP2	ENSG00000007062	PROM1	FGFBP2
chr12.3238530.3238736	4,72878	52112	NM_006675	TSPAN9	ENSG00000011105	TSPAN9	TSPAN9
chr19.48244989.48245642	10,21416	-3477	NM_015710	GLTSCR2	ENSG00000024422	EHD2	EHD2
chr7.90720923.90721162	8,32265	-172740	NM_003505	FZD1	ENSG00000058091	CDK14	CDK14
chr3.123406336.123406592	5,67454	-2027	NR_046629	MYLK-AS2	ENSG00000065534	MYLK	MYLK
chr3.123476205.123476400	4,53963	67812	NR_046629	MYLK-AS2	ENSG00000065534	MYLK	MYLK
chr3.123539346.123539596	7,7552	63678	NM_053026	MYLK	ENSG00000065534	MYLK	MYLK
chr7.83766205.83766403	5,48538	57913	NM_006080	SEMA3A	ENSG00000075213	SEMA3A	SEMA3A
chr1.21979897.21980285	8,5118	-1743	NM_001145658	RAP1GAP	ENSG00000076864	RAP1GAP	RAP1GAP
chr18.39516642.39517035	17,02361	-18324	NM_002647	PIK3C3	ENSG00000078142	PIK3C3	PIK3C3
chr1.198651011.198651224	5,10708	43020	NM_002838	PTPRC	ENSG00000081237	PTPRC	PTPRC
chr11.44590209.44591048	10,97077	3488	NM_001024844	CD82	ENSG00000085117	CD82	CD82
chr22.22280786.22280985	4,53963	26364	NM_014634	PPM1F	ENSG00000100034	PPM1F	PPM1F
chr18.19305614.19305853	10,97077	-15556	NM_020774	MIB1	ENSG00000101752	MIB1	MIB1
chr15.30134608.30134819	4,91793	-20008	NM_003257	TJP1	ENSG00000104067	TJP1	TJP1
chr6.45983706.45984094	6,99859	-274	NM_016929	CLIC5	ENSG00000112782	CLIC5	CLIC5
chr2.182321379.182321699	11,91652	-80	NM_000885	ITGA4	ENSG00000115232	ITGA4	ITGA4
chr2.183442596.183442779	4,53963	-55116	NM_001258312	PDE1A	ENSG00000115252	PDE1A	PDE1A
chr22.24094584.24094833	5,67454	-1430	NM_021916	ZNF70	ENSG00000128218	VPREB3	VPREB3
chr10.129862984.129863215	8,32265	17287	NM_130435	PTPRE	ENSG00000132334	PTPRE	PTPRE
chr10.129861249.129861719	10,21416	15671	NM_130435	PTPRE	ENSG00000132334	PTPRE	PTPRE
chr17.2717917.2718256	5,29623	18355	NM_001100398	RAP1GAP2	ENSG00000132359	RAP1GAP2	RAP1GAP2
chr18.28607327.28607563	12,48398	15336	NM_001941	DSC3	ENSG00000134762	DSC3	DSC3

chr13.80884101.80884384	9,64671	30843	NM_005842	SPRY2	ENSG00000136158	SPRY2	SPRY2
chr9.129944453.129944652	7,7552	-59509	NM_012098	ANGPTL2	ENSG00000136828	RALGPS1	ANGPTL2
chr9.129846544.129846765	4,72878	38389	NM_012098	ANGPTL2	ENSG00000136828	RALGPS1	ANGPTL2
chr9.129929210.129929420	5,29623	-44271	NM_012098	ANGPTL2	ENSG00000136828	RALGPS1	ANGPTL2
chr12.12043857.12043992	44,82883	-179953	NM_138723	BCL2L14	ENSG00000139083	ETV6	ETV6
chr12.11951886.11952090	5,86369	149200	NM_001987	ETV6	ENSG00000139083	ETV6	ETV6
chr12.12038842.12038968	33,47976	-184973	NM_138723	BCL2L14	ENSG00000139083	ETV6	ETV6
chr3.65692229.65692432	4,53963	-187160	NR_046575	MAGI1-AS1	ENSG00000151276	MAGI1	MAGI1
chr5.78723366.78723529	4,91793	86211	NM_001277077	HOMER1	ENSG00000152413	HOMER1	HOMER1
chr7.84707531.84707833	9,26841	43565	NM_152754	SEMA3D	ENSG00000153993	SEMA3D	SEMA3D
chr21.46351026.46351417	9,26841	-2434	NM_001127491	ITGB2	ENSG00000160255	ITGB2	ITGB2
chr21.46340766.46341036	7,18775	-24	NM_001303238	ITGB2	ENSG00000160255	ITGB2	ITGB2
chr3.195944373.195944735	7,18775	1171	NM_152672	SLC51A	ENSG00000163959	SLC51A	SLC51A
chr10.104561814.104562072	14,37549	26055	NM_017787	WBP1L	ENSG00000166272	WBP1L	WBP1L
chr10.104572329.104572569	6,43114	24841	NM_000102	CYP17A1	ENSG00000166272	WBP1L	WBP1L
chr2.239052858.239053246	15,1321	5689	NM_198582	KLHL30	ENSG00000168427	KLHL30	KLHL30
chr5.148227654.148227889	4,72878	21616	NM_000024	ADRB2	ENSG00000169252	ADRB2	ADRB2
chr17.39723467.39724058	27,42692	4547	NM_000226	KRT9	ENSG00000171403	KRT9	KRT9
chr2.109830652.109830810	4,53963	-72687	NR_036223	MIR4265	ENSG00000172985	SH3RF3	SH3RF3
chr1.111153325.111153847	13,24058	-4611	NM_004974	KCNA2	ENSG00000177301	KCNA2	KCNA2
chr17.75441507.75441873	7,18775	-4923	NM_001113496	SEPT9	ENSG00000184640	SEPT9	SEPT9
chr17.75437481.75437827	5,10708	-8959	NM_001113496	SEPT9	ENSG00000184640	SEPT9	SEPT9
chr6.36297630.36297904	4,91793	6895	NM_001010903	C6orf222	ENSG00000189325	C6orf222	C6orf222
chr1.156838603.156838970	12,48398	8116	NM_002529	NTRK1	ENSG00000198400	NTRK1	NTRK1
chr13.80966582.80966815	7,3769	-51613	NM_005842	SPRY2	ENSG00000202398	RNU6-61P	SPRY2
chr7.155165125.155165415	5,48538	-4642	NR_103545	BLACE	ENSG00000204960	BLACE	BLACE
chr7.155153821.155154252	6,24199	6591	NR_103545	BLACE	ENSG00000204960	BLACE	BLACE
chr9.116536186.116536423	7,7552	-102257	NM_133374	ZNF618	ENSG00000214669	RP11-18B16.1	ZNF618

chr7.155168098.155168558	7,18775	-7700	NR_103545	BLACE	ENSG00000218672	AC008060.7	BLACE
chr22.22296830.22297241	6,43114	10214	NM_014634	PPM1F	ENSG00000224086	LL22NC03-86G7.1	PPM1F
chr9.1938179.1938391	4,72878	-76934	NM_003070	SMARCA2	ENSG00000224092	RP11-443B9.1	SMARCA2
chr9.1858010.1858324	4,53963	-157052	NM_003070	SMARCA2	ENSG00000224092	RP11-443B9.1	SMARCA2
chr21.46347450.46347911	5,29623	1107	NM_001127491	ITGB2	ENSG00000227039	ITGB2-AS1	ITGB2
chr21.46345051.46345601	22,50899	3462	NM_001127491	ITGB2	ENSG00000227039	ITGB2-AS1	ITGB2
chr9.137462183.137462471	4,91793	-71324	NM_001278074	COL5A1	ENSG00000227150	RP11-473E2.3	COL5A1
chr13.80024108.80024295	5,86369	-31057	NM_001161407	NDFIP2	ENSG00000227354	RBM26-AS1	NDFIP2
chr13.80022638.80022841	6,99859	-32519	NM_001161407	NDFIP2	ENSG00000227354	RBM26-AS1	NDFIP2
chr1.53846552.53846777	8,32265	52760	NR_131923	LOC105378732	ENSG00000230138	RP11-117D22.2	RP11-117D22.2
chr1.192482429.192482673	6,05284	-62306	NM_002922	RGS1	ENSG00000232498	RP5-1011O1.2	RGS1
chr1.192485995.192486191	6,99859	-58764	NM_002922	RGS1	ENSG00000232498	RP5-1011O1.2	RGS1
chr1.199010643.199010979	12,67313	-22718	NR_110525	LINC01222	ENSG00000235492	LINC01221	LINC01221
chr13.46962401.46962915	22,13069	-1023	NM_001286766	KIAA0226L	ENSG00000241353	PPP1R2P4	KIAA0226L
chr12.93941679.93942020	12,86228	-21748	NM_003877	SOCS2	ENSG00000246985	SOCS2-AS1	SOCS2
chr4.19593637.19593893	6,05284	-659763	NM_001289135	SLIT2	ENSG00000248515	RP11-608O21.1	SLIT2
chr4.19975225.19975440	4,91793	-278195	NM_001289135	SLIT2	ENSG00000250611	RP11-339D20.1	SLIT2
chr12.3159439.3159763	14,75379	-26920	NM_001168320	TSPAN9	ENSG00000250899	RP11-253E3.3	TSPAN9
chr15.95128077.95128273	4,91793	229011	NM_001159644	MCTP2	ENSG00000259331	RP11-57P19.1	MCTP2
chr16.54407728.54408050	7,94435	-87511	NM_024336	IRX3	ENSG00000261803	RP11-434E6.2	IRX3
chr16.54400835.54401261	5,67454	-80670	NM_024336	IRX3	ENSG00000261803	RP11-434E6.2	IRX3
chr16.54490239.54490759	14,56464	-170121	NM_024336	IRX3	ENSG00000261803	RP11-434E6.2	IRX3
chr16.54508993.54509343	12,10568	-188790	NM_024336	IRX3	ENSG00000264079	AC009127.1	IRX3

Table S5: Reh and GM12878 overlapping peaks and correlated gene expression

	Reh <i>ETV6</i> CHIP-seq		Expression correlated	
	# peaks	# genes	# genes	gene overlap (%)
Total	1931	2223	51	2,29
GM overlap	386	457	14	3,06

Table S6: Motifs enrichment of *ETV6*-bound regions in Reh, GM12878 and K562 cell lines

La table complète sera téléchargeable avec la version en ligne de l'article.

Pour les Reh

Motif Name	Consensus	P-value	# of Target Sequences with Motif (of 1931)	% of Target Sequences with Motif	% of Background Sequences with Motif
RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	SAAACCACAG	1,00E-84	355.0	18.38%	5.63%
RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	NWAACCACADNN	1,00E-75	372.0	19.26%	6.62%
RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	AAACCACARM	1,00E-67	417.0	21.60%	8.58%
RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	GCTGTGGTTW	1,00E-65	336.0	17.40%	6.09%
ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	AACCGGAAGT	1,00E-59	670.0	34.70%	18.89%
Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	NNAYTTCCTGHN	1,00E-54	513.0	26.57%	13.17%
ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	ACAGGAAGTG	1,00E-52	539.0	27.91%	14.40%
EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	ATTCCTGTN	1,00E-36	353.0	18.28%	9.01%
Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	NRYTTCGGH	1,00E-35	473.0	24.50%	13.84%
GABPA(ETS)/Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	RACCGGAAGT	1,00E-34	418.0	21.65%	11.71%
ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	ACAGGAAGTG	1,00E-34	697.0	36.10%	23.63%
E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	GGCGGGAARN	1,00E-33	281.0	14.55%	6.64%
ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIP-Seq(GSE17954)/Homer	RCAGGATGTGGT	1,00E-31	105.0	5.44%	1.33%
PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	MGGAAGTGAAAC	1,00E-31	663.0	34.33%	22.51%
Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	MACAGGAAGT	1,00E-29	178.0	9.22%	3.51%
EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)/Homer	VACAGGAAAT	1,00E-28	268.0	13.88%	6.64%
bZIP:IRF(bZIP,IRF)/Th17-BatF-ChIP-Seq(GSE39756)/Homer	NAGTTTCABHTGACTNW	1,00E-23	194.0	10.05%	4.51%
Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	HACTTCGGY	1,00E-21	219.0	11.34%	5.60%
Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	NRYTTCGGY	1,00E-21	204.0	10.56%	5.12%

ETS(ETS)/Promoter/Homer	AACCGGAAGT	1,00E-19	148.0	7.66%	3.30%
EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	AVCAGGAAGT	1,00E-19	538.0	27.86%	19.30%
PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	AGAGGAAGTG	1,00E-16	248.0	12.84%	7.40%
MyoD(bHLH)/Myotube-MyoD-ChIP-Seq(GSE21614)/Homer	RRCAGCTGYTSY	1,00E-14	215.0	11.13%	6.38%
Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	BAACAGCTGT	1,00E-14	208.0	10.77%	6.11%
Tcf12(bHLH)/GM12878-Tcf12-ChIP-Seq(GSE32465)/Homer	VCAGCTGYTG	1,00E-13	279.0	14.45%	9.10%
ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	AVCCGGAAGT	1,00E-13	199.0	10.31%	5.88%
Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	NNVVCAGCTGBN	1,00E-12	434.0	22.48%	16.04%
ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	ACVAGGAAGT	1,00E-11	295.0	15.28%	10.17%
Tcf21(bHLH)/ArterySmoothMuscle-Tcf21-ChIP-Seq(GSE61369)/Homer	NAACAGCTGG	1,00E-11	257.0	13.31%	8.57%
Arnt:Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer	TBGCACGCAA	1,00E-11	146.0	7.56%	4.10%
Tlx?(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	CTGGCAGSCTGCCA	1,00E-10	160.0	8.29%	4.79%
SpIB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	AAAGRGGAAAGTG	1,00E-09	127.0	6.58%	3.58%
IRF1(IRF)/PBMC-IRF1-ChIP-Seq(GSE43036)/Homer	GAAAGTGAAAGT	1,00E-09	79.0	4.09%	1.84%
Ap4(bHLH)/AML-Tfap4-ChIP-Seq(GSE45738)/Homer	NAHCAGCTGD	1,00E-09	303.0	15.69%	10.96%
MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	AACAGCTG	1,00E-09	275.0	14.24%	9.77%
PAX5(Paired,Homeobox)/GM12878-PAX5-ChIP-Seq(GSE32465)/Homer	GCAGCCAAGCRTGACH	1,00E-09	116.0	6.01%	3.28%
Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Homer	GTCATGCHTGRCTGS	1,00E-08	124.0	6.42%	3.71%
EBF(EBF)/proBcell-EBF-ChIP-Seq(GSE21978)/Homer	DGTCCCYRGGGA	1,00E-08	100.0	5.18%	2.80%
KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	RGKGGGCGKGGC	1,00E-07	575.0	29.78%	24.23%
ETS:E-box(ETS,bHLH)/HPC7-Scl-ChIP-Seq(GSE22178)/Homer	AGGAARCAGCTG	1,00E-07	54.0	2.80%	1.20%
Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer	VNRVCAGCTGGY	1,00E-07	264.0	13.67%	9.89%
SPDEF(ETS)/VCaP-SPDEF-ChIP-Seq(SRA014231)/Homer	ASWTCCTGBT	1,00E-06	295.0	15.28%	11.37%
E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	DNRCAGCTGY	1,00E-06	426.0	22.06%	17.71%
HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	VCAGCTGBNN	1,00E-05	520.0	26.93%	22.54%
c-Myc(bHLH)/mES-cMyc-ChIP-Seq(GSE11431)/Homer	VVCCACGTGG	1,00E-05	112.0	5.80%	3.79%
Ptf1a(bHLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer	ACAGCTGTTN	1,00E-04	649.0	33.61%	29.31%
NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer	GCCATCTGTT	1,00E-04	193.0	9.99%	7.44%

NRF(NRF)/Promoter/Homer	STGCGCATGCGC	1,00E-04	28.0	1.45%	0.64%
n-Myc(bHLH)/mES-nMyc-ChIP-Seq(GSE11431)/Homer	VRCCACGTGG	1,00E-04	143.0	7.41%	5.37%
ZBTB18(Zf)/HEK293-ZBTB18.GFP-ChIP-Seq(GSE58341)/Homer	AACATCTGGA	1,00E-03	123.0	6.37%	4.50%
Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer	RCCACGTGGYYN	1,00E-03	134.0	6.94%	5.01%
PRDM1(Zf)/Hela-PRDM1-ChIP-Seq(GSE31477)/Homer	ACTTTCCTTTC	1,00E-03	208.0	10.77%	8.38%
MYB(HTH)/ERMYB-Myb-ChIPSeq(GSE22095)/Homer	GGCVGTTR	1,00E-03	291.0	15.07%	12.42%
AMYB(HTH)/Testes-AMYB-ChIP-Seq(GSE44588)/Homer	TGGCAGTTGG	1,00E-03	259.0	13.41%	11.06%

* Seuls les motifs significatifs sont montrés

Pour les GM12878

Motif Name	Consensus	P-value	# of Target Sequences with Motif (of 10817)	% of Target Sequences with Motif	% of Background Sequences with Motif
Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	NRYTTCCGGH	1e-1792	5045.0	46.64%	11.56%
ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	ACAGGAAGTG	1e-1740	4847.0	44.81%	10.88%
Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	NNAYTTCCTGHN	1e-1615	4464.0	41.27%	9.74%
ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	AACCGGAAGT	1e-1605	5333.0	49.31%	14.40%
ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	ACAGGAAGTG	1e-1582	5787.0	53.50%	17.29%
GABPA(ETS)/Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	RACCGGAAGT	1e-1494	4232.0	39.13%	9.35%
PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	AGAGGAAGTG	1e-1327	3043.0	28.13%	5.07%
EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	AVCAGGAAGT	1e-1214	4775.0	44.15%	14.32%
ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	ACVAGGAAGT	1e-1087	3347.0	30.94%	7.67%
Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	HACTTCCGGY	1e-1056	2801.0	25.90%	5.41%
EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	ATTCCTGTN	1e-1029	3194.0	29.53%	7.32%
ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	AVCCGGAAGT	1e-1014	2676.0	24.74%	5.12%
Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	NRYTTCCGGY	1e-972	2681.0	24.79%	5.36%
ETS(ETS)/Promoter/Homer	AACCGGAAGT	1e-916	2055.0	19.00%	3.19%
EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)/Homer	VACAGGAAAT	1e-894	2625.0	24.27%	5.57%
PU.1:IRF8(ETS:IRF)/pDC-Irf8-ChIP-Seq(GSE66899)/Homer	GGAAGTGAAAST	1e-886	1688.0	15.61%	2.09%
Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	MACAGGAAGT	1e-869	1823.0	16.85%	2.58%
SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	AAAGRGAAGTG	1e-840	1789.0	16.54%	2.57%
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	DATGASTCAT	1e-812	2492.0	23.04%	5.49%
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	NNATGASTCATH	1e-800	2205.0	20.39%	4.32%

* Seuls les 20 premiers motifs sont montrés

Pour les K562

Motif Name	Consensus	P-value	# of Target Sequences with Motif (of 3260)	% of Target Sequences with Motif	% of Background Sequences with Motif
ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	ACAGGAAGTG	1e-662	1696.0	52.02%	12.14%
Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	NRYYTCCGGH	1e-622	1702.0	52.21%	13.04%
GABPA(ETS)/Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	RACCGGAAGT	1e-615	1551.0	47.58%	10.63%
ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	AACCGGAAGT	1e-601	1859.0	57.02%	16.44%
EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	AVCAGGAAGT	1e-565	1780.0	54.60%	15.75%
Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	NNAYTTCCTGHN	1e-545	1487.0	45.61%	10.92%
ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	ACVAGGAAGT	1e-530	1309.0	40.15%	8.39%
ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	ACAGGAAGTG	1e-526	1893.0	58.07%	19.08%
ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	AVCCGGAAGT	1e-490	1124.0	34.48%	6.44%
PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	AGAGGAAGTG	1e-447	1017.0	31.20%	5.68%
ETS(ETS)/Promoter/Homer	AACCGGAAGT	1e-422	855.0	26.23%	4.03%
Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	HACTTCCGGY	1e-414	1061.0	32.55%	6.81%
Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	BBCTTATCTS	1e-395	950.0	29.14%	5.59%
Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	SAGATAAGRV	1e-380	889.0	27.27%	5.01%
Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	NBWGATAAGR	1e-379	1157.0	35.49%	8.94%
GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	AGATAASR	1e-360	1400.0	42.94%	13.74%
Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	NRYYTCCGGY	1e-345	967.0	29.66%	6.69%
EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	ATTTCTGTN	1e-313	978.0	30.00%	7.52%
Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	MACAGGAAGT	1e-309	630.0	19.33%	2.89%
SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	AAAGRGAAGTG	1e-294	624.0	19.14%	3.01%

* Seuls les 20 premiers motifs sont montrés

Table S7: Peak fold enrichment per motifs and expression

	# peaks	Mean FC	p-value
Motifs	ETS	560	7.26
	RUNX	254	6.58
	ETS+RUNX	392	7.53
	None	725	6.24
Expression correlated	74	9.25	-

* p-values are calculated using peak fold changes (FC) in each motif group vs the "expression correlated" peaks

Table S8: Motifs enrichment of ETS/RUNX motifs free *ETV6*-bound regions in Reh cells

La table complète sera téléchargeable avec la version en ligne de l'article.

Motif Name	Consensus	P-value	# of Target Sequences with Motif (of 725)	% of Target Sequences with Motif	% of Background Sequences with Motif
bZIP:IRF(bZIP,IRF)/Th17-BatF-ChIP-Seq(GSE39756)/Homer	NAGTTTCABTHTGACTNW	1,00E-24	108.0	14.90%	4.82%
E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	GGCGGGAARN	1,00E-20	117.0	16.14%	6.08%
Arnt:Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer	TBGCACGCAA	1,00E-10	69.0	9.52%	4.00%
PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	MGGAAGTGAAAC	1,00E-08	244.0	33.66%	24.11%
PRDM1(Zf)/Hela-PRDM1-ChIP-Seq(GSE31477)/Homer	ACTTTCAC TTTC	1,00E-07	108.0	14.90%	8.79%
IRF1(IRF)/PBMC-IRF1-ChIP-Seq(GSE43036)/Homer	GAAAGTGAAAGT	1,00E-06	40.0	5.52%	2.27%
NRF(NRF)/Promoter/Homer	STGCGCATGCGC	1,00E-05	13.0	1.79%	0.36%
Tlx?(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	CTGGCAGSCTGCCA	1,00E-05	59.0	8.14%	4.37%
KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	RGKGGGCGKGGC	1,00E-04	210.0	28.97%	22.28%
Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Homer	GTCATGCHTGRCTGS	1,00E-03	42.0	5.79%	3.44%

* Seuls les motifs significatifs sont montrés

Chapitre 4: Troisième article

Les modulateurs fonctionnels d'ETV6

4.1 Préface

Les altérations d'*ETV6* semblent jouer un rôle majeur dans l'étiologie de plusieurs désordres hématologiques. Par exemple, dans la majorité des cas de B-LAL, l'inactivation complète d'*ETV6* se fait par le biais de la t(12;21) suivie de la délétion de l'allèle résiduel. Or, certains cas semblent avoir conservé cet allèle résiduel d'*ETV6* mais affichent néanmoins un profil d'expression similaire à ceux l'ayant perdu, suggérant un mécanisme d'inactivation alternatif. Outre les événements délétères affligeant *ETV6* directement, il est possible que l'altération de certains gènes nécessaires à la régulation médiée par *ETV6* puisse affecter son réseau transcriptionnel de manière à contribuer à ces maladies hématologiques. L'identité de tels modulateurs d'*ETV6* demeure cependant largement inconnue.

L'identification de gènes pouvant moduler l'activité répressive d'*ETV6* dans des cellules leucémiques a été faite par un criblage fonctionnel à large échelle. L'impact des gènes les plus prometteurs a été évalué sur l'ensemble du réseau de régulation transcriptionnel d'*ETV6* défini à partir des études d'expression (Chapitre 2: Article 1) et de ChIP-seq (Chapitre 3: Article 2). Les gènes *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* et *SRP72* ont montré un effet important, généralisé et spécifique sur les cibles d'*ETV6*, supportant ainsi leur nouveau rôle comme

modulateurs fonctionnels d'*ETV6*. L'identification de ces modulateurs apporte une dimension jusqu'alors inexplorée à la régulation de l'activité d'*ETV6* et mène à plusieurs nouvelles pistes quant à leurs implications dans les pathologies humaines.

High throughput functional screening of *ETV6* transcriptional activity

Benjamin Neveu^{1,2}, Chantal Richer¹, Pauline Cassart¹, Maxime Caron¹, Camille Jimenez-Cortes^{1,3}, Pascal St-Onge¹, Claire Fuchs^{1,2}, Nicolas Garnier¹, Stéphane Gobeil^{4,5} & Daniel Sinnett^{1,6}

¹ Sainte-Justine University Health Center Research Center, Montreal, Qc, Canada

² Department of Biochemistry and Molecular Medicine, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada

³ Molecular biology program, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada

⁴ CHU de Québec-Université Laval Research Center, Quebec City, Qc, Canada

⁵ Department of Molecular medicine, Faculty of Medicine, Université Laval, Quebec City, Qc, Canada

⁶ Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada.

Corresponding author: Daniel Sinnett, Division of Hematology-Oncology, Sainte-Justine UHC Research Center, 3175 Côte Ste-Catherine, Montréal (Québec) Canada, H3T 1C5; Tel: (514) 345-4931/ext. 2990; Email: daniel.sinnett@umontreal.ca

Cet article sera soumis à Genome Biology.

4.2 Author contributions

DS is the principal investigator and takes primary responsibility for the paper. BN, CR, SG and DS contributed to the conception and design of the study. BN, CR and PC performed most of the experiments. MC analyzed the patients and TCGA expression data. CJC generated the data from the MOLT4 cell line. PSO was involved in the processing of the sequencing data. CF performed some of the Western blots. NG contributed to setting up the genome wide shRNA screening approach. BN performed data integration and analyses. BN, CR, SG and DS contributed to data interpretation. BN drafted the paper and SG and DS were involved in critical revision of the manuscript. All authors approved the final version.

4.3 Summary

Background: *ETV6* transcriptional activity is critical for proper blood cell development in the bone marrow. Despite the accumulating body of evidences linking *ETV6* malfunction to hematological malignancies, its regulatory network remains unclear.

Results: In an effort to uncover genes that modulate *ETV6* repressive activity, we performed a specifically designed, unbiased genome wide shRNA screen in pre-B acute lymphoblastic leukemia cells. After an extensive validation process, we identified 13 shRNAs that induced a significant overexpression of *ETV6* transcriptional target genes. Of them, silencing of *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* and *SRP72* led to a substantial abrogation of *ETV6* repressive activity. These genes were further shown to be associated with *ETV6*-mediated repression in various cancers.

Conclusions: Our work revealed critical modulators of *ETV6* function which could participate to cellular transformation through *ETV6* transcriptional network.

Keywords

Functional screen, shRNA, *ETV6*, transcription factor, acute lymphoblastic leukemia, regulatory network

4.4 Background

ETV6 is a member of the ETS superfamily of transcription factors [1] located on chromosome 12p13. ETS factors have been associated to multiple biological processes and pathways with, among others, critical implications in hematopoiesis [2]. Their dysregulation correlates with various hematological diseases such as lymphomas, leukemia and myelodysplastic syndrome. Chromosomal abnormalities of the 12p13 band are frequently found in these malignancies and so far more than 30 fusion partners have been identified in *ETV6*-derived translocations [3]. The mechanisms through which these fusions induce transformation are highly variable and depend on the function of the partner as well as the retained domains of *ETV6*. In several cases however, an expected impact of the translocation is to inhibit *ETV6* normal function as proposed for the t(1;12)(q21;p13) [4] and t(12;21)(p13;q22) [5-7] translocations.

The t(12;21)(p13;q22) translocation is the most frequent chromosomal abnormality (together with hyperdiploidy) to be associated with childhood B-cell acute lymphoblastic leukemia (ALL) [8-10]. It fuses all functional domains encoded by the *AML1* (*RUNX1*) gene to the first 5 exons of *ETV6* encoding its Pointed (PD) and repression domains. Despite the high recurrence of t(12;21) in B-cell ALL and the systematic expression of the *ETV6*-*AML1* fusion protein in t(12;21) positive leukemic patients [11, 12], the t(12;21) was demonstrated to be insufficient to induce leukemic transformation [13-15]. Interestingly, the non-rearranged residual allele of *ETV6* is afflicted by deleterious events in virtually every cases of t(12;21) positive ALL [12, 16-18], suggesting that *ETV6*

inactivation could indeed play a critical role in leukemogenesis. Moreover, several groups recently reported germline mutations in *ETV6* that were associated with familial hematological disorders [19-22]. Several of these mutations were located in the ETS DNA-binding domain of *ETV6* and impaired its nuclear localisation and/or repressor activity. These findings strongly support a role for *ETV6*-mediated transcription in these malignancies.

Despite the consequences of impaired *ETV6* function, very little is known on its regulatory network. To negatively regulate transcription, *ETV6* relies on known co-repressors such as the N-CoR and mSin3A complexes and the histone deacetylase *HDAC3* [23-25]. *ETV6* ability to interact with *HDAC3* was shown to be modulated through its binding with the *IRF-8* transcription factor [26]. This interaction was further associated to a change in *ETV6* binding and transcriptional activity. Such findings indicate that *ETV6*-mediated transcription can be regulated through complex interactions. However, these studies were conducted through candidate gene or interaction-based approaches and did not cover the full intricacy of *ETV6* function.

Small hairpin RNAs (shRNAs) are widely used to silence gene expression [27]. Accessibility to genome wide shRNAs libraries and advances in sequencing technologies enabled the high throughput interrogation of genes functions [28]. In this study, we designed a cellular screening system to functionally interrogate the impact of over 140,000 shRNAs on *ETV6*-mediated repression. A panel of high priority genes selected from the screen were further examined for their ability to efficiently regulate *ETV6* activity by monitoring its transcriptional network in pre-B

ALL cells [29] [Article 2]. Validation experiments successfully unveiled key modulators of *ETV6* repressive function in leukemic cells as well as in other cancers.

4.5 Results

4.5.1. Generating *ETV6*-dependent Blasticidin sensitive cells

To gain insights on the network of *ETV6* modulators, we modified a pre-B ALL cell line (Reh; **Figure 1**) in which the Blasticidin resistance gene (BlastR) is stably integrated and expressed through an artificial *ETV6*-responsive promoter (EBS₃tk) composed of 3 tandem EBS sites [30] followed by the HSV-TK promoter. Reh cells lack endogenous *ETV6* expression as a result of a t(12;21)(p13;q22) translocation and an additional 12p13 deletion. Because of the absence of functional endogenous *ETV6*, EBS₃tk is active and promotes Blasticidin resistance in this modified cell line. Clones were generated from the Blasticidin resistant cells and *ETV6* was re-introduced through lentiviral transduction. *ETV6* is expected to bind and represses the EBS₃tk promoter and consequently reduces BlastR expression. As shown in **Figure 2A**, BlastR expression in Reh EBS₃tk BlastR clone 2-1 is significantly reduced upon *ETV6* expression compared to the mock and pLenti control clones (2.7 fold compared to pLenti; Student's *t* test p-value ≤ 0.001). We derived a final clone from the Reh EBS₃tk BlastR clone 2-1 expressing *ETV6* to obtain a robust and homogenous cellular system which we refer to as clone 2-1 ETV6-5.

Although each interrogated Reh EBS₃tk BlastR clone significantly expressed *ETV6* after transduction (**Figure 2A**, inset), the inconsistent effect on BlastR repression suggests that the chromatin context where EBS₃tk BlastR inserted is rather influential. To identify the integration site of EBS₃tk BlastR in the *ETV6*-

responsive clone 2-1, we used data from a previous *ETV6* chromatin immunoprecipitation sequencing experiment (ChIP-seq) performed in these cells [Article 2]. When mapped on the specific plasmid sequence, several reads were found in the EBS₃tk promoter, suggesting that *ETV6* is recruited to the promoter region (Additional file 1). One read contained the partial sequence of the EBS₃tk promoter merged with a sequence from the first intron of the *NEGR1* gene, possibly indicating the integration site (Figure 2B and Additional file 1). PCR analysis of the Reh EBS₃tk BlastR clone 2-1 confirmed the integration site of EBS₃tk BlastR in the intron 1 of *NEGR1* (Figure 2C). Based on previous transcriptome data [29], this integration did not impact *NEGR1* expression compared to non-modified Reh cells (53.1 and 51.3 FPKM, respectively). Furthermore, *NEGR1* is not significantly modulated by *ETV6* re-expression (logFoldChange = -0.46, FDR = 0.50). Together, these results indicate that EBS₃tk BlastR insertion at an active locus enables stable BlastR expression, accessibility to *ETV6* and concomitant *ETV6*-mediated gene repression through the EBS₃tk promoter.

To screen for putative *ETV6* modulators, *ETV6*-mediated repression of BlastR should be reversible. We thus transduced the 2-1 ETV6-5 clone with a shETV6-GFP and a scETV6-GFP negative control. The GFP positive fraction of cells was retrieved and *ETV6* knockdown confirmed at the protein level (Figure 2D). As expected, BlastR expression was positively modulated by *ETV6* knockdown (Figure 2D). Additionally, nicotinamide (NAM), a chemical inhibitor of epigenetic repression [31-33], also increased BlastR expression in the 2-1 ETV6-5 clone

(**Figure 2E**), further indicating that the repression is reversible. Finally, impinging on *ETV6* repressor function should promote cell survival and proliferation of our clone in Blasticidin (BlastS) conditions. Accordingly, BlastR overexpression observed in shETV6-GFP cells (**Figure 2D**) considerably increased their ability to proliferate in the presence of Blasticidin (**Figure 2F**). Altogether, these results indicate that the Reh EBS₃tk BlastR 2-1 ETV6-5 clone features a selectable read-out of *ETV6* activity.

4.5.2. Identification of putative *ETV6* modulators through a genome wide shRNA screen

We designed a genome wide loss-of-function screen to identify *ETV6* modulators in the Reh EBS₃tk BlastR 2-1 ETV6-5 clone (**Figure 1** and [Additional file 2](#)). Cells were first transduced with the complete MISSION lentiviral shRNA library (~140,000 unique shRNAs divided in 15 pools). Those expressing shRNAs impairing *ETV6* repressive activity are expected to have increased BlastR expression and were thus selectively enriched through Blasticidin treatment. Deep sequencing allowed the identification of 126,042 unique integrated shRNAs which were evenly distributed among the 15 pools ([Additional file 3](#)). Using a series of thresholds (showed in **Figure 3A** and described in the *methods* section), we retained 2,858 positive shRNAs ([Additional file 3](#)) that were over-represented in the BlastS condition at both time points compared to their respective control population without selection. shRNAs with a non-significant over-representation (i.e. $\sum \text{Counts}(\text{BlastS vs } \emptyset \text{BlastS}) \leq 9,770$; see *methods* section) as well as those targeting unexpressed genes (FPKM ≤ 0.21 ; [34]) in Reh

cells [29] were discarded. At the end, we retrieved 1,241 shRNAs (i.e. selected shRNAs) that promote resistance to Blastcidin (**Figure 3B**, [Additional file 3](#) and [Additional file 4](#)). As an essential positive control, we found that a shRNA against *ETV6* was enriched in the BlastS conditions (**Figure 3C**). 1,075 genes were covered by a single shRNA (**Figure 3A** and [Additional file 4](#)) and, with few exceptions (see below), were not further considered because of potential off-target effects. Nonetheless, 4 and 77 genes were targeted by 3 and 2 distinct shRNAs, respectively (**Figure 3A**, [Additional file 4](#) and [Additional file 5](#)). Although multiple shRNAs against a particular gene did not systematically led to the same magnitude of selective advantage (exemplified in **Figure 3D, E**), possibly because of variable knockdown efficiencies, their impact are less likely driven by off-target effects and thus were prioritized as part of a discovery approach. Furthermore, 33 of these 81 genes were targeted by shRNAs from different pools (for example *TACC1* and *CAPN7*, **Figure 3D, E** respectively; see also [Additional file 5](#)), suggesting that their silencing conferred Blastcidin resistance independently of the competitive background, making them even stronger candidates.

To include a maximum of putative *ETV6* modulators for further validation, we first assigned the most potent shRNA of a given pair or trio to the greater difference in normalized counts between Blastcidin and control conditions (**Figure 3F, G**). Interestingly, although the differences between the best and second best shRNAs are mostly attributed to their representations in Blastcidin conditions (T_{28} BlastS and T_{42} BlastS), the best shRNAs are better represented in control

conditions (T_{28} and T_{42}) as well as in their initial populations prior to Blasticidin selection (T_{14}) (**Figure 3F**). This suggests that potent shRNAs that are weakly represented at the time of selection are disadvantaged and less susceptible to emerge, especially in such a strong selective environment. To attenuate these biases, the difference in normalized counts between Blasticidin and control conditions for each shRNA were corrected for their representation at T_{14} (**Figure 3H**). The relationship between the best and second best shRNAs is indeed less divergent after implementing this correction (compare **Figure 3H** and **Figure 3G**), indicating that 2 shRNAs targeting the same gene are likely to promote a similar impact on Blasticidin resistance. Accordingly, shRNAs from a pair or trio that were selected for further validation are those with the strongest T_{14} -corrected difference between Blasticidin and control conditions (see [Additional file 5](#)).

To compose with the off-target effects occurring among the 1,075 single shRNAs, we used a complementary approach based on structure predictions [35] to identify the potential partners of the 81 putative modulators confidently identified above (**Figure 3A**). Of the 136 proteins that were predicted to interact with 3 or more of these modulators, 9 were identified in the screen with a single shRNA ([Additional file 6](#)), indicating that they may be involved in the same pathways and are thus more likely to modulate *ETV6* function. These genes, together with the top 81 candidates, were thus included for further validation (total of 90 genes; [Additional file 5](#)).

4.5.3. Validation of putative *ETV6* modulators on known endogenous transcriptional targets

To validate the impact of these putative modulators on *ETV6*-mediated regulation, we constructed distinct Reh *ETV6* cell lines expressing a shRNA against each of these candidates and evaluated *ETV6* endogenous transcriptional network [29] by targeted RNA sequencing. Of the 84 putative *ETV6* target genes initially included in the design, 40 were confirmed as their expression was significantly reduced in Reh *ETV6* samples compared to Reh pLENTI control cells (**Figure 4A** and [Additional file 7](#)). Furthermore, 22 of them have an associated *ETV6* ChIP-seq signal [Article 2] and are thus considered as direct *ETV6* targets ([Additional file 7](#)). Inversely, the expression of 64 genes (from the apoptosis panel) was shown to be independent of *ETV6* in our previous transcriptome analysis [29] as well as in the present targeted RNA-seq experiment (**Figure 4B** and [Additional file 7](#)). As expected, genes in this group were also depleted in *ETV6* ChIP-seq signals [Article 2]. Only 2 of them were in fact associated with a peak, therefore the expression of the remaining 62 genes was used as negative controls ([Additional file 7](#)).

We first evaluated the number of *ETV6* targets (n=22) with increased expression in the Reh *ETV6* shRNA samples compared to the negative shRNA controls (SHC002 and SHC202). As shown in **Figure 4C**, virtually all shRNAs increased the expression of at least one *ETV6* target, suggesting that they could have been retrieved in the screen by abrogating *ETV6*-mediated repression of the BlastR gene. However, none of them affected the entire group of *ETV6* targets,

indicating that *ETV6* modulators could act in a gene-specific manner. To ensure that the candidate modulators are not broad modifiers of gene transcription but instead act specifically on *ETV6*, we evaluated their impact on *ETV6*-independent genes (n=62). Accordingly, 24 shRNAs led to a specific overexpression of *ETV6* targets compared to *ETV6*-unrelated genes (**Figure 4D**), indicating that their effects on gene expression are attributed to impaired *ETV6* repressive function.

The expression of the shRNAs-targeted genes was then evaluated. Of the 85 informative amplicons, 63 genes have confirmed reduced expression in their corresponding shRNA sample (**Figure 4E**). The expression of a target gene was not systematically the lowest in their corresponding shRNA sample. For instance, the expression of *COMMD9* was relatively stable among the 85 tested samples but greatly reduced in the sh*COMMD9* sample (**Figure 4F**), whereas *KLF5* expression was lowered within the sh*KLF5* sample but also in several other samples (**Figure 4G**). The shRNAs were thus categorized into 3 groups based on their silencing (**Figure 4H**). The "Best knockdown" group (n=31) led to the lowest expression of their target gene among all samples (as sh*COMMD9*, **Figure 4F**). These shRNAs gave a mean silencing of 71% (0.29 residual expression). The "Knockdown" group (n=32) silenced their target gene as well but without being the sample with the lowest expression (e.g. sh*KLF5*, **Figure 4G**). The effects of such shRNAs on *ETV6*-modulated genes could not be confidently associated to their target gene. Finally, the "No knockdown" group (n=22) failed to repressed its target gene.

Of the 24 shRNAs that significantly overexpressed *ETV6* targets (**Figure 4D**), 5 have unconfirmed knockdown, 18 have a knockdown, of which 12 are in the "best knockdown" group and 1 has a non-informative amplicon. The knockdown in this sample was however confirmed by qPCR and considered validated (*DYRK4*, 73% knockdown; [Additional file 8](#)). Therefore, we identified 13 shRNAs that efficiently knocked down their target genes and impaired *ETV6* transcriptional activity (**Figure 4I, J**). The overexpression levels and the number of *ETV6* targets affected by the silencing of these putative modulators are however variable (**Figure 4J, K**). To statistically determine the most influential *ETV6* modulators, we performed a paired Student's *t* test with the expression of *ETV6* targets in each shRNA sample compared to SHC controls. Among the 13 top shRNAs, 5 of them (shAKIRIN1, shCOMMD9, shDYRK4, shJUNB and shSRP72) show a significant p-value for this comparison, indicating that they greatly impact the expression of the *ETV6* targets group (**Table 1**). Moreover, an unsupervised clustering analysis based on the expression of the 22 *ETV6* targets revealed the broad impact of *AKIRIN1*, *DYRK4* and *SRP72* depletion on *ETV6* function as these shRNA samples clustered in a restricted hub together with the Reh pLENTI sample lacking *ETV6* expression (**Figure 4K**).

4.5.4. *ETV6*-modulated transcription of t(12;21)-associated genes

Leukemic cells from t(12;21) positive patients have a distinct and clear expression profile [36]. This molecular signature can be, at least partly, explained by the absence of *ETV6* expression as some of its target genes are specifically overexpressed in these cells [29]. Interestingly, 9 of the 22 *ETV6* target genes

included in the validation process were identified as specifically overexpressed in t(12;21) positive patients [29] (those presented in **Figure 5A** and [Additional file 9](#)). These genes, in addition to serve as an *ETV6* transcriptional readout, could possibly contribute to leukemogenesis.

We thus focused on the impact of *ETV6* modulators on the expression of these 9 t(12;21)-associated *ETV6* targets. First, a similar unsupervised clustering analysis showed that 4 modulators, *AKIRIN1*, *DYRK4*, *SRP72* and *JUNB* were indeed essential for their regulation by *ETV6* as these shRNA samples clustered with the pLENTI sample ([Additional file 9](#)). Although *JUNB* seemed not involved in *ETV6*-mediated repression to the same extent as the other top modulators (**Figure 4K**), the targets it regulates are enriched for t(12;21)-associated overexpressed genes. Of the 5 top modulators (**Table 1**), only *COMMD9* shows a moderate impact on t(12;21)-associated *ETV6* targets ([Additional file 9](#)).

After confirming that the 9 t(12;21)-associated *ETV6* targets were downregulated in Reh *ETV6* compared to Reh wild-type (WT) cells (**Figure 5A**), we silenced each of the top *ETV6* modulators in both cellular background (**Figure 5B**) and further evaluated their impact on gene expression (**Figure 5C** and [Additional file 10](#)). As expected (see above; [Additional file 9](#)), the silencing of the *ETV6* modulators *AKIRIN1*, *DYRK4*, *JUNB* and *SRP72* upregulated most *ETV6* targets whereas *COMMD9* knockdown had less impact (**Figure 5C** and [Additional file 10](#), Reh *ETV6*). Of note, *OSTalpha* is the only *ETV6* target to be overexpressed by the silencing of the 5 modulators. Even if similar knockdowns were obtained in Reh *ETV6* and Reh WT cells (**Figure 5B**), the overexpression of the t(12;21)-

associated *ETV6* target genes was mostly restricted to Reh *ETV6* cells (**Figure 5C** and [Additional file 10](#), Reh *ETV6* vs Reh WT), indicating that the impact of modulators on gene expression is not attributed to an *ETV6*-independent effect. The upregulations observed are indeed due to a lack of *ETV6*-mediated repression and further strengthen the relationship between the identified modulators and *ETV6* function.

We next investigated the expression status of the *ETV6* modulators in our pediatric leukemia samples. No clear association could be made between the expression of *AKIRIN1*, *COMMD9*, *DYRK4* and *JUNB* and t(12;21)-associated *ETV6* targets (data not shown). However, the expression of *SRP72* was significantly lower in a t(12;21) negative sample (**Figure 5D**; upper panel). Strikingly, the expression of the *ETV6* target *OSTalpha* was consequently upregulated in that same patient (**Figure 5D**, lower panel). This result strengthens the role of *SRP72* in *ETV6* repressive function in pre-B ALL patients.

4.5.5. Implications of *ETV6* and its modulators in various cancers

ETV6 transcriptional function is well associated with childhood pre-B ALL [10] but remains elusive for other cancers. We thus investigated *ETV6*-mediated transcription in several cancers with a particular attention on its newly identified modulators.

We evaluated the expression of known *ETV6* targets in our pre-T ALL patients. Interestingly, one of them showed a significant overexpression of *CLIC5* (**Figure 6A**, arrow), a known *ETV6* target gene strongly upregulated in t(12;21) positive

pre-B ALL patients (**Figure 6A**, t(12;21); [29]). Strikingly, an *ETV6* deletion was found in this patient [37], suggesting that *CLIC5* overexpression is attributed to *ETV6* depletion. Accordingly, we used the pre-T leukemic cell line MOLT4 to confirm that *ETV6* effectively modulates the expression of *CLIC5* in this context. As shown in **Figure 6B**, the expression of *CLIC5*, but not the other t(12;21)-associated *ETV6* targets, is significantly upregulated when *ETV6* is knocked down (**Figure 6B**, inset), suggesting that *CLIC5* is indeed a target of *ETV6* in pre-T cells. Next, we evaluated the implication of the 5 top *ETV6* modulators on *CLIC5* expression. shRNA-mediated silencing was validated for each modulator ([Additional file 11](#)) and knockdowns were consequent with those obtained in Reh cells (**Figure 5B**). Remarkably, repression of all 5 *ETV6* modulators induced a significant upregulation of *CLIC5* in MOLT4 cells (**Figure 6C**) without any marked downregulation of *ETV6* expression ([Additional file 11](#)). These results strongly support the role of these genes as modulators of *ETV6*-mediated repression and suggest that *ETV6* modulators might be conserved in various cellular contexts.

To further assess the impact of *ETV6* modulators in other cancers, we interrogated gene expression data from The Cancer Genome Atlas (TCGA database). Notably, only very few known pre-B ALL *ETV6* targets were negatively correlated with the expression of *ETV6* in other cancers (e.g. glioblastomas and gliomas, [Additional file 12](#), arrows), indicating that *ETV6* targets are rather tissue specific. Nevertheless, these conserved *ETV6* target genes ([Additional file 12](#), arrows) were also negatively correlated with the expression of several modulators, supporting a role for these modulators in

ETV6-mediated repression in other contexts. To circumvent the cancer specificity issue, genes with a strong negative correlation with *ETV6* expression were identified in each cancer. Strikingly, very strong associations were observed for several *ETV6* modulators and the expression of cancer-specific *ETV6*-correlated genes (e.g. glioblastomas and gliomas in **Figure 7** and breast cancer in [Additional file 13](#)). Inversely, such correlations were absent on genes that showed no dependency to *ETV6*, indicating that the correlations with the modulators were indeed conditional to *ETV6*-correlated expression. Moreover, the expression of the modulators did not correlate with the others nor with the expression of *ETV6* itself, suggesting that the correlations observed between a given modulator and *ETV6*-modulated genes are rather specific (**Figure 7** and [Additional file 13](#), represented by the diagonal in the modulator group). Finally, as for *ETV6* target genes, the impact of its modulators also seemed to be cancer specific. This analysis showed that both the number and the identity of the modulators associated with *ETV6*-correlated genes expression were highly variable among cancers. For instance, *SRP72* showed good associations with *ETV6*-correlated genes expression in glioblastomas and gliomas patients whereas no correlation was observed in breast cancer patients (**Figure 7** and [Additional file 13](#), respectively). On the contrary, *STIM1* was associated with *ETV6*-correlated genes expression only in breast cancer patients. Of note, the expression of *SRP72* and *STIM1* is similar between these cancers (not shown) and thus could not explain the divergences observed on *ETV6*-correlated genes expression. All together, these results demonstrate the strong interactions

between *ETV6*, its modulators and *ETV6*-correlated genes and highlight the critical role of these modulators on *ETV6* transcriptional function.

4.6 Discussion

ETV6 loss of function seems to be a determinant event in the development of hematological malignancies [19-22, 38]. Herein, we designed a cellular system in which we conducted a functional genome wide shRNA screen that uncovered an important number of genes involved in *ETV6* transcriptional activity. After extensive validations, we confidently identified 13 novel genes that participate in *ETV6*-mediated repression, of which *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* and *SRP72* demonstrated a broad impact on *ETV6* transcriptional network in leukemic cells. Alterations in such modulators would disrupt *ETV6* function and participate in hematological diseases.

The primary screen for *ETV6* modulators was done through an *ETV6*-responsive artificial promoter (EBS₃tk) integrated at a specific genomic locus. This experimental setting might not faithfully replicate the genomic environment of *ETV6* endogenous targets and to better evaluate the impact of the putative modulators on *ETV6* function, an important validation procedure was performed using the complete *ETV6* transcriptional network in Reh cells [29] [Article 2]. By including multiple *ETV6* target genes as well as negative control genes, each modulator was interrogated with regards to their specificity and range of effects over *ETV6* activity. We thus identified 13 modulators whose impact on transcription was restricted to *ETV6* targets. Of note, certain chromatin remodelers and common co-repressors that may be required for *ETV6* function could have been deceitfully discarded as their silencing is expected to promote the expression of unrelated genes as well. For instance, *CBX3* (or HP1 γ), a

known heterochromatin associated protein [39], was captured by 3 unique shRNAs in the primary screen but was not retained at this point as the knock-down of this gene led to the overexpression of 8 of the 22 *ETV6* targets but also of 7 of the 62 control genes (Student's t test p-value = 0.15). We also observed that most of the modulators seemed implicated in the regulation of a fraction of *ETV6* targets, highlighting the complexity of *ETV6*-mediated repression. However, some of the modulators appeared to be more widely involved in *ETV6* function. *ETV6* activity greatly depends on *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* and *SRP72* and are thus interesting genes regarding hematological diseases.

Interestingly, our results revealed that not all *ETV6* targets responded similarly towards the identified modulators. For example *DSC3*, a gene significantly repressed by *ETV6* and overexpressed in t(12;21) positive leukemic patients, seems to be mostly independent of the modulators. Inversely, *OSTalpha* showed the same dependency to *ETV6* but was significantly overexpressed following the silencing of 9 of the 13 modulators. *ETV6* requirement for certain modulators is perhaps important for the regulation of a specific transcriptional response following a given signal. For instance, *SEMA3A* and *SEMA3D*, two members of the Semaphorins Class III gene family, are both regulated by *ETV6* with the involvement of the same modulators. The specificity given to *ETV6* by its modulators may ensure that genes involved in the same pathway are adequately co-expressed. With this rationale, *ETV6* target genes that share similar dependencies towards modulators could be suggestive of an implication in a common pathway. Further investigation of *ETV6* targets functions would indeed

be facilitated. To date however, no associations can easily be made between most of *ETV6* target genes and their impact on leukemogenesis remain elusive.

With few exceptions (e.g. *CLIC5*), *ETV6* transcriptional targets seems highly variable across tissues. Aside from *ETV6* modulators and the peculiar chromatin environments, several other mechanisms may lead to *ETV6* specificity and contribute to such disparities. Among others, the ETV6-AML1 fusion protein expressed in Reh cells is known to interact with residual *ETV6* [7], modulate its binding [Article 2] and therefore dictate specificity. Although most *ETV6* targets were upregulated by the silencing of *ETV6* modulators specifically in an *ETV6* positive background, *NTRK1* behave differently. In fact, *NTRK1* expression was also upregulated in Reh WT cells following the silencing of *ETV6* modulators (e.g. shSRP72 induced a 9.92 and 6.68 fold overexpression of *NTRK1* in Reh *ETV6* and Reh WT cells, respectively). Interestingly, *NTRK1* associated ChIP-seq signal does not contain the *ETV6* consensus binding site [Article 2], suggesting that *ETV6* acts as a co-repressor at this locus. It is possible that ETV6-AML1 or other repressive ETS factors [1] endogenously repress *NTRK1* in Reh WT cells and that *ETV6* overexpression strengthens this repression (*ETV6* downregulates *NTRK1* in Reh cells). Because ETV6-AML1 retains both the PD and repression domains of *ETV6* and the repressive ETS factors share a similar DNA-binding domain [40], the identified modulators could potentially affect their repressive function as well, explaining *NTRK1* upregulation in Reh WT cells following the silencing of *ETV6* modulators. This is of great interest as *ETV6* is involved in multiple translocations with various fusion partners [3] and that ETS

factors are widely involved in human malignancies [2, 40-42]. Additional experiments will be needed to determine the implication of the novel *ETV6* modulators on other related proteins.

Certain modulators demonstrated a critical impact on *ETV6* activity. While their mechanisms of action in *ETV6*-mediated transcription remain to be established, several lines of evidences regarding their role in transcriptional regulation are worth to mention. Although *AKIRIN1* function in human cells remains largely unknown, it had been found to participate in transcription-associated chromatin remodeling in *Drosophila melanogaster* [43, 44]. *AKIRIN1* impact on gene expression was however highly restricted to *ETV6* targets, either suggesting an alternative function in human cells or an *ETV6*-specific chromatin remodeling activity. *COMMD9* was recently associated to *TFDP1/E2F1* transcriptional activity in lung cancer [45], indicating that it could take part in transcription regulation. Moreover, two other *COMMD* genes, *COMMD8* and *COMMD10*, were also retrieved from the genome wide screen with a single shRNA, suggesting that this gene family could be implicated in *ETV6* function. Interestingly, this whole family is known to be involved in NF- κ B-mediated regulation of transcription [46]. As for NF- κ B, it is possible that *COMMD* proteins could regulate *ETV6* transcriptional activity. The most significant overexpression of *ETV6* target genes was achieved when *DYRK4* was knocked down. Unfortunately, *DYRK4* is the less studied member of the serine/threonine dual specificity kinase family [47, 48] and no role in transcriptional regulation had been associated to this gene so far. Notably, another member of this family, *DYRK3*,

was also identified in the primary screen with a very potent shRNA. Interestingly, this member is known to phosphorylate histone H2B [48] and *SIRT1*, a protein deacetylase that targets histones and other transcription factors such as *E2F1* and NF- κ B [49]. Perhaps *DYRK3* (and maybe *DYRK4*) role in *ETV6*-mediated repression is associated with the activation of the histone deacetylase activity of *SIRT1*, thus promoting transcriptional repression. These interactions are further supported by the capture of a shRNA against *SIRT3* and *SIRT4* in the primary screen. Alternatively, *ETV6* phosphorylation is known to modulate its transcriptional function [50-52] and *DYRK3* and 4 might play a role in this process. Determining whether these kinases directly phosphorylate *ETV6*, its modulators or participate in *ETV6*-mediated repression through histones modifications would be of great interest because of their substantial impact on *ETV6* function. *JUNB* is a subunit of the AP-1 transcription factor which have been extensively studied in the context of multiple lymphoid malignancies such as lymphomas and T-cell leukemias [53]. *FOSL2*, another subunit of AP-1, was also captured in the screen with a single shRNA, suggesting that AP-1 might participate in *ETV6* function. Interestingly, although *ETV6* itself was not shown to bind AP-1, physical interactions between ETS and AP-1 factors have been described [54, 55]. These interactions were dependent on the conserved ETS domain, suggesting that similar interactions could potentially occur with *ETV6* and modulate its transcriptional activity.

Lastly, *SRP72* is potentially the most interesting *ETV6* modulator in the hematological context. One of our t(12;21) negative patient sample showed

approximately 50% less *SRP72* expression compared to the whole group, suggesting monoallelic expression. Expression of the *ETV6* target *OSTalpha* was also significantly higher in this leukemia patient, further supporting that *SRP72* downregulation impact *ETV6* repressive function. Furthermore, mutations in this gene have been associated to familial aplasia and myelodysplasia [56] both of which are also associated with *ETV6* mutations or deletions [20]. The mechanisms through which *SRP72* mutations induce these familial forms of aplasia remain largely unknown but, with regards to our findings, could involve deregulation of *ETV6* target genes. Interestingly, in addition to *SRP72*, *SRP14* was also found with a single shRNA in the genome wide screen and *SRP9* was identified as an *ETV6*-bound protein by a proteomic approach ([Additional file 14](#)). These results suggest that a complex formed with SRP proteins could bind *ETV6* to regulate its transcriptional output. SRP proteins canonical function is to form a complex that binds nascent peptide signals, induce translational arrest and transfer the ribosome to the translocon onto the endoplasmic reticulum where the translation would resume [57]. However, the quantity and localisation of the *ETV6* protein remain normal after the silencing of *SRP72* ([Additional file 15](#)), suggesting that *SRP72* may play a role in *ETV6*-mediated repression through an alternate pathway. Interestingly, *SRP72* have been associated to a novel transcriptional function through the binding of histone H4 [58]. Additional experiments should be performed to determine the functional interactions between SRP proteins and *ETV6*-mediated transcription and their implications in hematological malignancies

4.7 Conclusions

ETV6 repressive function has been extensively interrogated through a specifically designed genome wide shRNA screen. The individual assessment of a large number of putative modulators and their impact on *ETV6* transcriptional network confirmed the validity of this high throughput methodology and unveiled the high complexity of *ETV6*-mediated repression. Novel modulators of *ETV6* activity were thus identified and associated to the regulation of *ETV6* target genes in various contexts. Although most of the modulators were involved in a target-specific manner, some others were demonstrated to have a broad impact on *ETV6* transcriptional network. The critical role of *AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* and *SRP72* on *ETV6*-mediated regulation makes them interesting genes regarding the etiology of hematological diseases.

4.8 Methods

4.8.1. Constructs

The luciferase gene from the pGL4.10 plasmid (Promega) was excised by NcoI and XbaI enzymatic digestion. The same restriction sites in pGL4.10 were then used for the ligation of the Blastidicin resistance gene (BlastR) amplified by PCR ([Additional file 16](#)) from the pLENTI CMV TetR Blast (716-1) (Addgene). The EBS₃tk artificial promoter was amplified by PCR ([Additional file 16](#)) from the pGL3 EBS₃tk vector and subcloned in the multi cloning site of pGL4 BlastR using KpnI and NheI restriction enzymes to obtain the final pGL4 EBS₃tk BlastR.

The complete *ETV6* complementary DNA (cDNA) sequence (NM_001987) obtained from pcDNA3.1 *ETV6* was subcloned into the Gateway compatible vector pENTR3C and transferred into the pLENTI CMV puro DEST (w118-1) lentiviral vector (kindly provided by Dr. Christian Beauséjour) using LR clonase II reactions according to the manufacturer protocol (Thermo Fisher Scientific).

4.8.2. Lentiviral production and transduction

Production of lentiviral particles and cell transduction were performed as previously described [29]. Briefly, HEK293T cells were transfected with the lentiviral expression plasmid (pLENTI backbone) together with the third generation encapsidation plasmids, pRSV-Rev, pMD2.VSVG and pMDL (kindly provided by Dr. Christian Beauséjour) using Polyethylenimine (Polysciences). Media was renewed 16h post-transfection and cells were cultured for 30h. Viral particles from culture media were concentrated by ultracentrifugation (50 000 g)

and estimated using the p24 antigen ELISA (Advanced Bioscience Laboratories). The transduction was performed by adding these particles (10-100 ng per 1×10^5 cells) and 8 $\mu\text{g}/\text{mL}$ polybrene (Sigma) to cells for 24h. Media was changed and the selection agent (i.e. 1 $\mu\text{g}/\text{mL}$ puromycin) was added 48h post-transduction. Cells were cultured for at least 2 weeks prior to any experiments. (*Please refer to the "shRNAs preparation" section for the pLKO shRNAs particles.*)

4.8.3. Establishment and control of the screening cell line

The Reh (ATCC ® CRL-8286™) cell line and derived clones were maintained in RPMI 1640 (Wisent) 10% Fetal Bovine Serum (FBS; Wisent) in a 5% CO_2 incubator at 37°C. Reh cells were transfected with pGL4 EBS₃tk BlastR using the Lipofectamine 2000 reagent (Thermo Fisher Scientific). Media was changed the next day and 2 $\mu\text{g}/\text{mL}$ Blasticidin was added 48h post-transfection for selection. Cells were maintained for a month under Blasticidin selective pressure to retrieve cells with a stable integration. Using 6-wells plates, 500 cells from this Blasticidin resistant population were seeded in 1 mL of MethoCult H4230 media (STEMCELL Technologies) containing 2 $\mu\text{g}/\text{mL}$ Blasticidin and grown for an additional month to obtain distinct colonies (clones). These colonies were transferred and amplified in 96-wells liquid cultures with 2 $\mu\text{g}/\text{mL}$ Blasticidin. Blasticidin resistant clones were then transduced with pLENTI *ETV6* or pLENTI empty vector lentiviral particles. Cells were selected with 1 $\mu\text{g}/\text{mL}$ puromycin before assessing *ETV6* and BlastR expression by quantitative real-time PCR (qRT-PCR; see related section). The most responsive clone to *ETV6* expression (i.e. Reh EBS₃tk BlastR 2-1) was submitted to an additional round of clonal

selection in methylcellulose media (as above; with 1 $\mu\text{g}/\text{mL}$ puromycin) to obtain a strict and homogenous cellular system with single EBS₃tk BlastR and *ETV6* integrations (i.e. Reh EBS₃tk BlastR 2-1 ETV6-5).

A GFP selectable pLKO plasmid containing a shRNA against *ETV6* was used for lentiviral production and transduction of the *ETV6*-dependent Blasticidin sensitive clone 2-1 ETV6-5. A scrambled shRNA sequence was also included as a negative control. GFP positive cells were retrieved by flow cytometry using the FACSAria IIu cell sorter (Becton Dickinson) and stabilized in culture for several days before validating the *ETV6* knockdown by Western Blotting (see related section) and qRT-PCR (see related section). GFP positive cells were seeded at an initial concentration of 1×10^5 cells/mL and treated with increasing doses of Blasticidin (5, 10 and 25 $\mu\text{g}/\text{mL}$). Cells were then cultured for 12 days in these conditions and proliferation was evaluated with the Z1 Particle Counter (Beckman Coulter). Alternatively, both the original Reh EBS₃tk BlastR 2-1 clone and the *ETV6* expressing 2-1 ETV6-5 clone were treated with Nicotinamide (NAM; Sigma) at a final concentration of 20 mM for 72h. BlastR expression was then evaluated by qRT-PCR (see related section).

4.8.4. shRNAs preparation

Lentiviral particles were prepared from the MISSION TRC shRNA whole library (Sigma) according to the manufacturer's protocol. Briefly, the library was divided in pools each containing approximately 9600 different shRNAs (10 96-well plates/pool). For lentiviral production, HEK293T cells were transfected (FuGENE 6) with the shRNA plasmid pools, psPAX2 (Addgene) and pMD2.G (Addgene) at

a ratio of 1/0.75/0.25. Cell culture media were harvested and filtered (0.45 µm filter) 48 hours post-transfection. The multiplicity of infection (MOI) of each shRNAs pool was evaluated by transducing Reh cells with various dilutions of viral particles using 5 µg/mL polybrene. The next day, media was replaced and cells were selected with 1 µg/mL puromycin for 48h. Cell viability was assessed by propidium iodide (PI) staining (final concentration of 0.2 µg/mL in PBS). 1×10^4 stained cells were analyzed by flow cytometry on a LSRFortessa FACS (Becton Dickinson). Data was acquired with the FACSDiva software (Becton Dickinson) ([Additional file 17](#)).

For individual shRNAs, HEK293T cells were transfected with the pLKO shRNA construct together with the third generation encapsidation plasmids, pRSV-Rev, pMD2.VSVG and pMDL using Lipofectamine 2000 (Thermo Fisher Scientific). Media was renewed 16h post-transfection and harvested after an additional 30h of culture. Viral particles were used directly for transduction.

4.8.5. Genome wide shRNA screening

1×10^7 *ETV6*-dependent Blasticidin sensitive 2-1 *ETV6*-5 cells were transduced with each of the 15 shRNAs pool at a MOI of 0.3-0.4 (determined for each pool; see "shRNAs preparation" section) using 5 µg/mL polybrene. In these conditions, each unique shRNA is expected to be integrated by approximately 200 cells. Media was changed the next day and replaced with fresh media containing 1 µg/mL puromycin. Cells were maintained in culture for 14 days to allow knockdowns to modulate *ETV6* function and BlastR expression. Each pool of cells were then split into 3 parts: 1×10^7 cells were pelleted and used as the initial

time point of the experiment (T_{14}), 2×10^7 cells were maintained as is (no Blasticidin), and 5×10^7 cells were treated with 50 $\mu\text{g}/\text{mL}$ Blasticidin. After 14 days, 1×10^7 cells of each shRNAs pool (with and without Blasticidin selection) were pelleted and used as the first time point of selection (T_{28}). Cells were maintained for an additional 14 days in culture in these conditions and pellets were taken for each shRNA pool and used as the final time point of selection (T_{42}).

Genomic DNA (gDNA) was extracted from each sample using the QIAamp DNA Mini Kit according to the manufacturer's protocol (QIAGEN). Quantification was performed on Qubit using the Quant-it PicoGreen dsDNA Kit (Broad range; Thermo Fisher Scientific). 6 μg of gDNA was used as template for the first PCR with primers flanking the integrated shRNA sequences and containing Illumina-compatible overhang extremities: F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCGTAACCTTGAAAGTATTT
CGATTTCTTGGCTTTATATATCTTGTGG; R-
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGTGGATGAATACTGC
CATTTGTCTC. To evaluate the sensitivity of the methodology, duplicates of T_{14} samples were supplemented with respectively 20 and 200 copies of SHC002 and SHC202 lentiviral vectors prior to the first PCR ([Additional file 18](#)). Technical duplicates for each T_{14} samples also allowed the assessment of the reproducibility of the PCR and sequencing methods ([Additional file 19](#)). For each sample, 12 amplification reactions of 50 μL were performed as follow: gDNA: 10 ng/ μL , primers: 0.25 μM each, dNTPs: 0.3 mM and 1U KAPA HiFi DNA

polymerase (KAPA Biosystems) in 1X KAPA reaction buffer. Thermocycling method: 95°C, 4 min; [98°C, 20 sec; 70°C, 15 sec; 72°C, 1 min] X 15 cycles; 72°C, 5 min. PCR reactions from each gDNA sample were pooled and 90 µL were purified with 1.8X volume AMPure XP magnetic beads (Beckman Coulter) and eluted in 30 µL water. The second amplification reaction used Nextera XT index Kit v2 primers (Illumina) and was performed in 50 µL as follow: purified PCR#1: 30 µL, primers: 5 µL each, dNTPs: 0.3 mM and 1U KAPA HiFi DNA polymerase in 1X KAPA reaction buffer. Thermocycling method: 95°C, 4 min; [98°C, 20 sec; 68°C, 30 sec; 72°C, 30 sec] X 12 cycles; 72°C, 5 min. Reactions were once again purified with AMPure XP magnetic beads (as above). Amplicons size was confirmed by Bioanalyser on a DNA 1000 chip (Agilent) and libraries were sequenced using the HiSeq 2500 platform (paired-end 100 b.p.; Illumina) with 9×10^6 reads per library.

4.8.6. Data analysis

Reads from each sample were aligned on a custom reference containing all the target sequences of the MISSION Human shRNA library using the Bowtie2 alignment tool version 2.2.3. For a given shRNAs pool and for each condition (T₁₄, T₂₈ with and without Blasticidin, T₄₂ with and without Blasticidin), read counts per shRNA were obtained using BEDTools version 2.22.1. Note that for T₁₄ samples, the technical duplicates were merged at this point. Every shRNA counts in a sample were normalized per the total number of reads in that sample to reflect their relative abundance. shRNAs with no coverage (0) in at least one of the time points were removed from further analyses. To be considered a positive

hit, a shRNA had to have a greater normalized count in the Blastocidin condition compared to the untreated control condition at both time points (T_{28} and T_{42}). The normalized count also had to be greater at the end of the experiment (T_{42} with Blastocidin) compared to the initial count at T_{14} . These positive shRNA hits were further refined using additional thresholds. We filtered out weakly over-represented shRNAs ($\sum \text{Counts}(\text{BlastS} - \emptyset \text{BlastS})$ is less than 2 standard deviation of the added SHC202 control vector; $\leq 9,770$; [Additional file 18](#)) as well as shRNAs which target non-expressed genes [34] in Reh cells [29].

4.8.7. Validation of candidate shRNAs by targeted RNA sequencing

The shRNAs selected for validation ([Additional file 5](#)) and two negative controls from the MISSION library (SHC002 and SHC202) were used individually to produce lentiviral particles and transduce Reh *ETV6* cells (as described in the "shRNAs preparation" section). All shRNA particles were used at a 1/10 to 1/20 dilution to reach a MOI >1. All shRNA cell lines were maintained with 1 $\mu\text{g}/\text{mL}$ puromycin for 2 weeks.

Total RNA was extracted from these shRNA populations using the RNeasy Mini kit (QIAGEN). Two biological replicates of Reh *ETV6* cells and one of Reh pLENTI were also included as additional controls. 50 ng of RNA was then processed through the TruSeq Targeted RNA expression kit according to the manufacturer's protocol (Illumina). Amplicons included *ETV6* target genes [29], non-*ETV6* target genes (from the apoptosis panel; Illumina), genes targeted by shRNAs ([Additional file 5](#)) and housekeeping genes [59]. All probes (n=304) and

their corresponding genes are listed in [Additional file 20](#). Amplicons size was confirmed by Bioanalyser on a DNA 1000 chip (Agilent). Amplicons were sequenced on a HiSeq 2500 platform (paired-end 100 b.p.; Illumina) with 1.8×10^6 reads per sample. Mapping was performed using the Illumina BaseSpace SequenceHub Informatics suite and raw counts were obtained and further processed.

4.8.8. Targeted RNA sequencing analysis

Raw counts for each gene in a sample were first normalized per the number of total reads in that sample. Genes considered very weakly expressed (mean normalised counts < 10 and/or normalised pLENTI counts < 10) were discarded from the analyses. Expression in each sample was further normalised using 12 housekeeping genes [59]. For each gene, the relative expression in a given shRNA sample was calculated compared to the SHC002 and SHC202 controls by including the variability of the 4 *ETV6* replicates: $((\text{shRNA sample count} - \text{SHC002 and SHC202 mean count}) / \text{standard deviation of the 4 } ETV6 \text{ controls})$. The relative expression of a gene was also calculated as a simple ratio against the SHC002 and SHC202 controls: $(\text{shRNA sample count} / \text{SHC002 and SHC202 mean count})$. Using standard deviation-corrected relative expressions, shRNAs inducing a specific re-expression of validated *ETV6* targets, but not non-*ETV6* targets, according to a Welch's corrected Student *t* test were considered positives. Knockdown efficiency for each shRNA sample was calculated similarly. Heatmaps were generated with the pheatmap R package.

4.8.9. Protein extraction

1×10^7 cells were washed in PBS and disrupted for 15 min on ice in a hypotonic cytoplasm extraction buffer (CEB; 10 mM HEPES pH 7.5, 10mM KCl, 1X Protease Inhibitors Set III Animal-free (Calbiochem); 100 μ L/ 1×10^7 cells). NP-40 was added to a final concentration of 0.05% prior to centrifugation. Supernatant containing the cytoplasmic proteins was collected and the nuclei pellet was washed in CEB buffer. Nuclei were then disrupted in a hypertonic nuclear extraction buffer (NEB; 20 mM HEPES pH 7.5, 50 mM KCl, 300 mM NaCl, 5% Glycerol, 0.05% NP-40, 1X Protease Inhibitors Set III (with EDTA); 100 μ L/ 1×10^7 cells) for 1 hour with agitation at 4°C. The supernatant obtained after centrifugation contained soluble nuclear proteins.

4.8.10. Western blotting

Protein immunodetection was performed as previously described [29]. Briefly, 20 μ g of protein samples were diluted in Laemmli buffer and migrated on SDS-denaturing 10% polyacrylamide gels. Transfer was performed overnight at 4°C on polyvinylidene difluoride membranes. Membranes were then blocked in a milk solution prior to immunoblotting using antibodies listed in [Additional file 21](#). Immunodetection was performed by enhanced chemiluminescence with Western Lightning Plus-ECL (PerkinElmer) according to the manufacturer's protocol.

4.8.11. qRT-PCR

Total RNA was extracted from cells using the RNeasy Mini kit (QIAGEN) according to the manufacturer's protocol and quantified by Nanodrop (Thermo

Fisher scientific). 750 ng of RNA were used for reverse transcription using the M-MLV reverse transcriptase (Life Technologies). These cDNAs were then subjected to quantitative real-time PCR. Reactions were performed in triplicates using the SYBR Green PCR Master Mix (Life technologies) and the primers listed in [Additional file 16](#) on the ABI PRISM 7000 sequence detection system (Life Technologies). Relative expression was calculated by the $2^{-(\Delta\Delta Ct)}$ comparative method [60] using *GAPDH* and *$\beta 2M$* as reference genes.

4.8.12. Transcriptome analysis of cancer patients

Transcriptome data from childhood ALL patients and controls were obtained from a previous published study [61]. Additional childhood B-ALL patients (11 t(12;21) positive and 14 t(12;21) negative) and controls (3 CD19⁺CD10⁺ pre-B cells isolated from healthy cord blood samples) were also included. Briefly, white cells from bone marrow samples were enriched with Ficoll-Paque PLUS (GE Healthcare Life Sciences). DNA and RNA from these cells were then extracted using the Allprep DNA/RNA mini kit (Qiagen). 1 μ g of RNA with RIN ranging from 5 to 10 were treated with DNase (Ambion). Libraries were generated using the Illumina TruSeq Stranded Total RNA rRNA removal GOLD kit according to the manufacturer protocol. Sequencing was performed on the HiSeq 2500 platform (paired-end 100 b.p.; Illumina) with 8×10^7 expected reads per sample. The pipeline to produce the RNAseq binary alignment files (BAM) is based on the GATK Best Practice for germline SNPs and Indels in RNAseq. The gene expression values (FPKM) were calculated using the Cufflinks tool version 2.2.1 on hard-clipped BAM files.

TCGA normalized expression values (rundate '20160128') were obtained using the R package RCGAToolbox v.2.0.0 [62]. For each cancer, a log fold change was calculated for each gene using their median expression in the top and bottom 10% of *ETV6* expressing samples. 20 genes with the most negative fold changes were retained for further analysis. Additionally, 20 random genes with log fold changes near 0 (over -0.001 and under 0.001) were used as control genes uncorrelated to *ETV6* expression. Log fold changes for those cancer-specific *ETV6*-correlated and non-correlated genes as well as for known *ETV6* target genes were calculated similarly with the top and bottom 10% of samples according to the expression of each modulator.

4.9 Declarations

4.9.1. Ethics approval and consent to participate

The CHU Sainte-Justine Research Ethics Board approved the protocol. Informed consent was obtained from the parents of the patients to participate in this study.

4.9.2. Availability of data and material

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

4.9.3. Competing interests

The authors declare that they have no competing interests.

4.9.4. Funding

This study was supported by research funds provided by the Terry Fox Research Institute and the Canadian Institutes of Health Research. BN is the recipient of a Cole Foundation scholarship. DS holds the François-Karl Viau Research Chair in Pediatric Oncogenomics. Next generation sequencing was performed at the Integrated Clinical Genomic Center in Pediatrics at the CHU Sainte-Justine Research Center. Computations were performed on the Briarée supercomputer at the Université de Montréal, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the Fonds de recherche du Québec - Nature et technologies(FRQNT).

4.10 Figures

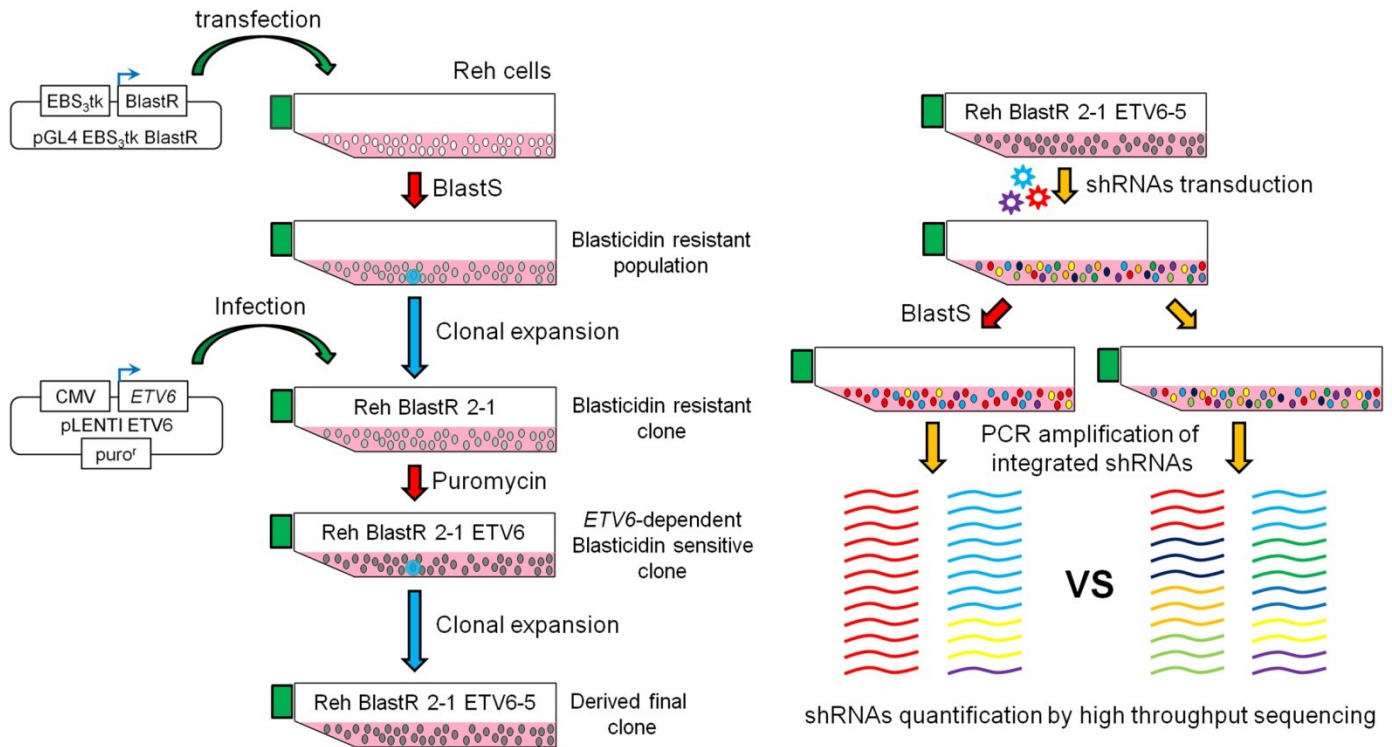


Figure 1: Overview of the genome wide shRNA screening design. To identify putative modulators of *ETV6*-mediated transcriptional repression, a pre-B acute lymphoblastic leukemia cell line (Reh) was modified sequentially to obtain an *ETV6*-dependent BlastS sensitive clone. These cells are expressing the BlastS resistance gene (BlastR) under the control of the *ETV6*-responsive artificial promoter (EBS₃tk). These cells were then transduced with shRNAs ($n \approx 140,000$) and selected with BlastS. Over-represented shRNAs in the BlastS condition compared to the mock condition are expected to target genes required for *ETV6*-mediated repression of BlastR.

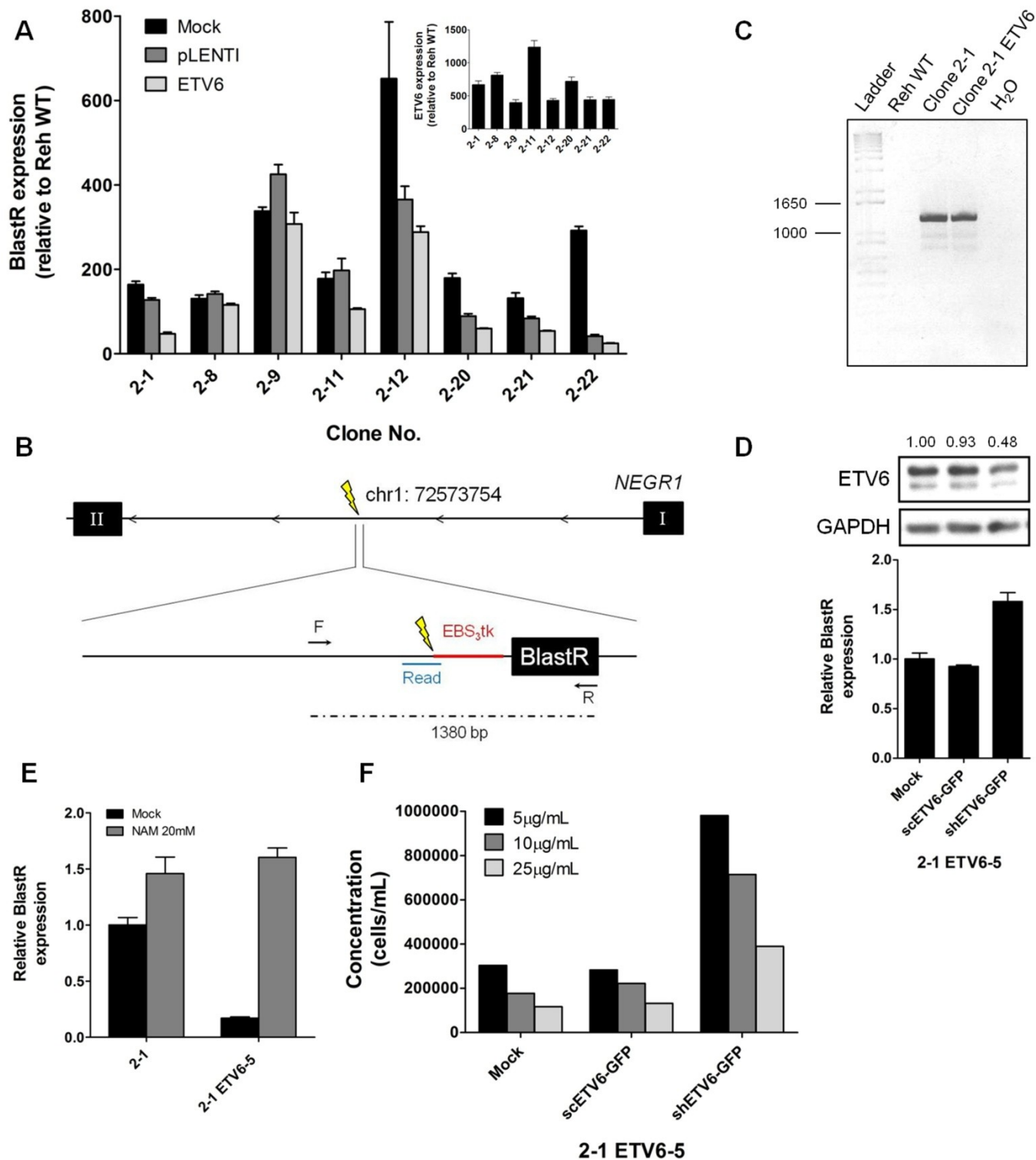


Figure 2: Establishment of an *ETV6*-dependent Blasticidin-sensitive leukemia cell system. A. BlastR expression was evaluated by qRT-PCR in each Reh EBS₃tk BlastR clone (clones 2-1 to 2-22) following *ETV6* expression (inset).

Data are shown as mean \pm SD, n=4 **B.** *ETV6* ChIP-seq data obtained from clone 2-1 located the insertion site of EBS₃tk BlastR (yellow mark). One sequencing read (indicated in blue) contained the partial sequence of EBS₃tk promoter together with 95 nucleotides of the first intron of the *NEGR1* gene. The F/R arrows depict the primers used in the PCR analysis. **C.** PCR analysis of the clone 2-1 confirmed the insertion site of EBS₃tk BlastR in first intron of the *NEGR1* gene. **D.** BlastR expression was evaluated by qRT-PCR in 2-1 ETV6-5 cells after shRNA-mediated silencing of *ETV6* (quantified Western blot) or **E.** after nicotinamide (NAM) treatment. Data are shown as mean \pm SD, n=4. **F.** 2-1 ETV6-5 scramble (negative shRNA control; scETV6-GFP) and shETV6-GFP sorted cells were maintained in culture for 12 days with various Blasticidin concentrations and counted to evaluate proliferation.

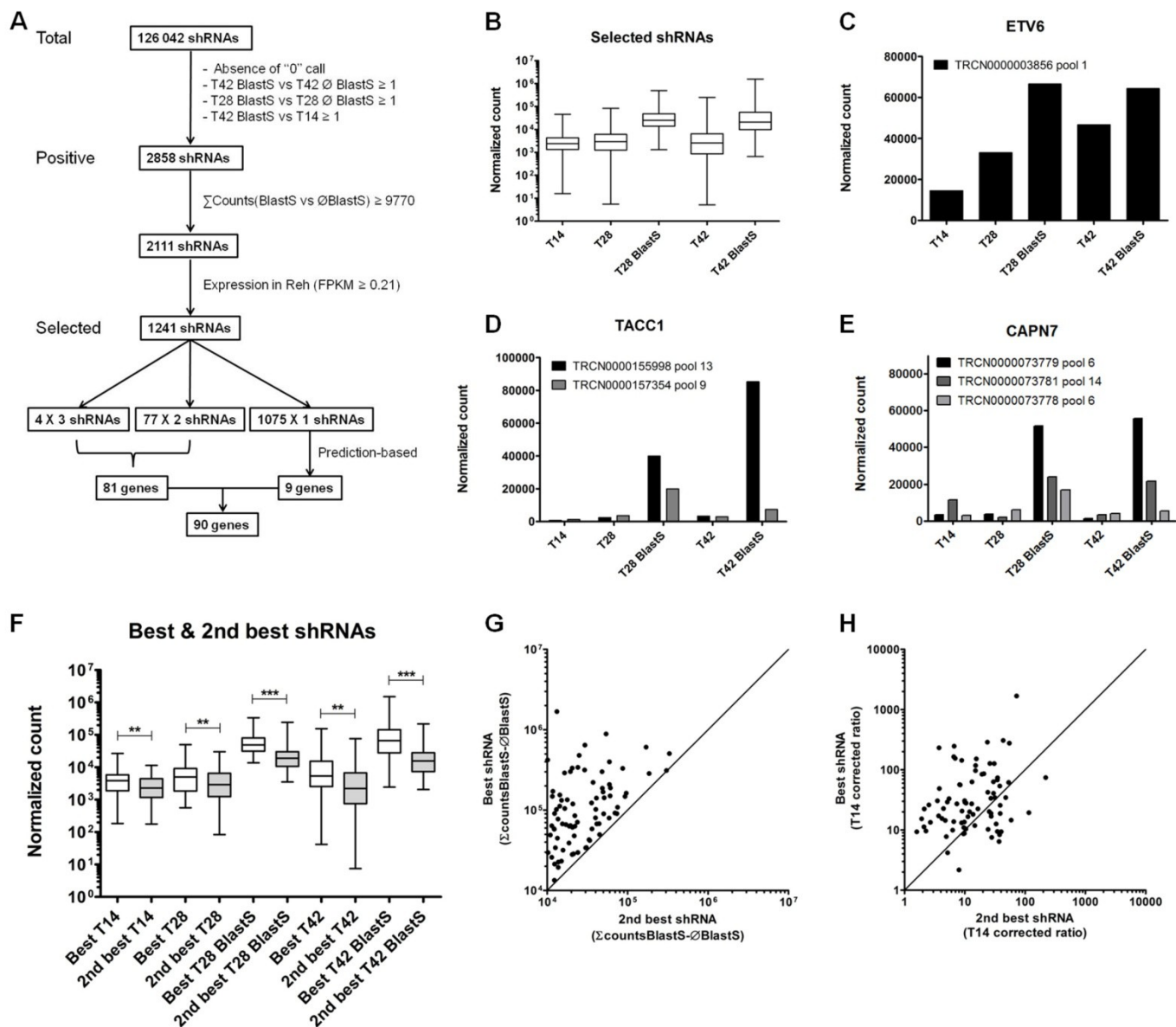


Figure 3: Genome wide shRNA screening to identify *ETV6* modulators. **A.** Flowchart illustrating the step-wise procedure for the analysis of the genome wide shRNA screen sequencing. **B.** Normalized read counts distribution of the 1241 selected shRNAs shown for the 5 time points. Whiskers represent minimum and maximum. **C-E.** Examples of selected shRNAs normalized counts for the 5 time points. **F.** Normalized counts distribution are shown separately for the best and second best shRNAs of a pair or trio. Whiskers represent minimum and maximum; statistical significance is calculated with a paired Student *t* test (n=81). **G.** Relation between the best (y axis) and second best (x axis) shRNAs of a pair or trio according to their respective Blasticidin count difference. **H.** The same relation between the best (y axis) and second best (x axis) shRNAs of a pair or trio when their Blasticidin count difference is corrected for their initial count at T₁₄.

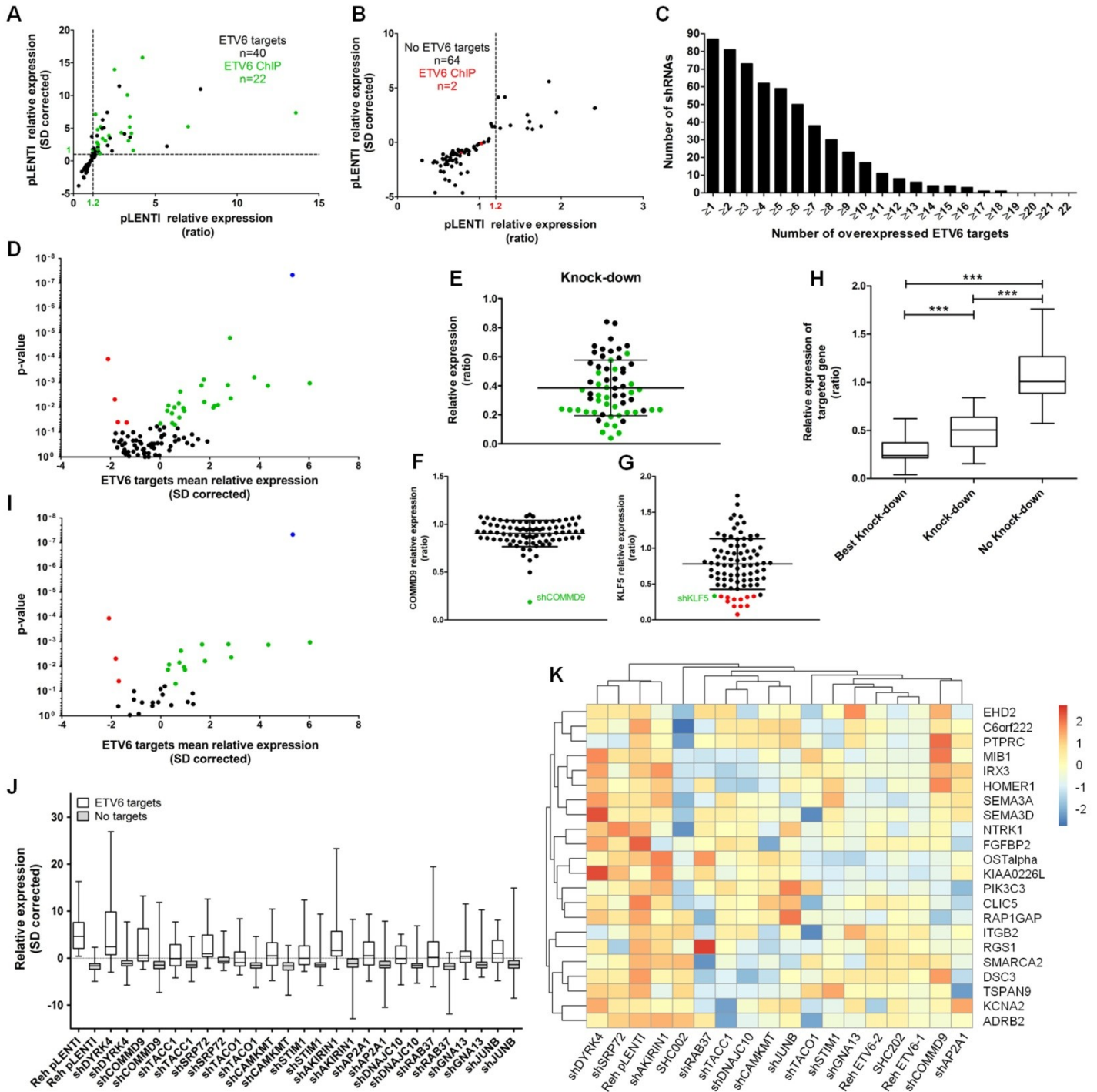


Figure 4: Impact of putative *ETV6* modulators on its transcriptional regulatory network. **A.** Graphical representation of the relative expression of *ETV6* target genes or **B.** *ETV6*-independent control genes in the pLENTI control sample compared to the *ETV6*-expressing controls. **C.** The number of shRNA samples (out of 89) is shown according to their impact on overexpressed *ETV6* target genes. **D.** The mean relative expression of *ETV6* targets (x axis) is

compared to the relative expression of control genes obtained with the same shRNA; the Student's *t* test p-values are shown on the y axis. Colored dots (green: overexpressed; red: underexpressed) are significant shRNAs. The blue dot represents the Reh pLENTI sample which lack *ETV6* expression and serves as a positive control. **E.** Residual expression of the targeted gene in each knocked down samples. The green dots are shRNAs which induce the lowest expression of its targeted gene among all samples. **F.** Relative *COMMD9* expression is shown for all shRNA samples. sh*COMMD9* sample is shown in green and displays the lowest expression. **G.** Relative *KLF5* expression is diminished in the sh*KLF5* sample (green) but several other shRNAs also reduce its expression. **H.** Relative expression of the shRNA-targeted genes according to the category of knockdown (Best knockdown, n=31; Knockdown, n=32; No knockdown, n=22). Whiskers represent minimum and maximum; statistical significance is calculated with an unpaired Student's *t* test. **I.** Similar representation as in panel D; shRNAs shown are restricted to those displaying the best knockdown of their targeted gene (with the inclusion of sh*DYRK4*). **J.** Distribution of the relative expressions of *ETV6* targets (n=22) and control genes (n=62) in the 13 significant shRNA samples. Reh pLENTI is also shown as a control. Whiskers represent minimum and maximum. **K.** Heatmap representation of the normalized read counts of *ETV6* targets (lanes) in the 13 significant shRNA samples (and control samples; columns).

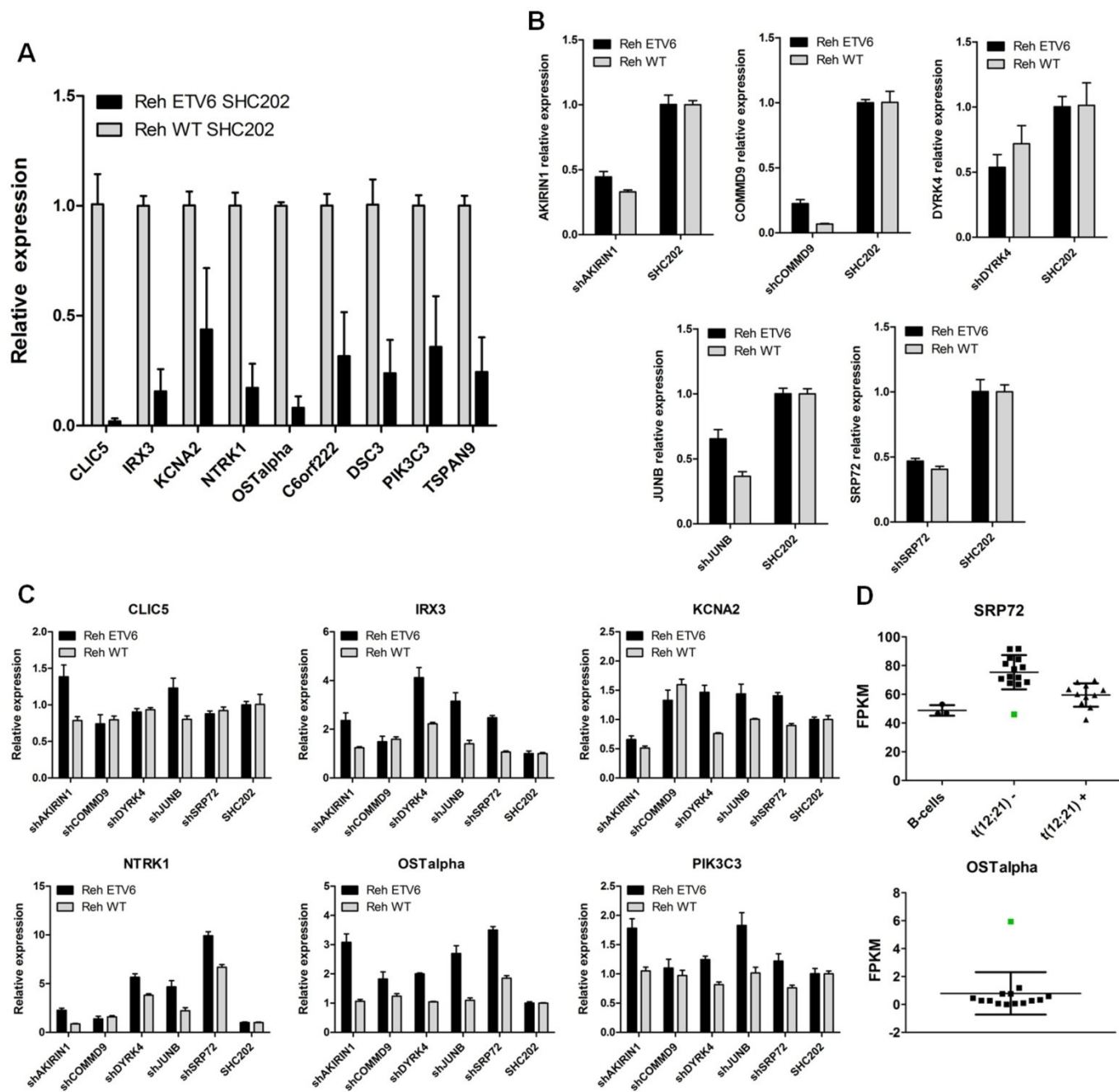


Figure 5: Impact of the top *ETV6* modulators on the transcription of *t(12;21)*-associated overexpressed genes. A. qRT-PCR analysis of 9 *t(12;21)*-associated *ETV6* target genes. Data are shown as mean \pm SD, $n=6$. **B.** Relative expression of the 5 top *ETV6* modulators after shRNAs transduction. Data are shown as mean \pm SD, $n=6$. **C.** Relative expression of the *t(12;21)*-associated *ETV6* target genes in both Reh WT and Reh *ETV6* cells upon silencing of *ETV6* modulators (see also [Additional file 10](#)). Data are shown as mean \pm SD, $n=6$. **D.** *SRP72* (upper panel) and *OSTalpha* (lower panel) expression in leukemic patient samples. The green dot represents the same patient.

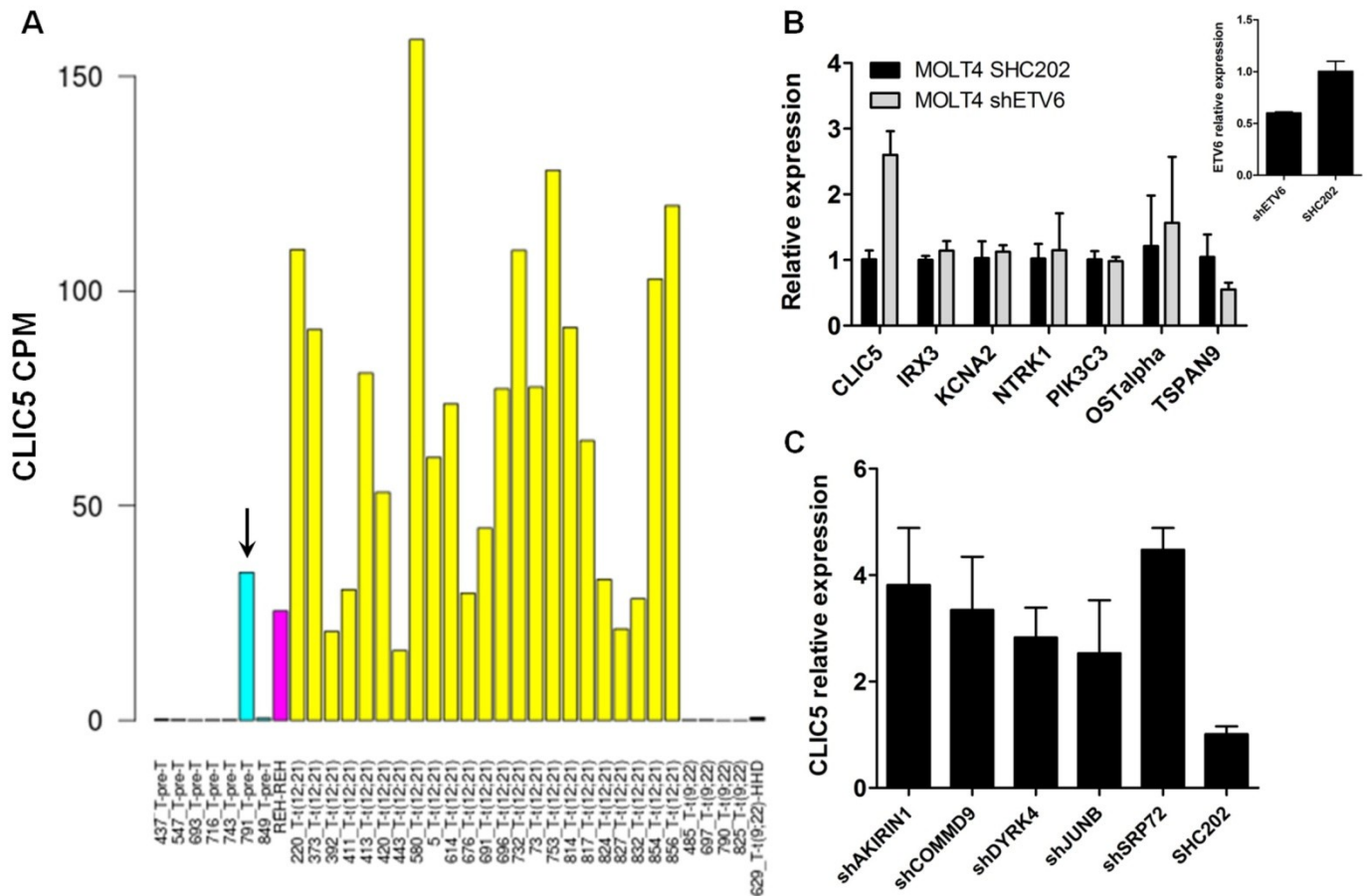


Figure 6: Top *ETV6* modulators in childhood pre-T ALL. **A.** Expression of *CLIC5* in leukemic patient samples according to their molecular subtypes. The arrow indicates a pre-T ALL patient with upregulated *CLIC5* expression. Expression in Reh cells is shown for reference (purple). **B.** qRT-PCR analysis of 9 t(12;21)-associated *ETV6* target genes after shRNA-mediated silencing of *ETV6* in MOLT4 cells (inset). Two of these genes were not expressed (*C6orf222* and *DSC3*; not shown). Data are shown as mean \pm SD, n=6. **C.** Relative expression of *CLIC5* after the knockdown of *ETV6* modulators in MOLT4 cells (shown in [Additional file 11](#)). Data are shown as mean \pm SD, n=6.

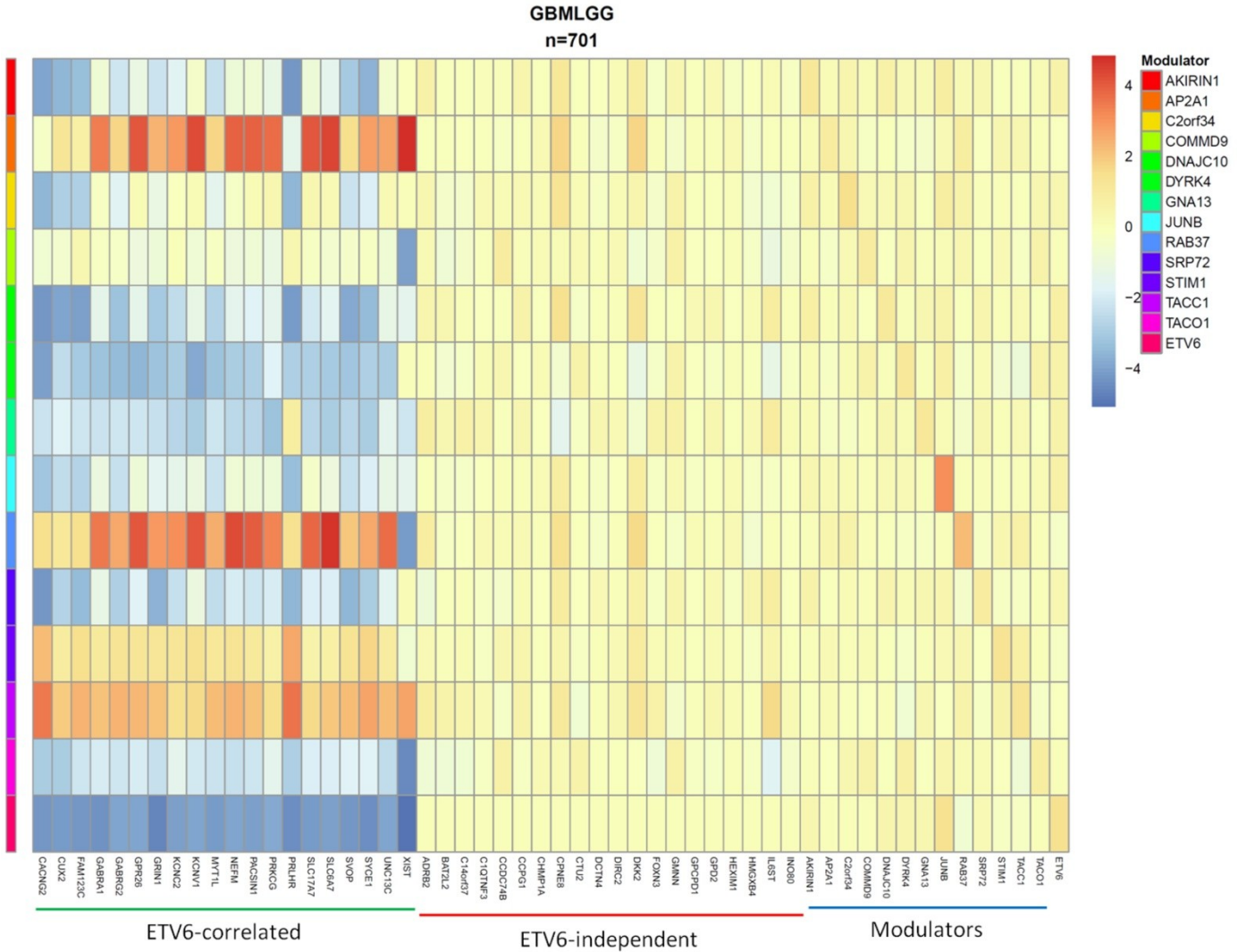


Figure 7: *ETV6* modulators in glioblastomas and gliomas. Heatmap depicting the log fold changes in gene expression (columns) between TCGA patient samples with the highest and lowest expression of a modulator (lanes). Genes are grouped as *ETV6*-correlated, *ETV6*-independent and *ETV6* modulators.

4.11 Tables

Table 1: Data summary of the 5 top *ETV6* modulators

Gene symbol		<i>AKIRIN1</i>	<i>COMMD9</i>	<i>DYRK4</i>	<i>JUNB</i>	<i>SRP72</i>
shRNA		TRCN0000134954	TRCN0000167758	TRCN0000197116	TRCN0000014943	TRCN0000151445
Genome wide shRNA screen	Pool	12	11	13	2	15
	T14	2822	1196	5864	178	5250
	T28	1563	1538	15253	1131	14583
	Normalized counts					
	T28 BlastS	41303	16880	56508	9936	163516
	T42	1306	1835	14864	656	17490
	T42 BlastS	24629	27894	37829	12352	199618
	∑BlastS-ØBlastS counts	63063	41401	64221	20500	331062
T14 corrected ratio	22.35	34.61	10.95	115.39	63.06	
KD efficiency (remaining expression)	38.73%	18.78%	26.84%	19.44%	48.82%	
Number of overexpressed targets (≥ 1 SD)	15	10	16	11	10	
Targeted sequencing validation	Mean targets relative expression					
	SD corrected	4.35	2.73	6.03	1.67	2.85
	Ratio	2.10	1.93	3.21	1.99	2.55
	Mean controls relative expression					
	SD corrected	-1.07	-1.13	-0.76	-1.04	-0.15
	Ratio	0.78	0.71	0.77	0.76	0.93
Student's <i>t</i> test p-value						
Targets vs no targets	0.00133	0.00127	0.00106	0.00129	0.00434	
Targets SHC vs shRNA (paired)	0.00669	0.01284	0.00281	0.02227	0.00564	

T14, T28 and T42 represent the timepoints (days after transduction), KD, knockdown; SD, standard deviation

4.12 References

1. Mavrothalassitis G, Ghysdael J. Proteins of the ETS family with transcriptional repressor activity. *Oncogene*. 2000; 19(55):6524-6532.
2. Ciau-Uitz A, Wang L, Patient R, Liu F. ETS transcription factors in hematopoietic stem cell development. *Blood Cells Mol Dis*. 2013; 51(4):248-255.
3. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: a review. *Leuk Res*. 2012; 36(8):945-961.
4. Otsubo K, Kanegane H, Eguchi M, Eguchi-Ishimae M, Tamura K, Nomura K et al. ETV6-ARNT fusion in a patient with childhood T lymphoblastic leukemia. *Cancer Genet Cytogenet*. 2010; 202(1):22-26.
5. Romana SP, Mauchauffe M, Le Coniat M, Chumakov I, Le Paslier D, Berger R et al. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood*. 1995; 85(12):3662-3670.
6. Golub TR, Barker GF, Bohlander SK, Hiebert SW, Ward DC, Bray-Ward P et al. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 1995; 92(11):4917-4921.
7. Gunji H, Waga K, Nakamura F, Maki K, Sasaki K, Nakamura Y et al. TEL/AML1 shows dominant-negative effects over TEL as well as AML1. *Biochem Biophys Res Commun*. 2004; 322(2):623-630.
8. Romana SP, Poirel H, Leconiat M, Flexor MA, Mauchauffe M, Jonveaux P et al. High frequency of t(12;21) in childhood B-lineage acute lymphoblastic leukemia. *Blood*. 1995; 86(11):4263-4269.
9. Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML et al. TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia*. 1995; 9(12):1985-1989.
10. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatr Clin North Am*. 2015; 62(1):47-60.
11. Agape P, Gerard B, Cave H, Devaux I, Vilmer E, Lecomte MC et al. Analysis of ETV6 and ETV6-AML1 proteins in acute lymphoblastic leukaemia. *Br J Haematol*. 1997; 98(1):234-239.

12. Poirel H, Lacronique V, Mauchauffe M, Le Coniat M, Raffoux E, Daniel MT et al. Analysis of TEL proteins in human leukemias. *Oncogene*. 1998; 16(22):2895-2903.
13. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet*. 2001; 130(2):93-104.
14. van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J et al. Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood*. 2011; 118(4):1041-1051.
15. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A*. 2002; 99(12):8242-8247.
16. Patel N, Goff LK, Clark T, Ford AM, Foot N, Lillington D et al. Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br J Haematol*. 2003; 122(1):94-98.
17. Lilljebjorn H, Sonesson C, Andersson A, Heldrup J, Behrendtz M, Kawamata N et al. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet*. 2010; 19(16):3150-3158.
18. Montpetit A, Larose J, Boily G, Langlois S, Trudel N, Sinnett D. Mutational and expression analysis of the chromosome 12p candidate tumor suppressor genes in pre-B acute lymphoblastic leukemia. *Leukemia*. 2004; 18(9):1499-1504.
19. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol*. 2015; 16(16):1659-1666.
20. Topka S, Vijai J, Walsh MF, Jacobs L, Maria A, Villano D et al. Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet*. 2015; 11(6):e1005262.
21. Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet*. 2015; 47(5):535-538.

22. Zhang MY, Churpek JE, Keel SB, Walsh T, Lee MK, Loeb KR et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet.* 2015; 47(2):180-185.
23. Lee YJ, Kim JH, Bae S, Rho SK, Choe SY. Mechanism of transcriptional repression by TEL/RUNX1 fusion protein. *Mol Cells.* 2004; 17(2):217-222.
24. Wang L, Hiebert SW. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene.* 2001; 20(28):3716-3725.
25. Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A et al. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood.* 2000; 96(7):2557-2561.
26. Kuwata T, Gongora C, Kanno Y, Sakaguchi K, Tamura T, Kanno T et al. Gamma interferon triggers interaction between ICSPB (IRF-8) and TEL, recruiting the histone deacetylase HDAC3 to the interferon-responsive element. *Mol Cell Biol.* 2002; 22(21):7439-7448.
27. Azlan A, Dzaki N, Azzam G. Argonaute: The executor of small RNA function. *J Genet Genomics.* 2016; 43(8):481-494.
28. Mohr SE, Smith JA, Shamu CE, Neumuller RA, Perrimon N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol.* 2014; 15(9):591-600.
29. Neveu B, Spinella JF, Richer C, Lagace K, Cassart P, Lajoie M et al. CLIC5: a novel ETV6 target gene in childhood acute lymphoblastic leukemia. *Haematologica.* 2016; 101(12):1534-1543.
30. Szymczyzna BR, Arrowsmith CH. DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. *J Biol Chem.* 2000; 275(37):28363-28370.
31. Schmidt MT, Smith BC, Jackson MD, Denu JM. Coenzyme specificity of Sir2 protein deacetylases: implications for physiological regulation. *J Biol Chem.* 2004; 279(38):40122-40129.
32. Bitterman KJ, Anderson RM, Cohen HY, Latorre-Esteves M, Sinclair DA. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human SIRT1. *J Biol Chem.* 2002; 277(47):45099-45107.

33. Sanders BD, Jackson B, Marmorstein R. Structural basis for sirtuin function: what we know and what we don't. *Biochim Biophys Acta*. 2010; 1804(8):1604-1616.
34. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics*. 2013; 14:778.
35. Singh R, Park D, Xu J, Hosur R, Berger B. Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. *Nucleic Acids Res*. 2010; 38(Web Server issue):W508-515.
36. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002; 1(2):133-143.
37. Spinella JF, Cassart P, Richer C, Saillour V, Ouimet M, Langlois S et al. Genomic characterization of pediatric T-cell acute lymphoblastic leukemia reveals novel recurrent driver mutations. *Oncotarget*. 2016; 7(40):65485-65503.
38. Bohlander SK. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol*. 2005; 15(3):162-174.
39. Mishima Y, Jayasinghe CD, Lu K, Otani J, Shirakawa M, Kawakami T et al. Nucleosome compaction facilitates HP1gamma binding to methylated H3K9. *Nucleic Acids Res*. 2015; 43(21):10200-10212.
40. Kar A, Gutierrez-Hartmann A. Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. *Crit Rev Biochem Mol Biol*. 2013; 48(6):522-543.
41. Feng FY, Brenner JC, Hussain M, Chinnaiyan AM. Molecular pathways: targeting ETS gene fusions in cancer. *Clin Cancer Res*. 2014; 20(17):4442-4448.
42. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer*. 2017; 17(6):337-351.
43. Nowak SJ, Aihara H, Gonzalez K, Nibu Y, Baylies MK. Akirin links twist-regulated transcription with the Brahma chromatin remodeling complex during embryogenesis. *PLoS Genet*. 2012; 8(3):e1002547.

44. Bonnay F, Nguyen XH, Cohen-Berros E, Troxler L, Batsche E, Camonis J et al. Akirin specifies NF-kappaB selectivity of Drosophila innate immune response via chromatin remodeling. *EMBO J.* 2014; 33(20):2349-2362.
45. Zhan W, Wang W, Han T, Xie C, Zhang T, Gan M et al. COMMD9 promotes TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer. *Cell Signal.* 2017; 30:59-66.
46. Bartuzi P, Hofker MH, van de Sluis B. Tuning NF-kappaB activity: a touch of COMMD proteins. *Biochim Biophys Acta.* 2013; 1832(12):2315-2321.
47. Papadopoulos C, Arato K, Lilienthal E, Zerweck J, Schutkowski M, Chatain N et al. Splice variants of the dual specificity tyrosine phosphorylation-regulated kinase 4 (DYRK4) differ in their subcellular localization and catalytic activity. *J Biol Chem.* 2011; 286(7):5494-5505.
48. Becker W, Weber Y, Wetzel K, Eirmbter K, Tejedor FJ, Joost HG. Sequence characteristics, subcellular localization, and substrate specificity of DYRK-related kinases, a novel family of dual specificity protein kinases. *J Biol Chem.* 1998; 273(40):25893-25902.
49. Guo X, Williams JG, Schug TT, Li X. DYRK1A and DYRK3 promote cell survival through phosphorylation and activation of SIRT1. *J Biol Chem.* 2010; 285(17):13223-13232.
50. Lopez RG, Carron C, Ghysdael J. v-SRC specifically regulates the nucleocytoplasmic delocalization of the major isoform of TEL (ETV6). *J Biol Chem.* 2003; 278(42):41316-41325.
51. Poirel H, Oury C, Carron C, Duprez E, Laabi Y, Tsapis A et al. The TEL gene products: nuclear phosphoproteins with DNA binding properties. *Oncogene.* 1997; 14(3):349-357.
52. Arai H, Maki K, Waga K, Sasaki K, Nakamura Y, Imai Y et al. Functional regulation of TEL by p38-induced phosphorylation. *Biochem Biophys Res Commun.* 2002; 299(1):116-125.
53. Papoudou-Bai A, Hatzimichael E, Barbouti A, Kanavaros P. Expression patterns of the activator protein-1 (AP-1) family members in lymphoid neoplasms. *Clin Exp Med.* 2016.

54. Bassuk AG, Leiden JM. A direct physical association between ETS and AP-1 transcription factors in normal human T cells. *Immunity*. 1995; 3(2):223-237.
55. Verger A, Duterque-Coquillaud M. When Ets transcription factors meet their partners. *Bioessays*. 2002; 24(4):362-370.
56. Kirwan M, Walne AJ, Plagnol V, Velangi M, Ho A, Hossain U et al. Exome sequencing identifies autosomal-dominant SRP72 mutations associated with familial aplasia and myelodysplasia. *Am J Hum Genet*. 2012; 90(5):888-892.
57. Akopian D, Shen K, Zhang X, Shan SO. Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem*. 2013; 82:693-721.
58. Li J, Zhou F, Zhan D, Gao Q, Cui N, Li J et al. A novel histone H4 arginine 3 methylation-sensitive histone H4 binding activity and transcriptional regulatory function for signal recognition particle subunits SRP68 and SRP72. *J Biol Chem*. 2012; 287(48):40641-40651.
59. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013; 29(10):569-574.
60. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001; 25(4):402-408.
61. Lajoie M, Drouin S, Caron M, St-Onge P, Ouimet M, Gioia R et al. Specific expression of novel long non-coding RNAs in high-hyperdiploid childhood acute lymphoblastic leukemia. *PLoS One*. 2017; 12(3):e0174124.
62. Samur MK. RTCGAToolbox: a new tool for exporting TCGA Firehose data. *PLoS One*. 2014; 9(9):e106397.

4.13 Supplementary material

4.13.1. Supplemental figures



TAACTC does not align on chr1
but maps on EBS₃tk sequence

Figure S1 (Additional file 1): Identification of EBS₃tk BlastR insertion site using ChIP-seq data. *ETV6* ChIP-seq performed in 2-1 *ETV6* cells [Article 2] captured 3 reads on the EBS₃tk promoter. One of this read combined a partial EBS₃tk sequence with 95 nucleotides mapping at the genomic coordinates chr1: 72573660-72573754, locating the breakpoint in the first intron of the *NEGR1* gene.

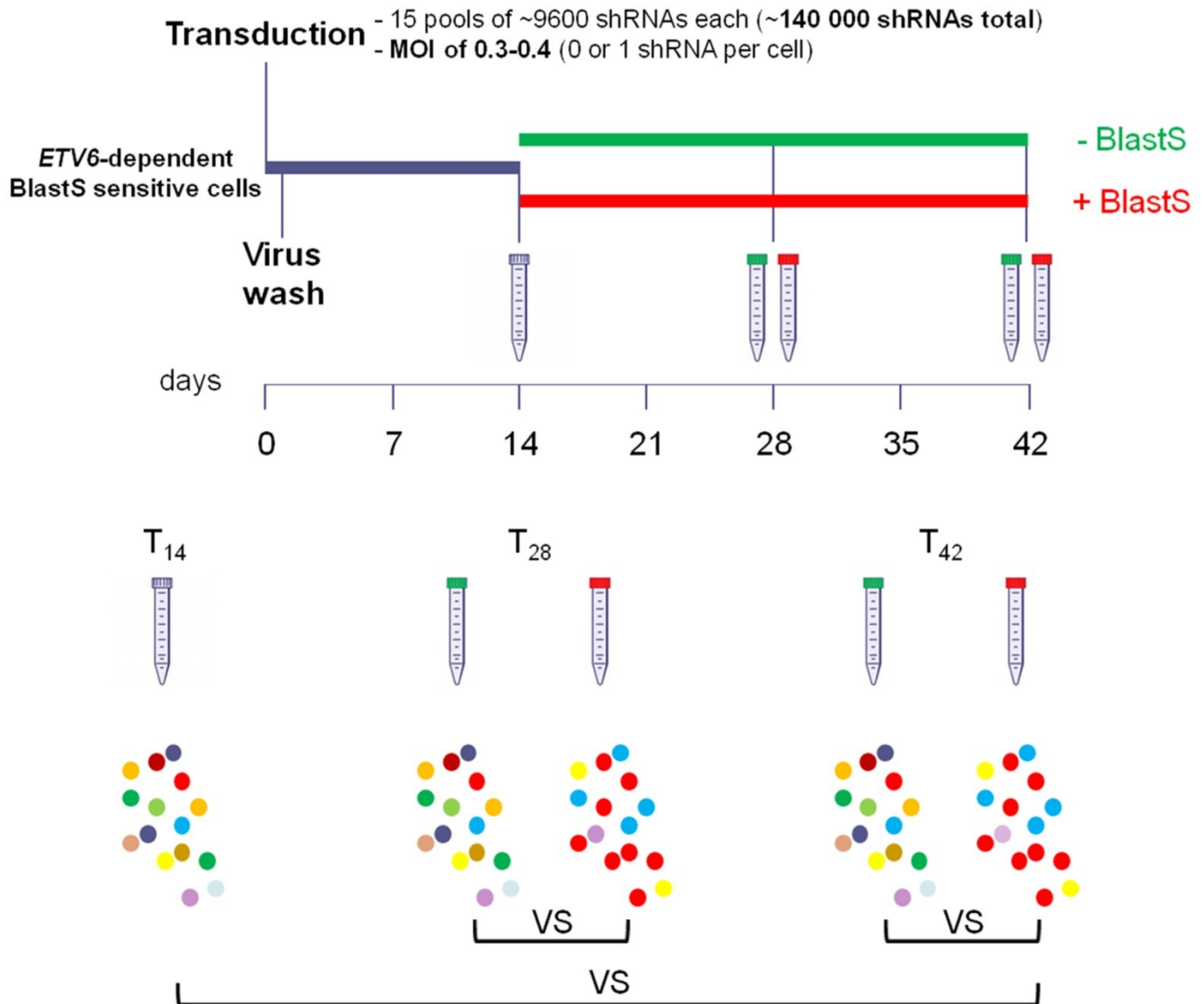


Figure S2 (Additional file 2): Experimental design of the genome wide shRNA screen. *ETV6*-dependent Blastasticin (BlastS) sensitive cells (i.e. Reh EBS₃tk BlastR 2-1 *ETV6*-5 clone) were transduced with 15 pools of shRNAs lentiviral particles. Virus were washed the next day and cells were maintained in culture for 14 days. Each pool of cells were separated into 3 fractions: pelleted and used as the initial time point of the experiment (T₁₄), maintained in culture and treated with Blastasticin. After 14 days, cells of each shRNAs pool (with and without Blastasticin selection) were pelleted and used as the first time point of selection (T₂₈). Cells were cultured for an additional 14 days in these conditions and pelleted (end point; T₄₂). For each pools, shRNAs were quantified by high throughput sequencing and samples were compared as indicated (VS).

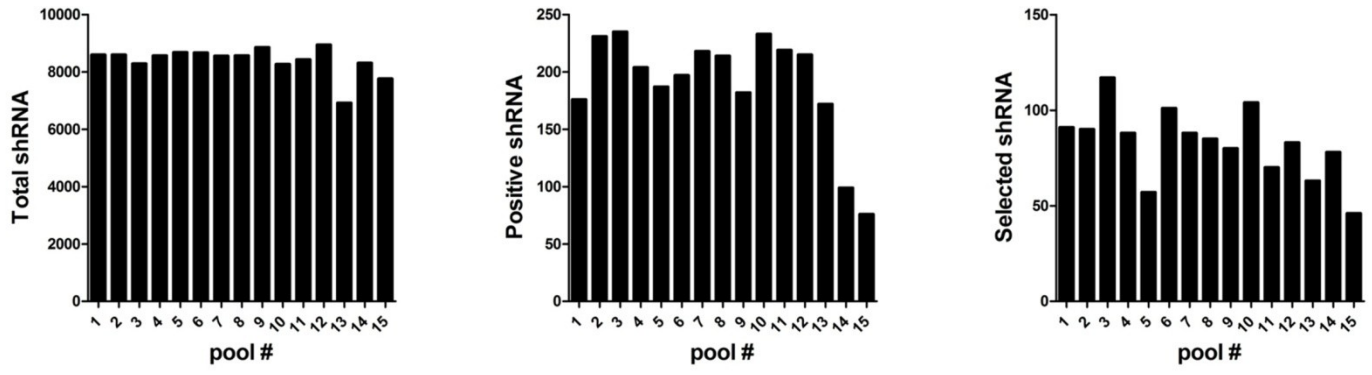


Figure S3 (Additional file 3): shRNAs distribution. The number of shRNAs per pool is given for the total shRNAs, positive shRNAs and selected shRNAs.

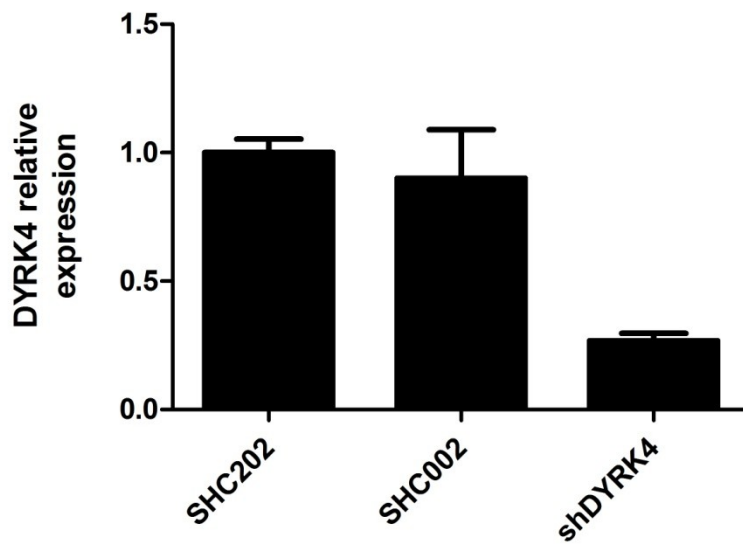


Figure S4 ([Additional file 8](#)): *DYRK4* knockdown. qPCR showing the knockdown efficiency of the shDYRK4 included in the validation experiment. *DYRK4* relative expression is calculated for SHC controls and shDYRK4 Reh ETV6 samples. Data are shown as mean \pm SD, n=3.

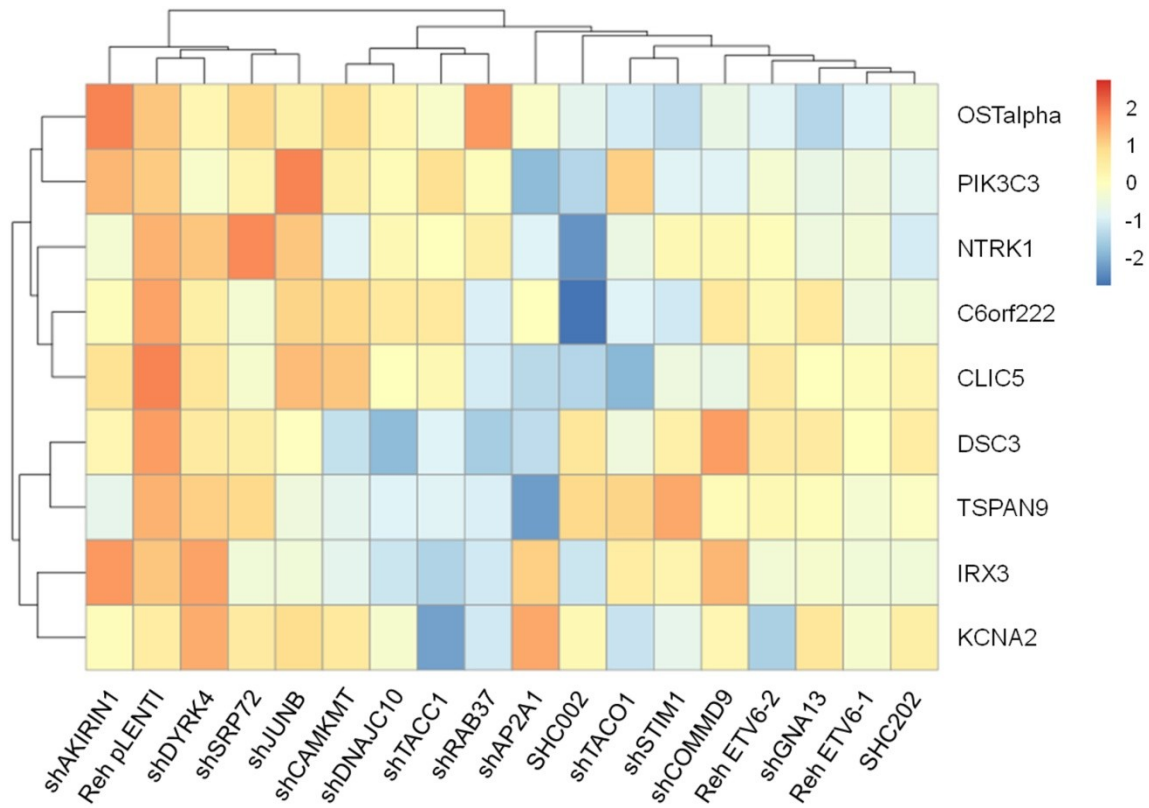


Figure S5 (Additional file 9): *ETV6* modulators on t(12;21)-associated genes. Heatmap representation of the normalized read counts of the 9 t(12;21)-associated *ETV6* targets (lanes) in the 13 significant shRNA samples (and control samples; columns).

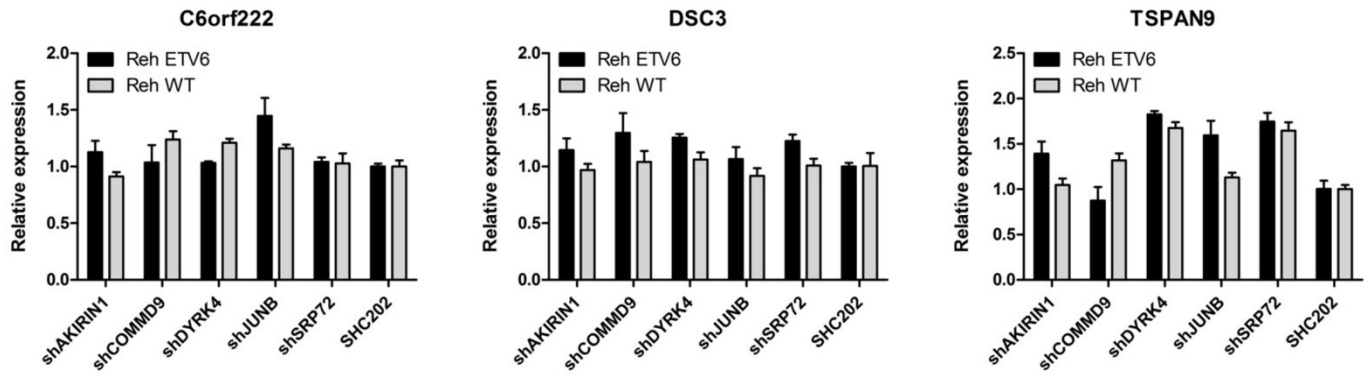


Figure S6 (Additional file 10): Top *ETV6* modulators on t(12;21)-associated genes. Relative expressions of the t(12;21)-associated *ETV6* target genes *C6orf222*, *DSC3* and *TSPAN9* were calculated in both Reh and Reh *ETV6* backgrounds upon silencing of *ETV6* modulators. Data are shown as mean \pm SD, n=6.

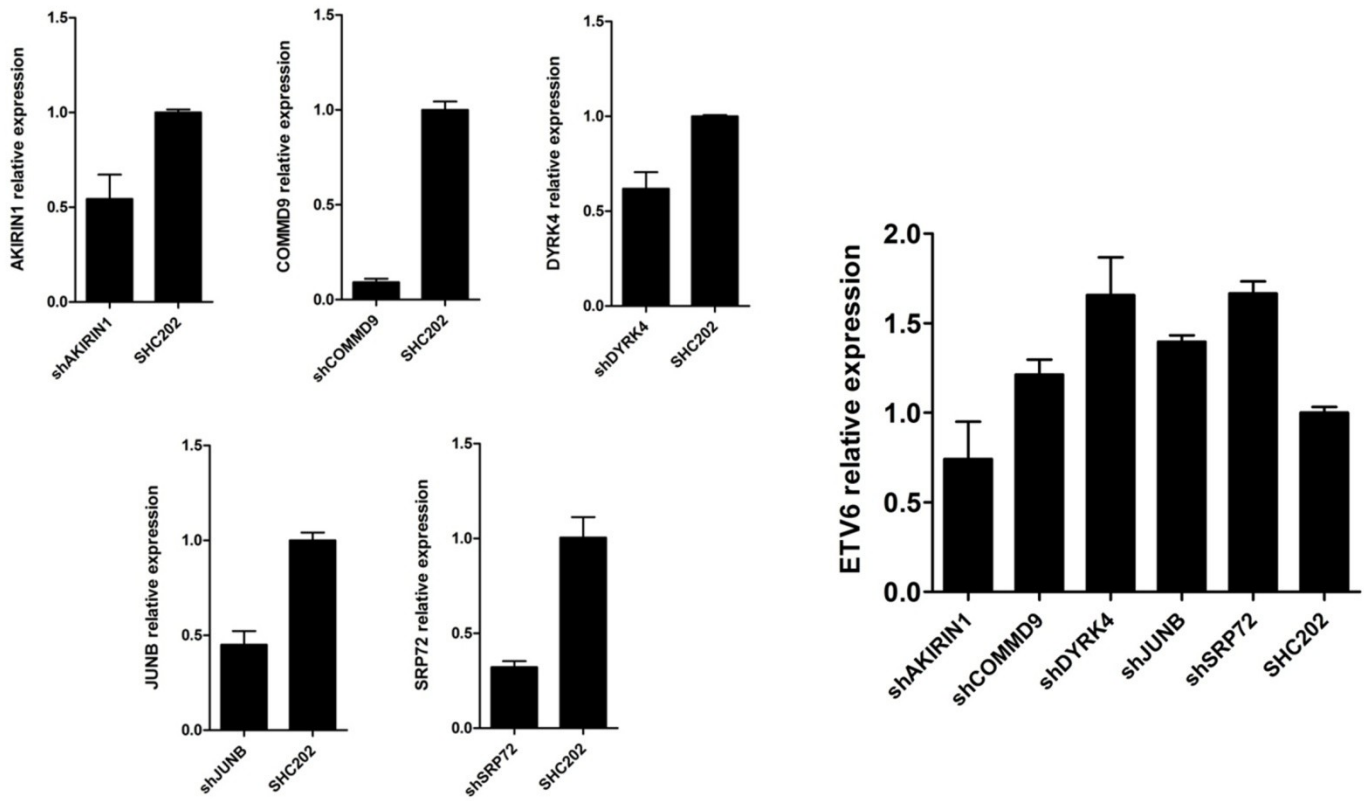


Figure S7 (Additional file 11): Knockdown of the top *ETV6* modulators in a pre-T ALL cell line. Relative expression of the top *ETV6* modulators after shRNAs transduction in MOLT4 cells. Relative expression of *ETV6* is also shown in these samples. Data are shown as mean \pm SD, n=6.

GBMLGG
n=701

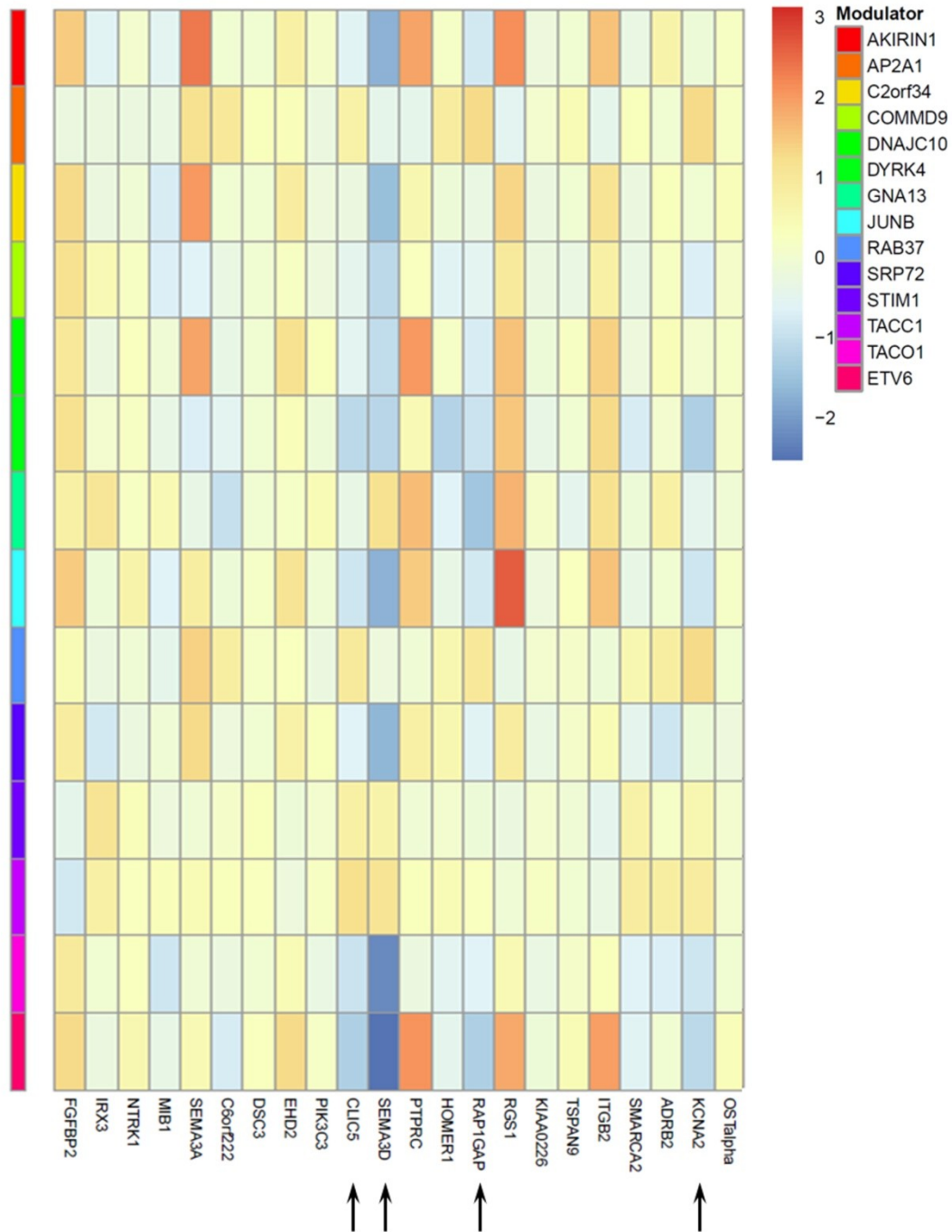


Figure S8 (Additional file 12): **ETV6** modulators in glioblastomas and gliomas. Heatmap depicting the log fold changes in the expression of known *ETV6* target genes (columns) between TCGA patient samples with the highest and lowest expression of a modulator (lanes). Arrows show genes that correlate with *ETV6* expression.

BRCA
n=1212

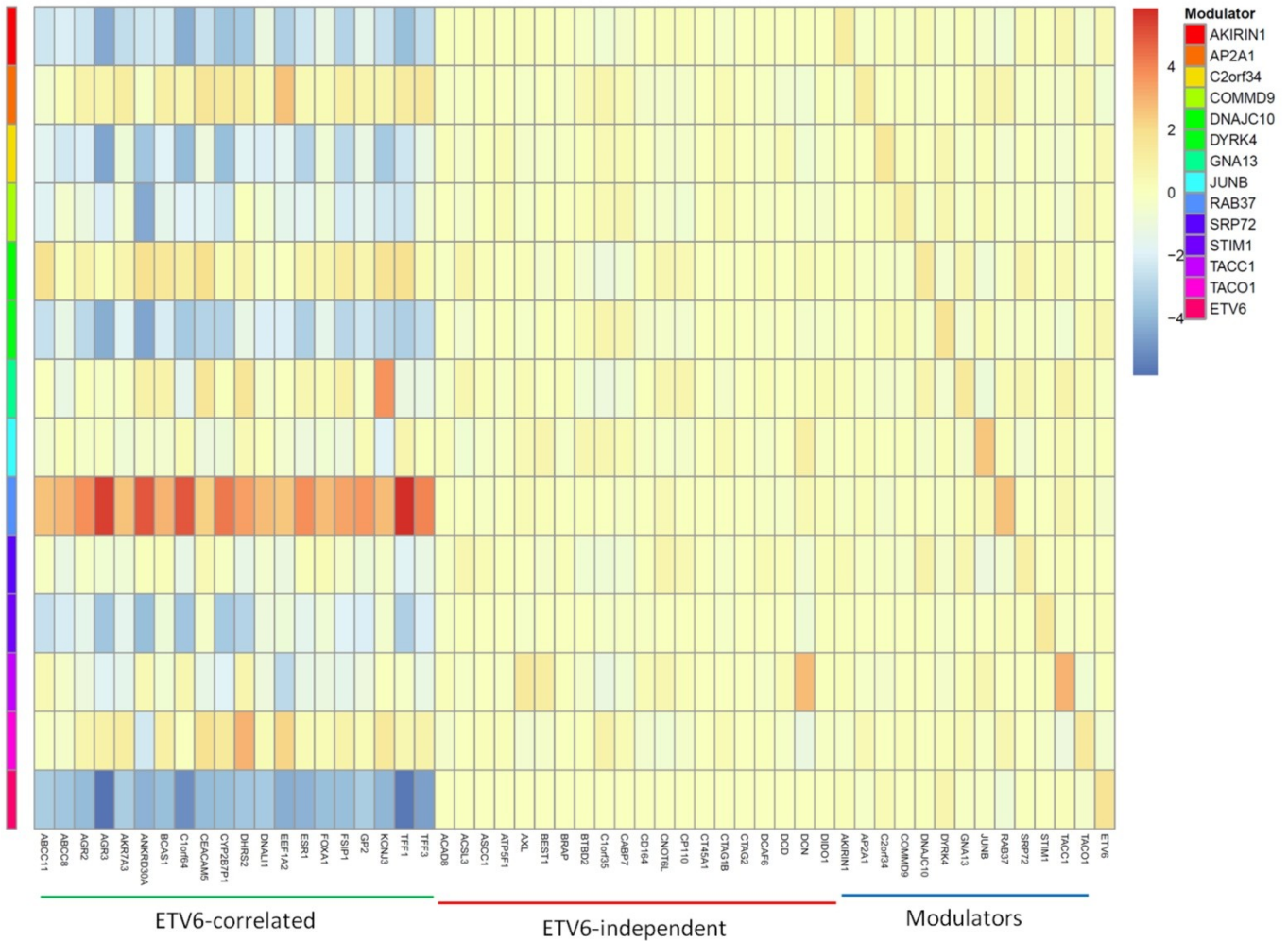


Figure S9 (Additional file 13): *ETV6* modulators in breast cancer. Heatmap depicting the log fold changes in gene expression (columns) between TCGA breast cancer patient samples with the highest and lowest expression of a modulator (lanes). Genes are grouped as *ETV6*-correlated, *ETV6*-independent and *ETV6* modulators.

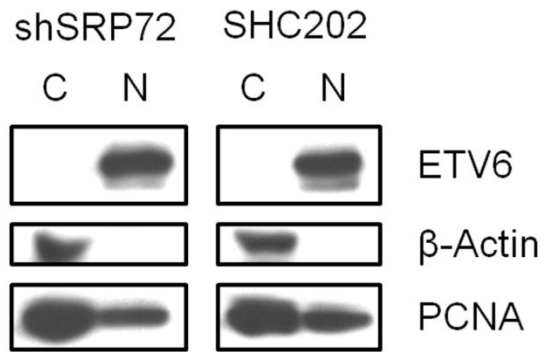


Figure S10 ([Additional file 15](#)): **Impact of *SRP72* on the *ETV6* protein.** *ETV6* expression and localization (C: cytoplasm; N: nucleus) was evaluated by Western blotting in Reh *ETV6* shSRP72 and SHC202 control cells.

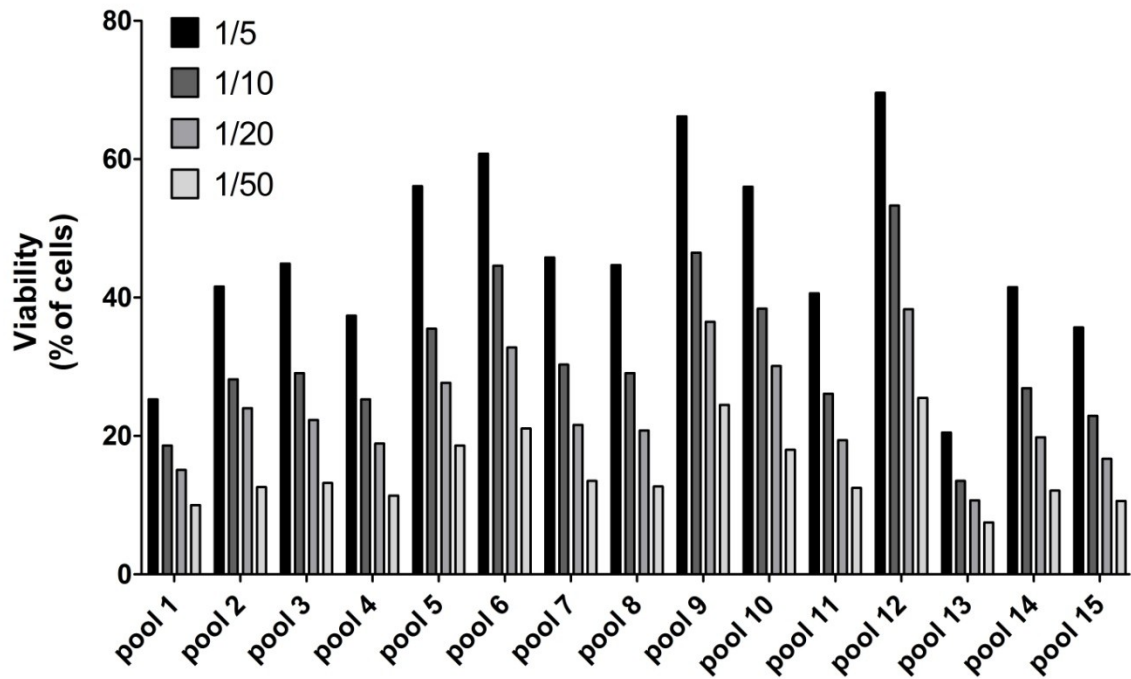


Figure S11 (Additional file 17): Assessment of the multiplicity of infection (MOI) for each shRNA pool. Reh WT cells were transduced with various dilutions of viral particles from the 15 shRNA pools and selected 48h with puromycin. Viability was evaluated by flow cytometry of propidium iodide-stained cells.

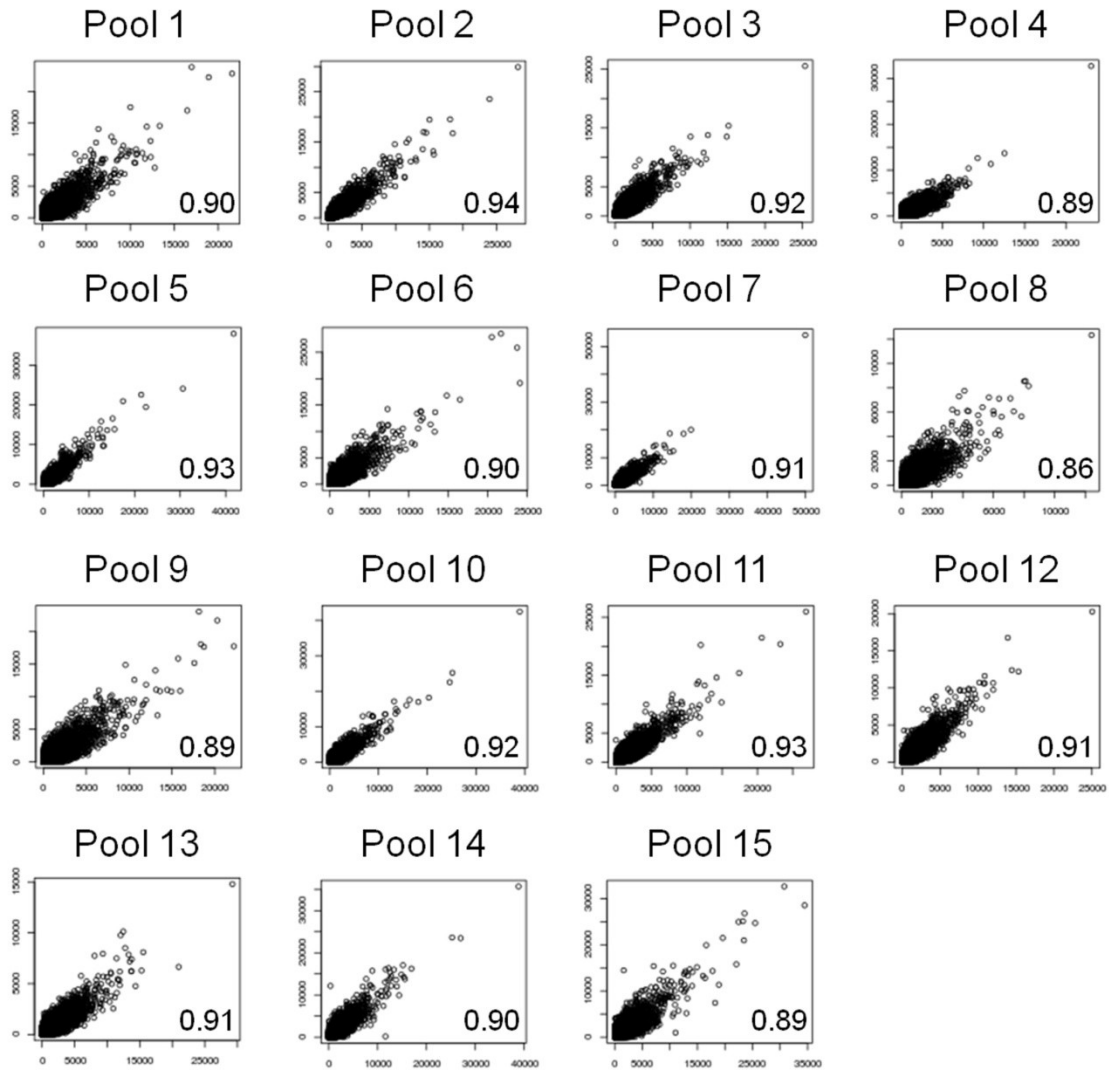


Figure S12 (Additional file 19): Technical reproducibility of the quantification of shRNAs from the genome wide screen. Raw shRNA counts of T₁₄ duplicated samples were plotted against each other's for each pool. The Pearson correlation coefficient (r) is given at the bottom right of each panel.

4.13.2. Supplemental tables

Table S1 ([Additional file 4](#)): List of the 1,241 selected shRNAs

Cette table sera téléchargeable avec la version en ligne de l'article.

Table S2 (Additional file 5): shRNAs selected for validation

Gene	Screen recurrence	From different pools (Yes/No)	Selected shRNA ID	Normalized counts					Σ Counts (BlastS-noBlastS)	T14 corrected ratio
				T14	T28	T28_BlastS	T42	T42_BlastS		
AARS	2	No	TRCN0000045689	2055,26	4257,49	22548,83	1613,39	6515,63	23193,58	11,28
ACACA	3	No	TRCN0000029028	3136,27	7761,21	48846,88	15619,84	142921,49	168387,32	53,69
AKIRIN1	2	Yes	TRCN0000134954	2821,89	1562,70	41302,93	1305,99	24628,91	63063,15	22,35
AP2A1	2	No	TRCN0000065111	1506,01	1013,70	13799,37	1192,75	14256,95	25849,87	17,16
ARMCX3	2	No	TRCN0000123241	362,29	1419,77	10087,02	23,25	4993,38	13637,39	37,64
ATG12	2	No	TRCN0000007394	844,15	5235,42	81347,96	4145,00	113428,73	185396,27	219,63
ATG14	2	Yes	TRCN0000142849	827,02	1482,98	29188,13	1079,36	53482,06	80107,85	96,86
ATG3	2	No	TRCN0000146846	8643,29	6030,86	52659,10	6267,89	49549,00	89909,35	10,40
ATP13A1	2	No	TRCN0000051610	5702,24	3967,37	54585,43	3047,90	100854,49	148424,64	26,03
ATP8A1	2	No	TRCN0000038439	1354,27	564,79	20525,01	1121,75	25900,93	44739,41	33,04
BARD1	2	No	TRCN0000369045	1152,69	2897,71	3528,64	392,76	10718,30	10956,47	9,51
BDH1	2	No	TRCN0000028477	1783,43	657,89	19725,20	503,96	11076,80	29640,16	16,62
C4B	3	Yes	TRCN0000194235	8481,98	22991,46	155572,47	32970,87	380882,92	480493,06	56,65
CAMKMT	2	No	TRCN0000148319	1596,82	1867,99	13707,24	613,04	10127,69	21353,90	13,37
CAPN7	3	Yes	TRCN0000073779	3355,67	3585,55	51681,22	1387,45	55560,30	102268,52	30,48
CASC5	2	No	TRCN0000145591	4675,18	4352,24	36836,59	11762,30	113774,70	134496,76	28,77
CBX3	3	Yes	TRCN0000364151	5853,70	8352,80	197945,42	11633,81	1506722,99	1684681,80	287,80
CCDC84	2	No	TRCN0000142380	1452,25	835,73	15847,69	2011,12	8260,26	21261,09	14,64
CHURC1	2	No	TRCN0000150386	2652,82	4271,69	10324,72	2246,01	9932,68	13739,71	5,18
CKAP2	2	Yes	TRCN0000167132	5582,69	6130,38	44033,53	5632,85	79915,11	112185,41	20,10
CNKSR1	2	No	TRCN0000037940	1466,53	1460,08	77895,85	2993,19	266700,77	340143,36	231,94

COMMD9	2	Yes	TRCN0000167758	1196,29	1537,69	16879,96	1834,99	27894,20	41401,48	34,61
CRK	2	No	TRCN0000021848	697,03	1253,71	14747,16	36,10	27441,92	40899,28	58,68
CTNND1	2	Yes	TRCN0000122987	9824,02	14987,71	112290,17	57139,67	260781,22	300944,01	30,63
CUEDC1	2	No	TRCN0000007700	1260,60	1870,90	8805,33	3381,61	27136,97	30689,79	24,35
DAPK2	2	No	TRCN0000001719	3905,64	7787,55	57188,99	3079,74	24509,72	70831,42	18,14
DIS3L	2	Yes	TRCN0000153834	776,79	464,67	18979,17	1319,25	4515,26	21710,52	27,95
DNAJC10	2	No	TRCN0000062330	342,06	395,67	6671,34	51,74	6435,53	12659,47	37,01
DNMBP	2	No	TRCN0000148479	2813,51	1961,61	49305,96	2073,85	17918,04	63188,54	22,46
DPF1	2	No	TRCN0000013170	6850,79	8352,77	125363,20	19767,75	236575,86	333818,54	48,73
DYRK4	2	Yes	TRCN0000197116	5863,54	15252,53	56508,26	14863,51	37828,69	64220,92	10,95
ERP27	2	Yes	TRCN0000149165	5037,85	8103,57	44989,60	4199,98	33486,34	66172,38	13,14
ETV6	1	N/A	TRCN0000003856	14559,75	33065,82	66593,09	46676,62	64289,03	51139,69	3,51
GNA13	2	No	TRCN0000036887	6835,99	15694,65	44934,82	20055,48	54605,99	63790,68	9,33
GRWD1	2	No	TRCN0000135333	2679,85	5923,59	88170,80	5055,92	73286,61	150477,90	56,15
ICMT	2	No	TRCN0000035271	1499,21	1643,17	22368,73	1686,79	14936,78	33975,55	22,66
JUNB	2	No	TRCN0000014943	177,66	1130,95	9935,83	656,10	12351,69	20500,48	115,39
KLF5	2	Yes	TRCN0000013637	7524,06	26928,61	48994,27	30248,55	66539,95	58357,06	7,76
L3MBTL3	1	N/A	TRCN0000152202	340,73	1201,68	17469,81	5,86	28528,58	44790,85	131,45
LACTB2	2	No	TRCN0000167705	2137,92	4169,86	44235,01	2531,66	53064,00	90597,49	42,38
LMBR1L	2	No	TRCN0000063469	1145,74	2385,46	30116,55	8361,47	24890,91	44260,53	38,63
LPHN3	2	Yes	TRCN0000011695	2324,45	4781,20	28796,01	5859,24	40808,20	58963,78	25,37
LZTR1	2	No	TRCN0000181026	3853,80	1029,42	29849,51	3279,78	79268,57	104808,87	27,20
MAP4K4	2	Yes	TRCN0000219682	616,91	1098,33	84689,72	2574,39	108134,66	189151,67	306,61
MCM3AP	2	Yes	TRCN0000330179	2611,37	5753,28	168301,05	8434,75	489140,31	643253,33	246,33
MKRN2	2	No	TRCN0000004382	3993,19	3118,64	68380,12	1730,27	45103,20	108634,41	27,20
MMP17	2	Yes	TRCN0000049974	1073,53	1552,40	55897,16	485,60	38352,07	92211,24	85,90
MYT1L	2	No	TRCN0000020865	503,27	200,05	10380,07	271,31	3798,25	13706,96	27,24

NDUFB2	2	No	TRCN0000064764	2640,67	5877,81	69118,39	7167,37	34594,87	90668,08	34,34
NUDC	2	No	TRCN0000147755	3379,64	2834,34	26993,62	2116,66	11138,05	33180,67	9,82
OSBPL10	2	Yes	TRCN0000147511	1078,74	1068,00	80025,42	5300,62	80678,78	154335,57	143,07
OTUD6B	2	Yes	TRCN0000129141	4056,96	1078,76	28648,89	1138,41	32038,56	58470,28	14,41
PHC2	2	Yes	TRCN0000162731	13632,42	23548,99	48004,53	24263,62	187277,56	187469,48	13,75
PI4K2B	2	Yes	TRCN0000195310	4394,37	7227,00	51343,83	10412,06	33654,58	67359,34	15,33
PIK3CB	2	Yes	TRCN0000039982	4786,12	10867,55	115341,54	23627,60	220489,12	301335,51	62,96
POGZ	2	No	TRCN0000005707	1062,56	2198,74	20121,73	5845,68	17636,05	29713,36	27,96
POLR3C	2	No	TRCN0000052969	1351,44	1823,10	13876,31	1074,96	2437,08	13415,34	9,93
PRKCH	2	No	TRCN0000006294	1637,73	1106,43	81333,54	4162,27	132121,11	208185,95	127,12
RAB37	2	No	TRCN0000047904	3499,01	3747,13	59819,12	3304,27	89739,82	142507,54	40,73
RARA	2	Yes	TRCN0000020373	9403,25	16929,12	133146,85	31831,12	231183,74	315570,34	33,56
RPL39L	2	No	TRCN0000117655	2209,11	6040,95	47047,04	3636,49	125431,33	162800,93	73,70
SAMHD1	1	N/A	TRCN0000145408	794,36	5155,73	21429,06	8317,56	11811,82	19767,59	24,88
SENP2	2	Yes	TRCN0000355816	2196,16	6647,88	323452,74	25586,16	316447,39	607666,09	276,70
SETD4	2	Yes	TRCN0000134946	1877,46	6766,32	44454,76	10631,97	263554,22	290610,70	154,79
SFMBT1	1	N/A	TRCN0000146760	16435,71	16253,56	20065,26	11731,04	18553,25	10633,92	0,65
SGMS1	2	No	TRCN0000134521	1359,57	2446,70	33006,13	4819,11	19177,19	44917,51	33,04
SLC39A14	2	Yes	TRCN0000344145	6291,31	26523,45	60462,69	24247,94	75494,54	85185,85	13,54
SLC39A6	2	No	TRCN0000376609	3502,86	7226,45	60358,31	6738,62	55646,74	102039,98	29,13
SRP72	2	Yes	TRCN0000151445	5249,69	14582,64	163516,28	17489,52	199617,83	331061,95	63,06
STIM1	2	No	TRCN0000146686	1528,38	2274,71	9600,41	2303,43	7219,68	12241,94	8,01
TACC1	2	Yes	TRCN0000155998	784,30	2502,66	39922,65	3354,12	85222,71	119288,58	152,10
TACO1	2	No	TRCN0000149384	5961,27	4762,95	39138,73	3290,05	91686,21	122771,93	20,59
TAZ	2	Yes	TRCN0000083056	7429,51	30634,93	257041,25	66367,59	345260,65	505299,37	68,01
TFCP2	2	Yes	TRCN0000274172	14380,09	47213,83	337550,71	154456,35	752400,27	888280,80	61,77
TMX4	2	No	TRCN0000064042	3149,80	1826,29	126634,00	5705,03	280179,30	399281,98	126,76

TRIM41	1	N/A	TRCN0000034046	5409,96	4676,59	8919,97	6216,84	33566,83	31593,37	5,84
TXK	2	Yes	TRCN0000121107	1577,26	2761,30	48576,54	5397,34	92261,71	132679,61	84,12
TXNIP	2	No	TRCN0000059058	3580,02	8209,79	175385,81	12436,70	267222,80	421962,11	117,87
UBA6	2	No	TRCN0000007710	1343,47	3076,25	61203,16	3033,83	166350,27	221443,35	164,83
UBE2G2	1	N/A	TRCN0000004023	4578,19	6343,27	18441,43	4349,73	9374,21	17122,65	3,74
UBE2K	1	N/A	TRCN0000007455	1455,13	788,49	21324,62	605,53	24494,31	44424,90	30,53
UBE2M	2	No	TRCN0000007260	2514,18	1615,14	27261,18	4293,00	17655,54	39008,59	15,52
UBE2S	1	N/A	TRCN0000007676	2352,65	1859,60	37246,32	3025,20	60084,40	92445,93	39,29
UHRF1	1	N/A	TRCN0000004353	2873,75	2559,07	23351,07	1378,79	8654,15	28067,36	9,77
WWC3	2	Yes	TRCN0000129101	574,09	947,58	11824,42	575,73	5438,02	15739,13	27,42
YPEL5	2	Yes	TRCN0000130158	185,05	1595,18	111384,98	5489,68	206510,91	310811,03	1679,58
ZBTB26	2	No	TRCN0000107414	2408,72	1118,21	52072,08	690,10	42513,97	92777,73	38,52
ZBTB7B	2	No	TRCN0000020206	246,37	110,69	5889,88	854,36	6832,49	11757,31	47,72
ZNF761	2	Yes	TRCN0000146860	845,07	245,67	15833,72	356,20	18646,21	33878,07	40,09
ZRANB1	2	No	TRCN0000073813	6366,12	4831,41	32215,89	4701,54	38468,29	61151,22	9,61

Table S3 (Additional file 6): Struct2Net prediction analysis*La table complète sera téléchargeable avec la version en ligne de l'article.*

Unique ID	Recurrence	Single shRNA in screen	Unique ID	Recurrence	Single shRNA in screen
UBQLN3	7		UBE2D4	3	
UBQLN1	6		UBE2D1	3	
UBQLN4	6		UBE2D2	3	
UBL7	6		TMUB1	3	
UBL4B	6		UBE2G2	3	√
UBQLNL	6		UBE2R2	3	
ANUBL1	5		LOC731049	3	
PARK2	5		UBE2S	3	√
UHRF2	5		CDC34	3	
UBL4A	5		UBE2M	3	
UBB	5		UBE2U	3	
NEDD8	5		UBE2G1	3	
UBTD1	5		UBE2T	3	
UHRF1	5	√	UBE2N	3	
LOC388720	5		UBE2A	3	
RPS27A	5		UBE2F	3	
RPS27AP11	5		AKTIP	3	
UBTD2	5		SUMO1	3	
UBQLN2	4		UBE2NL	3	
UBC	4		UBE2B	3	
OASL	4		UBE2L6	3	
RAD23B	4		FAU	3	
ISG15	4		UBE2W	3	
RAD23A	4		SUMO2	3	
NUB1	4		SUMO3	3	
FAF1	4		COBLL1	3	
UBL3	4		SHARPIN	3	
PHC2	4		NPLOC4	3	
PHC1	4		UBXN6	3	
PHC1B	4		TXN2	3	
PHC3	4		C19orf47	3	
SHANK2	4		PPFIBP1	3	
SHANK3	4		PPFIBP2	3	
SAMD14	4		STARD13	3	
CNKSR2	4		STIM1	3	
SCMH1	4		ARL11	3	
SAMD7	4		ARL6	3	
SFMBT2	4		SASH3	3	
SCML2	4		ARL8A	3	

SAMD10	4		ARL14	3	
CNKSR3	4		ETS1	3	
ANKS6	4		PLEKHG4	3	
SFMBT1	4	√	GAS2L2	3	
SAMD11	4		ETV6	3	√
SCML1	4		STIM2	3	
SAMD3	4		ETS2	3	
SAMD12	4		PPFIA1	3	
L3MBTL3	4	√	KIAA1026	3	
SAMD8	4		FAM59A	3	
L3MBTL4	4		STK39	3	
L3MBTL	4		OXR1	3	
TP73	4		RASA1	3	
TP63	4		DGKH	3	
SGMS1	4		DGKD	3	
SAMHD1	4	√	CNKSR1	3	
TRIM7	4		RIN1	3	
PCGF6	3		DACH1	3	
ARL5A	3		TRIM5	3	
TRIM28	3		TRIM75	3	
UBAP2	3		TRIM47	3	
UBE2E3	3		TRIM68	3	
UBE2E1	3		TRIM41	3	√
UBE2K	3	√	TRIM22	3	
UBE2H	3		TRIM69	3	
UBE2E2	3		TRIML1	3	
UBE2I	3		TRIM35	3	
UBE2C	3		TRIM64B	3	
UBE2D3	3		TRIM38	3	

* Seules les protéines ayant été prédites pour interagir avec plus de 3 modulateurs potentiels sont montrées

For each of the 81 putative *ETV6* modulators identified with 2 or 3 distinct shRNAs in the genome wide screen, the binding partners were predicted by Struct2Net. The recurrence of every putative binding partners identified was determined. Those with the potential to interact with at least 3 putative modulators (n=136) that were also captured in the genome wide screen with a unique shRNA are indicated (n=9). Note that 4 putative modulators (*ZBTB7B*, *ZBTB26*, *ZNF761* and *KLF5*) were discarded from this analysis as their number of predicted binding partners exceeded 600 and led to excessive background (the average number of predicted binding partners is 35 for the remaining 77 modulators).

Table S4 (Additional file 7): *ETV6* targets and control genes used in the validation process

Category	ChIP-seq signal	Gene Name	Normalized count					SD Reh <i>ETV6</i>
			Reh <i>ETV6</i>				Reh pLENTI	
			SHC202	SHC002	Control 1	Control 2	Control	
<i>ETV6</i> targets	ChIP-seq negative	DDIT4L	51,72	59,25	60,60	74,39	169,44	9,45
		IL8	129,77	35,69	88,28	121,16	561,68	42,62
		OSR2	470,25	467,54	422,09	343,20	865,39	59,32
		LRMP	7851,48	6360,01	8097,02	8524,10	13614,01	940,75
		ABCA9	275,83	339,06	304,55	228,80	523,57	46,66
		IKZF2	2262,64	3012,26	2233,73	2699,36	4396,70	373,53
		KCNJ2	0,00	39,97	31,87	23,67	94,98	17,25
		LIMA1	151,80	177,02	194,85	160,05	244,49	19,09
		FRK	0,00	41,40	17,76	21,41	81,50	16,98
		CNN2	2608,86	1403,34	2840,23	3050,44	5028,74	737,31
		HMGCS1	1947,55	1117,11	2677,24	2702,17	3881,93	749,37
		C20orf197	0,00	19,99	7,84	2,25	27,56	8,94
		IFNGR1	3264,43	1990,09	3066,94	3189,07	4127,00	597,26
		IGSF10	747,03	906,53	1025,45	865,60	1102,84	114,89
		CERKL	11,97	13,56	26,12	40,01	42,80	13,04
	ABCA1	888,77	1323,40	827,99	1066,22	1342,64	222,19	
	TTC16	6,70	9,99	12,01	0,00	14,07	5,26	
	SPTBN2	197,29	102,07	142,61	154,97	188,20	39,20	
	ChIP-seq positive	FGFBP2	687,65	443,27	669,18	697,66	2542,22	121,35
		OSTalpha	320,84	274,10	257,02	259,23	692,43	29,68
		IRX3	430,02	267,68	433,58	444,07	1241,79	84,32
		NTRK1	66,08	18,56	124,85	194,98	661,35	76,20
		MIB1	16729,17	15516,72	16775,45	17452,27	22351,13	805,39
		SEMA3A	60,34	38,55	83,06	80,59	205,79	20,73
		C6orf222	126,90	26,41	123,81	191,60	474,32	68,11
		DSC3	771,93	831,58	620,60	784,45	1228,90	91,38
		EHD2	184,84	112,78	281,57	283,46	643,76	82,61
		PIK3C3	7995,62	7074,53	8585,97	8872,94	11930,73	794,13
		CLIC5	884,46	386,17	809,70	1032,41	1974,09	277,24
		SEMA3D	83,32	24,98	56,42	77,20	171,20	26,32
PTPRC		3462,68	2162,83	3882,91	4283,47	6989,35	919,97	
HOMER1		981,20	638,14	915,75	676,81	1384,27	170,85	
RAP1GAP		374,95	508,23	336,94	263,17	685,98	102,69	
RGS1	28,73	11,42	16,72	41,14	65,08	13,25		
KIAA0226L	836,58	1001,47	723,51	757,96	1177,30	123,79		

		TSPAN9	300,73	423,29	278,96	324,04	490,74	63,74
		ITGB2	118,76	219,14	156,72	183,71	243,32	42,44
		SMARCA2	767,62	1018,60	864,03	852,64	1044,21	104,51
		ADRB2	22,03	73,52	11,49	22,54	76,81	27,89
		KCNA2	134,08	118,49	94,55	50,72	140,13	36,34
ETV6- independent control genes	ChIP-seq negative	BAK1	1359,50	1387,64	1578,66	1750,92	1784,71	182,61
		BCL10	2218,11	2125,71	1832,02	2332,49	2441,96	214,12
		RELA	3170,57	2727,45	3206,94	3257,82	3510,80	244,79
		NOL3	255,24	294,80	231,42	168,50	254,46	52,90
		BNIP3L	11526,77	7240,85	15602,69	12065,95	11916,07	3427,63
		BIRC3	66,56	38,55	59,03	41,70	51,59	13,51
		AKT1	3507,69	4799,63	3600,82	3409,42	3740,63	651,53
		MCL1	4954,34	9028,22	5425,52	5498,46	5743,45	1883,23
		MYC	7800,24	8612,78	7578,81	7647,80	7577,41	477,63
		HIP1	1132,52	912,96	985,22	1044,24	962,13	92,93
		PIK3R2	182,45	119,92	194,33	203,44	156,54	37,73
		CD70	2094,56	5505,59	2494,93	2076,64	2424,96	1653,09
		RIPK1	1321,19	949,36	1438,13	1478,73	1099,32	241,10
		TRAF5	494,67	649,56	601,79	515,08	485,46	72,90
		BIRC2	1430,37	2327,72	1524,33	1361,51	1356,71	449,47
		TNFRSF10A	317,49	314,79	319,18	284,02	263,25	16,66
		DFFA	1639,64	2958,01	1622,02	1769,52	1574,23	643,85
		NFKB2	450,13	735,22	529,70	430,54	428,00	139,31
		CYCS	16384,87	24542,79	16012,25	16941,70	14657,05	4066,24
		IKBKB	2624,66	2796,69	2594,18	2832,91	2155,26	120,14
		TRAF3	1216,80	2313,45	1300,75	1121,44	1078,22	555,08
		BNIP2	5663,06	5509,16	4879,63	4858,84	4031,44	418,72
		IKBIP	267,69	188,44	172,39	102,00	125,47	68,01
		BRAF	703,93	1047,15	651,94	553,96	540,57	214,48
		ABL1	6117,51	5011,63	6635,90	6442,39	4604,84	725,72
		DAXX	7365,91	12526,58	6827,09	6556,22	5976,80	2825,20
		TRAF7	2823,39	4122,94	2966,12	2676,81	2334,08	661,05
		AVEN	875,85	1359,09	707,84	575,94	595,69	342,36
		FADD	1050,15	2074,32	1042,16	995,77	875,35	523,03
		CD40	465,46	1431,89	522,39	648,63	469,63	449,76
		NAIP	289,71	219,85	189,63	153,85	147,16	57,67
		DEDD	2903,36	3860,26	2777,02	2951,82	2210,96	496,93
		DIABLO	4621,05	5987,41	4057,91	4009,02	3245,20	921,86
		BCLAF1	5040,54	4816,76	5345,60	4460,42	3445,13	373,04
		BIRC6	6301,39	10393,02	5592,69	5250,50	4498,14	2379,67
		CFLAR	560,75	803,03	488,96	366,86	351,20	183,77
RIPK2	893,08	2396,96	971,12	1055,51	769,82	714,94		
BAG4	1837,89	2526,87	1857,61	1591,44	1278,73	401,00		

BAX	208,31	463,26	267,46	337,00	196,41	109,61	
CASP2	9954,18	20510,50	9713,81	9795,45	7532,85	5345,61	
CASP3	2127,12	2893,77	2366,42	2071,57	1541,40	375,18	
TP53	36704,57	64151,89	34537,70	33076,97	25718,87	14764,81	
BCL2L2	142,70	191,30	92,99	167,37	89,12	42,05	
TRAF2	1841,72	2804,54	1946,94	1736,83	1276,39	488,94	
HMGB1	60473,07	71461,98	62114,58	55649,00	39142,92	6620,36	
TNFRSF21	4197,26	5290,02	3560,60	3490,00	2456,03	833,45	
TRAF6	1035,79	1698,14	1112,69	1182,31	731,71	299,97	
CASP8AP2	4165,65	7304,38	4154,03	3842,21	2745,08	1632,10	
BCL2L13	1494,54	2354,85	1631,94	1692,87	1019,59	383,28	
BIRC5	11770,52	18951,55	11243,36	10805,88	7306,54	3859,35	
TNFRSF10D	179,57	241,98	202,69	218,65	121,37	26,30	
TP53I3	531,54	488,96	445,07	646,38	293,15	86,46	
CASP8	2375,65	3743,20	2260,90	1989,30	1366,68	784,21	
CASP9	504,25	1239,88	547,99	637,93	360,58	342,80	
NFKBIA	12751,23	21923,84	12515,38	12120,62	7694,67	4737,86	
BFAR	1209,14	1523,98	930,37	1000,85	592,17	266,33	
TRADD	88,11	108,50	75,75	74,39	44,56	15,80	
CASP7	2358,89	3505,50	2132,91	1840,52	1117,50	728,93	
PTEN	8625,81	7694,12	8489,85	6605,81	3565,32	928,00	
PEA15	570,33	997,90	570,97	538,74	284,36	219,46	
BAD	1977,24	2993,70	1783,96	1488,31	843,69	653,63	
DEDD2	60,34	104,22	47,54	42,83	17,00	27,99	
ChIP-seq	FAS	201,12	368,32	239,25	305,44	271,46	73,76
positive	TP53BP2	2527,45	4511,25	2996,94	2581,01	2344,05	928,74

Table S4 suite

Category	ChIP-seq signal	Gene Name	<i>ETV6</i> repression (Ratio)				Mean
			SHC202	SHC002	Control 1	Control 2	
<i>ETV6</i> targets	ChIP-seq negative	DDIT4L	3,28	2,86	2,80	2,28	2,80
		IL8	4,33	15,74	6,36	4,64	7,77
		OSR2	1,84	1,85	2,05	2,52	2,07
		LRMP	1,73	2,14	1,68	1,60	1,79
		ABCA9	1,90	1,54	1,72	2,29	1,86
		IKZF2	1,94	1,46	1,97	1,63	1,75
		KCNJ2		2,38	2,98	4,01	3,12
		LIMA1	1,61	1,38	1,25	1,53	1,44
		FRK		1,97	4,59	3,81	3,45
		CNN2	1,93	3,58	1,77	1,65	2,23
		HMGCS1	1,99	3,47	1,45	1,44	2,09
		C20orf197		1,38	3,52	12,22	5,71
		IFNGR1	1,26	2,07	1,35	1,29	1,49
		IGSF10	1,48	1,22	1,08	1,27	1,26
		CERKL	3,58	3,16	1,64	1,07	2,36
		ABCA1	1,51	1,01	1,62	1,26	1,35
		TTC16	2,10	1,41	1,17		1,56
		SPTBN2	0,95	1,84	1,32	1,21	1,33
	ChIP-seq positive	FGFBP2	3,70	5,74	3,80	3,64	4,22
		OSTalpha	2,16	2,53	2,69	2,67	2,51
		IRX3	2,89	4,64	2,86	2,80	3,30
		NTRK1	10,01	35,64	5,30	3,39	13,58
		MIB1	1,34	1,44	1,33	1,28	1,35
		SEMA3A	3,41	5,34	2,48	2,55	3,45
		C6orf222	3,74	17,96	3,83	2,48	7,00
		DSC3	1,59	1,48	1,98	1,57	1,65
		EHD2	3,48	5,71	2,29	2,27	3,44
		PIK3C3	1,49	1,69	1,39	1,34	1,48
		CLIC5	2,23	5,11	2,44	1,91	2,92
		SEMA3D	2,05	6,85	3,03	2,22	3,54
PTPRC	2,02	3,23	1,80	1,63	2,17		
HOMER1	1,41	2,17	1,51	2,05	1,78		
RAP1GAP	1,83	1,35	2,04	2,61	1,96		
RGS1	2,27	5,70	3,89	1,58	3,36		
KIAA0226L	1,41	1,18	1,63	1,55	1,44		
TSPAN9	1,63	1,16	1,76	1,51	1,52		
ITGB2	2,05	1,11	1,55	1,32	1,51		

		SMARCA2	1,36	1,03	1,21	1,22	1,20
		ADRB2	3,49	1,04	6,68	3,41	3,66
		KCNA2	1,05	1,18	1,48	2,76	1,62
ETV6- independent control genes	ChIP-seq negative	BAK1	1,31	1,29	1,13	1,02	1,19
		BCL10	1,10	1,15	1,33	1,05	1,16
		RELA	1,11	1,29	1,09	1,08	1,14
		NOL3	1,00	0,86	1,10	1,51	1,12
		BNIP3L	1,03	1,65	0,76	0,99	1,11
		BIRC3	0,78	1,34	0,87	1,24	1,06
		AKT1	1,07	0,78	1,04	1,10	1,00
		MCL1	1,16	0,64	1,06	1,04	0,97
		MYC	0,97	0,88	1,00	0,99	0,96
		HIP1	0,85	1,05	0,98	0,92	0,95
		PIK3R2	0,86	1,31	0,81	0,77	0,93
		CD70	1,16	0,44	0,97	1,17	0,93
		RIPK1	0,83	1,16	0,76	0,74	0,87
		TRAF5	0,98	0,75	0,81	0,94	0,87
		BIRC2	0,95	0,58	0,89	1,00	0,85
		TNFRSF10A	0,83	0,84	0,82	0,93	0,85
		DFFA	0,96	0,53	0,97	0,89	0,84
		NFKB2	0,95	0,58	0,81	0,99	0,83
		CYCS	0,89	0,60	0,92	0,87	0,82
		IKBKB	0,82	0,77	0,83	0,76	0,80
		TRAF3	0,89	0,47	0,83	0,96	0,79
		BNIP2	0,71	0,73	0,83	0,83	0,77
		IKBIP	0,47	0,67	0,73	1,23	0,77
		BRAF	0,77	0,52	0,83	0,98	0,77
		ABL1	0,75	0,92	0,69	0,71	0,77
		DAXX	0,81	0,48	0,88	0,91	0,77
		TRAF7	0,83	0,57	0,79	0,87	0,76
		AVEN	0,68	0,44	0,84	1,03	0,75
		FADD	0,83	0,42	0,84	0,88	0,74
		CD40	1,01	0,33	0,90	0,72	0,74
		NAIP	0,51	0,67	0,78	0,96	0,73
		DEDD	0,76	0,57	0,80	0,75	0,72
		DIABLO	0,70	0,54	0,80	0,81	0,71
BCLAF1	0,68	0,72	0,64	0,77	0,70		
BIRC6	0,71	0,43	0,80	0,86	0,70		
CFLAR	0,63	0,44	0,72	0,96	0,68		
RIPK2	0,86	0,32	0,79	0,73	0,68		
BAG4	0,70	0,51	0,69	0,80	0,67		
BAX	0,94	0,42	0,73	0,58	0,67		
CASP2	0,76	0,37	0,78	0,77	0,67		

CASP3	0,72	0,53	0,65	0,74	0,66	
TP53	0,70	0,40	0,74	0,78	0,66	
BCL2L2	0,62	0,47	0,96	0,53	0,65	
TRAF2	0,69	0,46	0,66	0,73	0,63	
HMGB1	0,65	0,55	0,63	0,70	0,63	
TNFRSF21	0,59	0,46	0,69	0,70	0,61	
TRAF6	0,71	0,43	0,66	0,62	0,60	
CASP8AP2	0,66	0,38	0,66	0,71	0,60	
BCL2L13	0,68	0,43	0,62	0,60	0,59	
BIRC5	0,62	0,39	0,65	0,68	0,58	
TNFRSF10D	0,68	0,50	0,60	0,56	0,58	
TP53I3	0,55	0,60	0,66	0,45	0,57	
CASP8	0,58	0,37	0,60	0,69	0,56	
CASP9	0,72	0,29	0,66	0,57	0,56	
NFKBIA	0,60	0,35	0,61	0,63	0,55	
BFAR	0,49	0,39	0,64	0,59	0,53	
TRADD	0,51	0,41	0,59	0,60	0,53	
CASP7	0,47	0,32	0,52	0,61	0,48	
PTEN	0,41	0,46	0,42	0,54	0,46	
PEA15	0,50	0,28	0,50	0,53	0,45	
BAD	0,43	0,28	0,47	0,57	0,44	
DEDD2	0,28	0,16	0,36	0,40	0,30	
ChIP-seq positive	FAS	1,35	0,74	1,13	0,89	1,03
	TP53BP2	0,93	0,52	0,78	0,91	0,78

Table S4 suite

Category	ChIP-seq signal	Gene Name	<i>ETV6</i> repression (SD corrected)				Mean
			SHC202	SHC002	Control 1	Control 2	
<i>ETV6</i> targets	ChIP-seq negative	DDIT4L	12,46	11,67	11,52	10,06	11,43
		IL8	10,13	12,34	11,11	10,34	10,98
		OSR2	6,66	6,71	7,47	8,80	7,41
		LRMP	6,13	7,71	5,86	5,41	6,28
		ABCA9	5,31	3,95	4,69	6,32	5,07
		IKZF2	5,71	3,71	5,79	4,54	4,94
		KCNJ2	5,51	3,19	3,66	4,13	4,12
		LIMA1	4,85	3,53	2,60	4,42	3,85
		FRK	4,80	2,36	3,75	3,54	3,61
		CNN2	3,28	4,92	2,97	2,68	3,46
		HMGCS1	2,58	3,69	1,61	1,57	2,36
		C20orf197	3,08	0,85	2,21	2,83	2,24
		IFNGR1	1,44	3,58	1,77	1,57	2,09
		IGSF10	3,10	1,71	0,67	2,07	1,89
		CERKL	2,36	2,24	1,28	0,21	1,53
		ABCA1	2,04	0,09	2,32	1,24	1,42
		TTC16	1,40	0,77	0,39	2,67	1,31
		SPTBN2	-0,23	2,20	1,16	0,85	0,99
	ChIP-seq positive	FGFBP2	15,28	17,30	15,43	15,20	15,80
		OSTalpha	12,52	14,09	14,67	14,59	13,97
		IRX3	9,63	11,55	9,59	9,46	10,06
		NTRK1	7,81	8,44	7,04	6,12	7,35
		MIB1	6,98	8,49	6,92	6,08	7,12
		SEMA3A	7,02	8,07	5,92	6,04	6,76
		C6orf222	5,10	6,58	5,15	4,15	5,24
		DSC3	5,00	4,35	6,66	4,86	5,22
		EHD2	5,56	6,43	4,38	4,36	5,18
		PIK3C3	4,96	6,12	4,21	3,85	4,78
		CLIC5	3,93	5,73	4,20	3,40	4,31
		SEMA3D	3,34	5,56	4,36	3,57	4,21
PTPRC	3,83	5,25	3,38	2,94	3,85		
HOMER1	2,36	4,37	2,74	4,14	3,40		
RAP1GAP	3,03	1,73	3,40	4,12	3,07		
RGS1	2,74	4,05	3,65	1,81	3,06		
KIAA0226L	2,75	1,42	3,67	3,39	2,81		
TSPAN9	2,98	1,06	3,32	2,62	2,49		
ITGB2	2,93	0,57	2,04	1,40	1,74		

		SMARCA2	2,65	0,25	1,72	1,83	1,61
		ADRB2	1,96	0,12	2,34	1,95	1,59
		KCNA2	0,17	0,60	1,25	2,46	1,12
		BAK1	2,33	2,17	1,13	0,19	1,45
		BCL10	1,05	1,48	2,85	0,51	1,47
		RELA	1,39	3,20	1,24	1,03	1,72
		NOL3	-0,01	-0,76	0,44	1,62	0,32
		BNIP3L	0,11	1,36	-1,08	-0,04	0,09
		BIRC3	-1,11	0,97	-0,55	0,73	0,01
		AKT1	0,36	-1,63	0,21	0,51	-0,14
		MCL1	0,42	-1,74	0,17	0,13	-0,26
		MYC	-0,47	-2,17	0,00	-0,15	-0,70
		HIP1	-1,83	0,53	-0,25	-0,88	-0,61
		PIK3R2	-0,69	0,97	-1,00	-1,24	-0,49
		CD70	0,20	-1,86	-0,04	0,21	-0,37
		RIPK1	-0,92	0,62	-1,41	-1,57	-0,82
		TRAF5	-0,13	-2,25	-1,60	-0,41	-1,09
		BIRC2	-0,16	-2,16	-0,37	-0,01	-0,68
		TNFRSF10A	-3,26	-3,09	-3,36	-1,25	-2,74
		DFFA	-0,10	-2,15	-0,07	-0,30	-0,66
		NFKB2	-0,16	-2,21	-0,73	-0,02	-0,78
		CYCS	-0,42	-2,43	-0,33	-0,56	-0,94
ETV6- independent control genes	ChIP-seq negative	IKBKB	-3,91	-5,34	-3,65	-5,64	-4,63
		TRAF3	-0,25	-2,23	-0,40	-0,08	-0,74
		BNIP2	-3,90	-3,53	-2,03	-1,98	-2,86
		IKBIP	-2,09	-0,93	-0,69	0,35	-0,84
		BRAF	-0,76	-2,36	-0,52	-0,06	-0,93
		ABL1	-2,08	-0,56	-2,80	-2,53	-1,99
		DAXX	-0,49	-2,32	-0,30	-0,21	-0,83
		TRAF7	-0,74	-2,71	-0,96	-0,52	-1,23
		AVEN	-0,82	-2,23	-0,33	0,06	-0,83
		FADD	-0,33	-2,29	-0,32	-0,23	-0,79
		CD40	0,01	-2,14	-0,12	-0,40	-0,66
		NAIP	-2,47	-1,26	-0,74	-0,12	-1,15
		DEDD	-1,39	-3,32	-1,14	-1,49	-1,84
		DIABLO	-1,49	-2,97	-0,88	-0,83	-1,54
		BCLAF1	-4,28	-3,68	-5,09	-2,72	-3,94
		BIRC6	-0,76	-2,48	-0,46	-0,32	-1,00
		CFLAR	-1,14	-2,46	-0,75	-0,09	-1,11
		RIPK2	-0,17	-2,28	-0,28	-0,40	-0,78
		BAG4	-1,39	-3,11	-1,44	-0,78	-1,68
		BAX	-0,11	-2,43	-0,65	-1,28	-1,12
		CASP2	-0,45	-2,43	-0,41	-0,42	-0,93

CASP3	-1,56	-3,60	-2,20	-1,41	-2,19	
TP53	-0,74	-2,60	-0,60	-0,50	-1,11	
BCL2L2	-1,27	-2,43	-0,09	-1,86	-1,41	
TRAF2	-1,16	-3,13	-1,37	-0,94	-1,65	
HMGB1	-3,22	-4,88	-3,47	-2,49	-3,52	
TNFRSF21	-2,09	-3,40	-1,33	-1,24	-2,01	
TRAF6	-1,01	-3,22	-1,27	-1,50	-1,75	
CASP8AP2	-0,87	-2,79	-0,86	-0,67	-1,30	
BCL2L13	-1,24	-3,48	-1,60	-1,76	-2,02	
BIRC5	-1,16	-3,02	-1,02	-0,91	-1,53	
TNFRSF10D	-2,21	-4,59	-3,09	-3,70	-3,40	
TP53I3	-2,76	-2,26	-1,76	-4,09	-2,72	
CASP8	-1,29	-3,03	-1,14	-0,79	-1,56	
CASP9	-0,42	-2,57	-0,55	-0,81	-1,08	
NFKBIA	-1,07	-3,00	-1,02	-0,93	-1,51	
BFAR	-2,32	-3,50	-1,27	-1,53	-2,15	
TRADD	-2,76	-4,05	-1,97	-1,89	-2,67	
CASP7	-1,70	-3,28	-1,39	-0,99	-1,84	
PTEN	-5,45	-4,45	-5,31	-3,28	-4,62	
PEA15	-1,30	-3,25	-1,31	-1,16	-1,75	
BAD	-1,73	-3,29	-1,44	-0,99	-1,86	
DEDD2	-1,55	-3,12	-1,09	-0,92	-1,67	
ChIP-seq	FAS	0,95	-1,31	0,44	-0,46	-0,10
positive	TP53BP2	-0,20	-2,33	-0,70	-0,26	-0,87

ETV6 repression of genes are calculated with the Reh pLenti expression values either as a ratio or corrected by the standard deviation (SD) of the Reh *ETV6* samples. ChIP-seq signals were obtained from a previous publication [Article 2].

Table S5 (Additional file 14): Proteomic analysis of ETV6-bound proteins

Category	Gene Name	Molecular Weight	Exclusive unique peptide count			
			Reh ETV6		Reh ETV6-HA	
			Cytoplasm	Nuclear	Cytoplasm	Nuclear
Nuclear interaction	RUNX1	50 kDa	0	0	0	12
	MKI67	359 kDa	1	0	0	7
	RUVBL2	51 kDa	0	0	0	4
	SRP9	10 kDa	0	0	0	3
	DIMT1	35 kDa	0	0	0	3
	ESCO2	68 kDa	0	0	0	3
	FMR1	66 kDa	2	0	1	3
	SMARCA4	189 kDa	0	0	0	3
	UTP3	55 kDa	1	0	0	2
	CBFB	10 kDa	0	0	0	2
	RBM14	69 kDa	0	0	0	2
	LP1	15 kDa	0	0	0	2
	CBX3	21 kDa	0	0	0	2
	HMGN4	10 kDa	0	0	0	2
	CDCA5	8 kDa	0	0	0	2
	RASAL3	104 kDa	0	0	1	2
	WDR36	99 kDa	0	0	0	2
	SLC25A5	33 kDa	0	0	0	2
	RPS29	7 kDa	0	0	0	2
	RUVBL1	30 kDa	0	0	1	2
CDCA7	51 kDa	0	0	0	2	
Cytosolic interaction	SLAIN2	45 kDa	0	0	3	0
	WWP2	86 kDa	0	0	2	0
	TTN	3014 kDa	0	4	2	0
Bait	ETV6	53 kDa	0	0	8	7

Anti-HA magnetic beads were added directly to nuclear and cytoplasm protein extracts from Reh pCCL ETV6-HA and Reh pCCL ETV6 cells (negative immunoprecipitation control). Precipitates were submitted to trypsin digestion prior to proteins detection by mass spectrometry.

Table S6 (Additional file 16): Oligonucleotide sequences

Purpose	Name	Sequence (5' - 3')
Cloning	BlastR_F	CACCATGGCCAAGCCTTTGTCTC
	BlastR_R	TGTCTAGATTAGCCCTCCCACACATAACC
	EBS3tk_F	CAGGTACCGGCCGCACTAGAACTAGTGGA
	EBS3tk_R	TGGCTAGCAGATCTGCGGCACGCTGT
qRT-PCR	qBlastR_F	TCCTGGGATCAAAGCCATAG
	qBlastR_R	TGTCTAGATTAGCCCTCCCACACATAACC
	ETV6_F	TGAACCACATCATGGTCTCTG
	ETV6_R	CGTAATCCCAAAGCAGTCTACA
	GAPDH_F	GATCATCAGCAATGCCTC
	GAPDH_R	TGTGGTCATGAGTCCTTCCA
	B2M_F	TACTCTCTTTCTGGCCTG
	B2M_R	GGATGGATGAAACCCAGACA
	AKIRIN1_F	GGACTGTTTCCAGGCACAGAA
	AKIRIN1_R	GCCCGTCTCCAGTTTGTCTA
	COMMD9_F	CCTCCTCAGACAGCATCAGC
	COMMD9_R	GGTTTGTCTCCGCATAGGCT
	DYRK4_F	TCTCAAGCCCGAAAATATAGTGC
	DYRK4_R	ACCGGCTTTGGATGTACGTG
	JUNB_F	AACAGCCCTTCTACCACGAC
	JUNB_R	CAGGCTCGGTTTCAGGAGTT
	SRP72_F	ACCAAAGGAACAAGGACAGGG
	SRP72_R	GGCAGCCATCTTTCTGGATCT
	CLIC5_F	TGGCTGACTGCAATCTGTTG
	CLIC5_R	CATCTCAGCCGGGATATCATAG
	IRX3_F	GAAGGCGGAACAGATCGCT
	IRX3_R	GAGAGCCGATAAGACCAGGG
	KCNA2_F	CACGTCTCTGGTCAGCCAAG
	KCNA2_R	CTCAGGGTGCTGCTACTGG
	NTRK1_F	CACATCGTGCGTTCTTTCG
	NTRK1_R	CAGCTTGGCATCAGGTCCA
	OSTalpha_F	GGGCATCATTTCCCGTCAAG
	OSTalpha_R	CGTTGGCCAAGACTGAGAAG
	C6orf222_F	TTGCACCCAGTTCCCATAACA
	C6orf222_R	TGGCTAAAGCTGCGAACCAT
	DSC3_F	ACAGAAGCACCTGGAGACGA
	DSC3_R	AATGGTTTCCTGCCCTCCATT
	PIK3C3_F	GCAGTACAACGCGAAAGTGG
	PIK3C3_R	TGGGGTTCTAAAGGCAACGG
TSPAN9_F	GTGCCTCCTCCTCAGCTTTT	
TSPAN9_R	TCAGGTCCTTCTTGGCGTTC	

Table S7 (Additional file 18): Raw and normalized counts of SHC002 and SHC202 added plasmids

Pool #	SHC202	SHC002	counts T14_PCR2	Normalized SHC202	Normalized SHC002	Ratio
Pool 1	1683	599	9090962	1851,29	658,90	2,81
Pool 2	3628	237	9439744	3843,32	251,07	15,31
Pool 3	2204	237	8741256	2521,38	271,13	9,30
Pool 4	2686	272	9006156	2982,40	302,02	9,88
Pool 5	1241	324	8557816	1450,14	378,60	3,83
Pool 6	3930	560	8997886	4367,69	622,37	7,02
Pool 7	881	104	9221520	955,37	112,78	8,47
Pool 8	1450	44	6730238	2154,46	65,38	32,95
Pool 9	3868	206	9555694	4047,85	215,58	18,78
Pool 10	3207	909	9017916	3556,25	1007,99	3,53
Pool 11	3800	574	8731080	4352,27	657,42	6,62
Pool 12	2245	573	9364978	2397,23	611,85	3,92
Pool 13	6692	663	5021278	13327,28	1320,38	10,09
Pool 14	12530	1791	9265620	13523,11	1932,95	7,00
Pool 15	16092	1709	9990826	16106,78	1710,57	9,42
Mean	4409	587	8715531	5162,45	674,60	9,93
			Standard deviation	4885		
			2* SD	9770		

Table S8 ([Additional file 20](#)): Probes and target sequences of genes monitored by Targeted RNA sequencing (Illumina "Manifest")

Cette table sera téléchargeable avec la version en ligne de l'article.

Table S9 (Additional file 21): Antibodies used for immunodetection

<i>Primary antibodies</i>	Host	Antibody # (Provider)	Working dilution
ETV6	mouse	ab54705 (Abcam)	1/1000
GAPDH	goat	sc-31915 (Santa Cruz)	1/1000
Beta-Actin	mouse	sc-81178 (Santa Cruz)	1/1000
PCNA	Rabbit	ab92552 (Abcam)	1/10000
<i>Secondary antibodies (HRP-coupled)</i>			
Anti-mouse	rabbit	sc-358914 (Santa Cruz)	1/5000
Anti-goat	chicken	sc-2961 (Santa Cruz)	1/5000
Anti-rabbit	donkey	sc-2077 (Santa Cruz)	1/5000

Chapitre 5: Discussion

5.1 Récapitulatif et mise en contexte

Grâce à la combinaison de diverses approches, nous avons analysé en profondeur le facteur de transcription *ETV6*. Nous avons ainsi identifié de nouvelles cibles transcriptionnelles d'*ETV6* par l'analyse de transcriptomes de lignées cellulaires et de patients leucémiques. Cette analyse a été complétée par la cartographie des sites de liaison d'*ETV6* suite à des expériences d'immunoprécipitation de la chromatine couplées au séquençage.

Le gène *CLIC5* a particulièrement retenu notre attention. Celui-ci est fortement lié et réprimé par *ETV6*, d'où sa surexpression dans les cellules des patients porteurs de la translocation t(12;21). Nous avons démontré le rôle de *CLIC5* dans la résistance au stress oxydatif via son implication dans la voie d'apoptose lysosomale.

Le séquençage massif des régions liées par *ETV6* a permis de construire son profil de liaison à l'ADN. En plus de confirmer sa liaison directe au niveau des sites EBS, nous avons également observé la capacité d'*ETV6* à agir comme cofacteur. Nos analyses ont permis de mettre en évidence l'impact potentiel de la protéine chimérique *ETV6-RUNX1* sur le profil de liaison à l'ADN d'*ETV6* ainsi que sur sa fonction transcriptionnelle. En combinant ces résultats avec ceux

obtenus dans diverses autres lignées hématopoïétiques, nous avons démontré l'importance du contexte cellulaire sur la régulation médiée par *ETV6*.

Notre intérêt s'est ensuite porté sur les modulateurs de la fonction d'*ETV6*. Pour ce faire, nous avons développé un système cellulaire qui a permis d'effectuer un criblage fonctionnel afin de mettre en évidence les gènes impliqués dans la répression médiée par *ETV6*. Cette stratégie a mené à une première identification non biaisée de facteurs nécessaires à la répression d'un promoteur ciblé par *ETV6*. En prenant avantage de nos nouvelles connaissances sur le réseau transcriptionnel d'*ETV6*, nous avons confirmé le rôle de ces nouveaux facteurs sur la répression de ses cibles endogènes. Certains de ces facteurs ont un rôle dominant sur la fonction d'*ETV6* alors que d'autres agissent seulement dans certains contextes particuliers, suggérant des mécanismes d'action différents. Cette étude pose les premières bases quant à l'identité des gènes impliqués dans la régulation fonctionnelle d'*ETV6* dans un contexte de B-LAL.

Les résultats obtenus ont permis de caractériser le facteur de transcription *ETV6* et de mettre en évidence la complexité de sa régulation dans le cadre des maladies hématologiques comme la leucémie. La mise en place et l'élaboration de méthodologies d'analyses à large échelle font partie intégrante de ma thèse de doctorat et donc plusieurs aspects techniques seront abordés et discutés plus en détail.

5.2 Le choix du modèle cellulaire

Une des forces de cette étude est assurément l'intégration des données obtenues à travers les différentes approches expérimentales utilisées. Ceci est rendu possible par l'utilisation d'un modèle cellulaire unique, la lignée cellulaire Reh, qui possède la translocation $t(12;21)(p13;q22)$ et la perte de l'allèle résiduel d'*ETV6*. Il était impératif que notre modèle soit adapté et reproductible de la maladie dans laquelle le gène *ETV6* est étudié sans quoi les résultats obtenus ne pourraient être transposables aux patients. Bien qu'*ETV6* ait déjà été associé à d'autres cancers [165], sa perte de fonction observée de manière récurrente suivant la translocation $t(12;21)$ suggère un rôle critique d'*ETV6* à cette étape clé du développement de la B-LAL. Ainsi, la caractérisation d'*ETV6* dans la lignée Reh permet d'avoir un portrait de sa fonction au moment précis de la transition de l'état préleucémique à leucémique. De plus, il existe un fort chevauchement entre les gènes réprimés suite à la réexpression d'*ETV6* dans la lignée Reh et ceux surexprimés spécifiquement dans le sous-groupe de patients leucémiques porteurs de la $t(12;21)$ [166]. Ceci suggère que la surexpression de ces gènes observée dans ce sous-type moléculaire est attribuable à l'inactivation d'*ETV6* et supporte l'utilisation de la lignée cellulaire Reh dans l'étude d'*ETV6* dans la B-LAL.

La réexpression d'*ETV6* dans la lignée cellulaire Reh a potentiellement engendré des différences d'expression très marquées dues à l'absence d'*ETV6* endogène. Le fort impact d'*ETV6* dans cette lignée a facilité l'identification de ses gènes cibles avec le potentiel de récupérer des gènes plus faiblement modulés par des

interactions secondaires. Cette absence d'*ETV6* endogène permet également d'évaluer l'impact de formes mutantes d'*ETV6* (p. ex. *ETV6* Δ *ETS*) sans avoir d'interférence ou d'effet compensatoire d'*ETV6*. Une forme mutante sans effet dominant négatif n'aura que très peu d'effet dans un modèle où la forme sauvage est déjà exprimée. De façon similaire, les expériences de ChIP où l'expression d'une forme étiquetée (*ETV6*-HA) est nécessaire à l'immunoprécipitation sont favorisées par l'absence d'*ETV6* qui viendrait compétitionner les mêmes sites de liaison. Finalement, l'absence d'*ETV6* endogène de la lignée Reh a permis de construire de façon séquentielle un système complexe de criblage fonctionnel.

Inversement, l'analyse fonctionnelle des gènes cibles d'*ETV6* dans ce modèle est rendue plus difficile. Les gènes cibles sont déjà surexprimés en raison de la perte d'*ETV6* dans ces cellules et nous empêchent peut-être d'apprécier tout le potentiel que pourraient avoir ces gènes lorsque surexprimés de manière exogène. Certaines cibles d'*ETV6* ne sont pas du tout exprimées dans les précurseurs B normaux (p. ex. *CLIC5*) et donc leur induction suivant l'inactivation d'*ETV6* a potentiellement un effet beaucoup plus important que ce que nous pourrions observer par surexpression. De plus, dans une maladie à l'étiologie aussi complexe que la B-LAL, il est possible que la surexpression d'un gène ait un effet seulement dans un certain contexte cellulaire (niveau de différenciation, microenvironnement de la niche, signalisation extracellulaire) qui ne peut être fidèlement reproduit *in vitro*. Cependant, l'évaluation des impacts phénotypiques des cibles d'*ETV6* identifiées dans le cadre de la B-LAL (cellules Reh) est difficilement justifiable dans un contexte différent où ces gènes pourraient avoir

une tout autre fonction ou encore dépendre de l'expression d'autres gènes. En ce sens, plusieurs gènes cibles d'*ETV6* sont surexprimés conjointement par la perte de son expression et il est possible que la contribution d'un gène à un phénotype donné ne puisse être perçue qu'en présence d'autres gènes surexprimés.

5.3 Régulation directe et indirecte

Les cibles d'*ETV6* n'ont pas toutes un profil aussi clair que celui de *CLIC5*. En considérant les 24 gènes réprimés par l'expression d'*ETV6* et surexprimés dans les patients t(12;21), environ la moitié (13) ont pu être associés à la présence physique d'*ETV6*. En d'autres mots, 11 gènes ne semblent pas avoir *ETV6* à proximité tel qu'illustré par l'absence de signaux ChIP-seq, mais sont néanmoins surexprimés dans les patients t(12;21) et répondent négativement à la présence d'*ETV6*. Comme aucune étude n'a encore pu démontrer l'implication d'*ETV6* dans des changements d'expression à grande distance et qu'il est peu probable qu'*ETV6* eu été présent à proximité de ces gènes sans avoir été récupéré en ChIP-seq, ces résultats démontrent qu'une partie de la signature des patients t(12;21) est attribuable à des effets "secondaires" à la perte d'*ETV6*. En ce sens, *ETV6* modifierait l'expression de ses gènes cibles qui à leur tour auraient un effet sur l'expression d'autres gènes. Bien que ces gènes ne puissent être formellement assignés comme cibles d'*ETV6*, leur dérégulation pourrait néanmoins avoir un rôle dans la pathophysiologie de la B-LAL.

Parmi les 13 gènes reconnus par *ETV6* et régulés par ce dernier, il est étonnant de constater que seulement 5 d'entre eux ont une dépendance de leur régulation au domaine ETS de liaison à l'ADN (*CLIC5*, *ANGPTL2*, *WBP1L*, *SLC51A* et *NDFIP2*). Or, 4 de ces 5 gènes réprimés par *ETV6* de manière ETS-dépendante sont associés à des régions ETS-unique (**Tableau VI**), renforçant fortement la validité de ces résultats. Seule la région liée par *ETV6* à proximité du gène *SLC51A* (*OSTalpha*) ne possède pas de motifs ETS. L'analyse en profondeur de cette région a néanmoins identifié les 3 motifs associés aux facteurs IRF, dont celui en tandem avec un site EBS, suggérant que ce gène soit réprimé par un complexe *ETV6*-IRF qui lierait conjointement l'ADN à cet endroit. Un complexe entre *ETV6* et *IRF8* (*ICSBP*) avait d'ailleurs été identifié [152] et corrobore l'enrichissement de motifs IRF.

La technique de CHIP permet non seulement de récupérer l'ADN lié directement par la protéine d'intérêt mais aussi par un complexe contenant cette même protéine. Or, les 8 gènes régulés de manière ETS-indépendante seraient liés à *ETV6* par l'entremise d'un tiers facteur de transcription qui recruterait *ETV6* pour participer à leur répression. À cet égard, la moitié de ce groupe de gènes cibles (4) contient les motifs RUNX aux régions liées par *ETV6* (**Tableau VI**). L'occurrence de ce motif est fort probablement attribuable à la présence de la protéine de fusion *ETV6*-RUNX1. L'hétérodimérisation entre *ETV6* et *ETV6*-RUNX1 a d'ailleurs été démontrée [98] et corroborée dans la lignée Reh par l'enrichissement de peptides correspondant à RUNX1 dans un immunoprécipité d'*ETV6*-HA. Aussi, certains autres motifs consensus dont ceux des facteurs

IRF1, *E2F6* et *KLF14* ont été identifiés de manière récurrente dans les régions liées et régulées par *ETV6*. Il est important de mentionner que le facteur *E2F6* et des membres de la famille KLF ont en plus été identifiés dans le criblage fonctionnel par shRNA, suggérant que l'activité transcriptionnelle d'*ETV6* peut effectivement dépendre de ces facteurs. L'ensemble de nos résultats illustre bien le rôle d'*ETV6* non seulement comme facteur de transcription mais aussi comme cofacteur. Bien que cette fonction soit atypique des facteurs ETS qui requièrent normalement une liaison directe aux sites EBS [167], *ETV6* est un membre particulier de par son activité répressive et semblerait aussi être recruté de manière différente.

Tableau VI: Sommaire des motifs consensus identifiés au niveau des séquences liées par *ETV6* ayant été associées à des gènes réprimés dans la lignée Reh et à la signature transcriptionnelle des patients t(12;21)(p13;q22) positifs

Catégorie	Gène cible	Motifs RUNX	Motifs ETS	Motifs additionnels				
ETS-dépendant	<i>CLIC5</i>	0	3	PU.I-IRF	KLF14			
	<i>ANGPTL2*</i>	0	1	-				
	<i>WBP1L*</i>	0	1	PRDM1				
	<i>SLC51A</i>	0	0	bZIP:IRF	PU.I-IRF	PRDM1	IRF1	KLF14
	<i>NDFIP2*</i>	0	4	PU.1-IRF				
ETS-indépendant	<i>KLHL30</i>	1	0	TLX	PAX8			
	<i>IRX3*</i>	1	3	ARNT:Ahr	PU.I-IRF	KLF14		
	<i>NTRK1</i>	0	0	ARNT:Ahr	PU.I-IRF			
	<i>C6orf222</i>	0	2	-				
	<i>KCNA2</i>	0	0	KLF14				
	<i>PIK3C3</i>	1	3	PU.I-IRF	KLF14			
	<i>TSPAN9*</i>	1	0	-				
<i>DSC3</i>	0	0	bZIP:IRF					

* Si plusieurs signaux ChIP-seq ont été identifiés pour un même gène, celui ayant l'enrichissement le plus grand a été considéré.

5.4 L'interaction d'*ETV6* avec *ETV6-RUNX1*

Classiquement, la fusion d'*ETV6* avec *RUNX1* modifie la fonction d'activateur transcriptionnel de ce dernier vers un répresseur, dérégulant ainsi l'expression de ses gènes cibles [97]. Nos résultats font état du rôle d'*ETV6* comme cofacteur qui, dans le cadre particulier de la t(12;21), semblerait agir comme partenaire de la protéine chimérique *ETV6-RUNX1*. Plusieurs évidences complémentaires suggèrent qu'*ETV6* puisse effectivement favoriser la répression des gènes ciblés par cette chimère. En plus de leur interaction physique, nous avons observé l'enrichissement de motifs RUNX-unique identifiés au niveau des séquences liées par *ETV6*. Non seulement ces séquences sont potentiellement liées par *ETV6* via la protéine de fusion, mais le recrutement d'*ETV6* à ces loci est également associé à une répression transcriptionnelle. De plus, il existe une récurrence entre les gènes réprimés par *ETV6* et *ETV6-AML1* dans les cellules Reh [168]. Bien que ce chevauchement soit assez restreint (n=10), la répression des gènes déjà ciblés par *ETV6-RUNX1* suivant la surexpression d'*ETV6* suggère un mécanisme coopératif. Nous avons également mis en évidence l'impact d'*ETV6* et de certains de ses modulateurs fonctionnels (*AKIRIN1* et *GNA13*) sur l'expression d'une cible potentielle de *ETV6-RUNX* soit *ITGB2* [168], supportant l'apport d'*ETV6* et de certaines protéines associées à la répression médiée par *ETV6-RUNX1*.

L'interaction *ETV6/ETV6-RUNX1* permettrait également leur recrutement à des sites additionnels. En plus des motifs RUNX-unique identifiés, certains étaient également présents à proximité de motifs ETS (motifs ETS+RUNX), suggérant

une liaison conjuguée à l'ADN. À cet égard, il a été démontré dans un modèle cellulaire murin (Ba/F3) que plusieurs cibles de la fusion ETV6-RUNX1 ne semblaient être régulées ni par *RUNX1* ni par *ETV6*, suggérant des cibles distinctes pour cette chimère [99]. Une partie de ces nouvelles cibles serait donc attribuable à cette interaction avec *ETV6*. Au même titre que les cibles dérégulées de *RUNX1*, ces gènes nouvellement ciblés et réprimés par le complexe *ETV6/ETV6-RUNX1* pourraient contribuer au caractère préleucémique de la t(12;21).

Cependant, il a aussi été démontré que ETV6-RUNX1 empêchait la régulation médiée par la forme sauvage d'*ETV6* et inversement [98]. Bien que nos expériences n'aient pas été construites de manière à pouvoir évaluer spécifiquement les rôles de dominant négatif de ETV6-RUNX1 et d'*ETV6*, nous avons observé qu'un petit nombre de gènes réprimés par ETV6-RUNX1, comme *PPM1F* [168], étaient considérablement surexprimés suite à la réexpression d'*ETV6* dans les Reh, corroborant cette facette de leur interaction. L'expression de ETV6-RUNX1 [169] et l'ablation d'*ETV6* [86, 133, 134] chez la souris engendrent toutefois des phénotypes distincts. Bien que ce mécanisme de dominant négatif puisse effectivement avoir lieu, il ne semble pouvoir expliquer qu'une fraction des changements transcriptionnels et phénotypiques associés à la t(12;21).

5.5 Les cibles d'*ETV6* dans la leucémogénèse

Un objectif important de ma thèse était l'identification de nouvelles cibles transcriptionnelles d'*ETV6* pour pouvoir ensuite les associer à certains phénotypes caractéristiques des cellules tumorales. Bien que nous ayons identifié bon nombre de gènes modulés par *ETV6*, leurs contributions à la leucémogénèse restent difficiles à évaluer. Nous avons concentré nos efforts sur *CLIC5*.

5.5.1. *ETV6*, *CLIC5* et le stress oxydatif dans l'hypothèse virale

Malgré la qualité et la validité des cibles transcriptionnelles d'*ETV6* que nous avons identifiées, il ne semble pas y avoir d'évidence pour l'enrichissement d'une voie de signalisation ou processus biologique (résultats non publiés) [170, 171]. Ce résultat rend encore plus difficile l'évaluation du rôle d'*ETV6* dans la leucémogénèse et suggère que certains gènes, individuellement ou conjointement, et non pas une voie définie, contribuent dans une certaine mesure à la maladie. Nous avons ainsi choisi de concentrer nos efforts sur la cible la plus solide d'*ETV6* soit le gène *CLIC5*. Il était intéressant de constater que *CLIC5* diminuait la sensibilité des cellules au stress oxydatif en réduisant l'impact du peroxyde d'hydrogène sur les lysosomes. Ce nouveau rôle pour *CLIC5* est bien loin de ceux le faisant intervenir dans la fonction normale des podocytes (cellules rénales) [172, 173] et des cellules ciliées de l'oreille interne [174] par son interaction avec des protéines de la famille ERM associées au cytosquelette [175-177]. Il s'agit par ailleurs de sa première fonction biologique reliée à son activité de canal ionique [178, 179]. Bien que classiquement

associée aux tumeurs solides, la résistance au stress oxydatif et à l'hypoxie est maintenant considérée comme un mécanisme déterminant dans les leucémies où les cellules hématopoïétiques sont également soumises à ces stress [180, 181]. De plus, la chimère ETV6-RUNX1 induit, par un mécanisme encore indéterminé, l'augmentation des espèces d'oxygène réactives (*reactive oxygen species; ROS*) [182]. La surexpression de *CLIC5* favoriserait donc les cellules leucémiques présentes dans leur niche.

L'infection virale est une composante environnementale associée à la transition de l'état préleucémique à leucémique [17, 18]. Les processus inflammatoires suivant les infections virales ou bactériennes contribuent de manière importante à l'efficacité de la réponse immunitaire tout en modulant l'hématopoïèse [183-185]. En effet, les cellules de la niche de la moelle osseuse comme les cellules stromales mésenchymateuses possèdent les récepteurs à certaines cytokines et chemokines pro-inflammatoires leur permettant de transduire ce signal au niveau de leur microenvironnement et d'amorcer des changements au niveau de la maturation et de la différenciation des cellules précurseurs hématopoïétiques (**Figure XIII**) [186]. Il est important de mentionner que l'activation des cellules stromales mésenchymateuses par l'IFN γ et/ou le TNF (*tumor necrosis factor*) induit l'enzyme NOS (*nitric oxide synthase*) responsable de la production d'oxyde nitrique impliqué dans cette transduction de signal [187]. Or l'oxyde nitrique, en plus d'être une molécule nécessaire à cette cascade signalétique, est également un ROS. Il a également été démontré que le TNF α induisait la production de ROS par la mitochondrie et initiait l'apoptose [188]. Ainsi, les cytokines pro-

inflammatoires comme l'IFN γ et le TNF produites en réponse à une infection engendrent un environnement chargé en ROS. Ce changement d'environnement favoriserait le développement et la sélection des cellules plus résistantes au stress oxydatif. Les avantages conférés par la perte de fonction d'*ETV6* et la surexpression de *CLIC5* s'inscrivent parfaitement dans ce scénario.

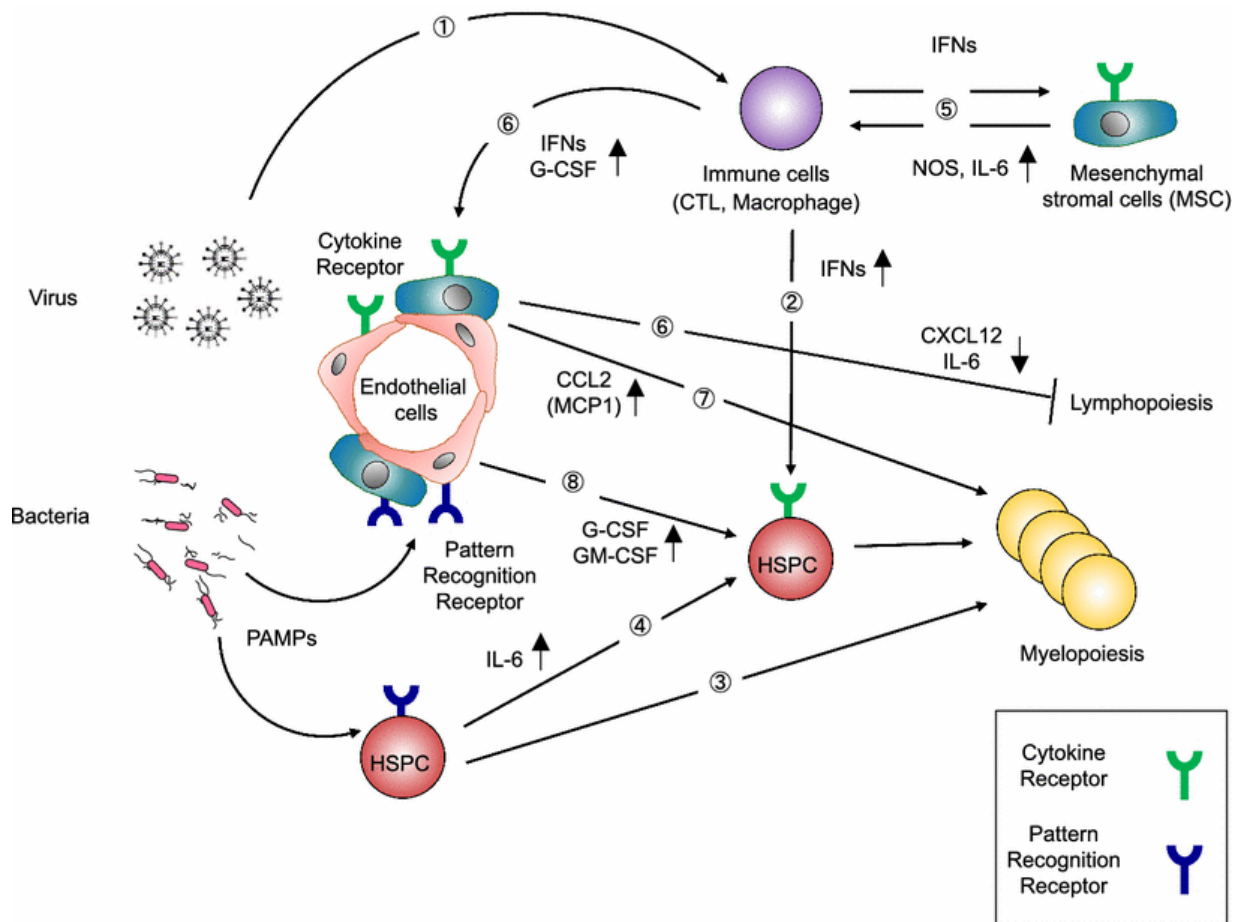


Figure XIII: Hématopoïèse induite par des effets directs et indirects de l'infection. 1. L'infection virale est reconnue par des cellules immunitaires notamment les lymphocytes T cytotoxiques (CTL) et les macrophages qui, après activation, sécrètent des interférons (IFNs) ou le *granulocyte colony-stimulating factor* (G-CSF). 2. Les interférons activent la prolifération des progéniteurs et cellules souches hématopoïétiques (HSPCs) tout en diminuant leur capacité d'autorenouvellement. 3. Les motifs moléculaires associés aux pathogènes (PAMPs) dérivés des infections bactériennes agissent directement sur les récepteurs de reconnaissance de motifs moléculaires (PRRs) exprimés

sur les HSPCs et favorisent leur différenciation en cellules myéloïdes. **4.** Parallèlement, les HSPCs sécrètent des cytokines pro-inflammatoires comme l'interleukine-6 (*IL-6*) qui les activent de manière paracrine et autocrine. **5.** Les interférons produits par les cellules immunitaires suite à l'infection stimulent aussi la production d'*IL-6* et de *nitric-oxide synthase (NOS)* par les cellules stromales mésenchymateuses (MSCs) qui modulent la différenciation et la fonction des cellules immunitaires. **6.** Le *G-CSF* sécrété par les cellules immunitaires diminue quant à lui la production de *CXCL12* et d'*IL-6* par les MSCs, ce qui entraîne une réduction de la différenciation en lymphocytes B. **7.** L'activation des PRRs sur les MSCs conduit à la production de *CCL2* qui promeut le trafic des monocytes. **8.** Finalement, le *G-CSF* ou le *granulocyte macrophage colony-stimulating factor (GM-CSF)* produits par les cellules endothéliales en réponse à l'activation des PRRs favorisent la différenciation myéloïde des HSPCs. Reproduite de H. Takizawa et M.G. Manz, 2017 [186].

En plus de son recrutement par *ICSBP* (ou *IRF8*) pour moduler l'expression des cibles transcriptionnelles dépendantes de L'IFN γ [152], plusieurs gènes liés et régulés par *ETV6* contiennent les motifs de liaison des facteurs IRF. Le site de liaison composite ETS-IRF est d'ailleurs commun dans le profil de liaison d'*ETV6*. Par exemple, la région promotrice de *CLIC5A* liée par *ETV6*, en plus de ses motifs ETS, contient la séquence ETS-IRF (**Tableau VI**). L'enrichissement de ces motifs a également été vu dans les deux autres lignées cellulaires testées (GM12878 et K562). Il est à noter que le motif ETS-IRF est le premier motif "non-ETS" à être enrichi au niveau des GM12878, suggérant ainsi un mécanisme de régulation conservé. Finalement, *CLIC5* semble jouer un rôle dans cette voie en induisant une résistance au ROS, une composante importante de la réponse inflammatoire. Bien que cette résistance puisse être favorable dans plusieurs autres contextes, il est intéressant de constater le rôle potentiel d'*ETV6* et de sa cible *CLIC5* dans la réponse à une infection.

5.5.2. Les cibles d'*ETV6* associées à la signature t(12;21)

En plus de *CLIC5*, 8 autres gènes ont été identifiés comme cibles d'*ETV6* par plusieurs approches complémentaires et font partie de la signature spécifique des patients t(12;21) positifs. La surexpression de ces gènes suivant la perte d'*ETV6* dans ces patients pourrait contribuer à la transformation cellulaire. Même si ces gènes ne semblent pas avoir de rôles connus dans la leucémogénèse, certains sont associés à d'autres cancers ou bien encore à une voie pouvant ultimement y contribuer.

NTRK1 est un récepteur tyrosine kinase au *NGF* (*nerve growth factor*) qui une fois activé, induit la prolifération, la différenciation et la survie des cellules neuronales via les cascades signalétiques PI3K, RAS/MAPK/ERK et PLC γ [189]. Bon nombre d'altérations de *NTRK1* (majoritairement des translocations) ont été reportées dans plusieurs types de cancers, suggérant un rôle important dans le processus oncogénique [189]. Des gains de copies et la surexpression du gène ont d'ailleurs été observés chez des patients atteints de cancer du foie, du poumon et du sein [190]. Dans le cancer du sein, les gains de copies ont été associés à une occurrence de métastases cérébrales [191]. L'activation constitutive de cette kinase par un variant d'épissage [192] ou par une délétion en phase [193] a également été reportée dans un cas de neuroblastome et de leucémie myéloïde aiguë, respectivement. Ces études suggèrent que la forte activité de *NTRK1* contribue à l'oncogénèse. Il a été estimé qu'environ 55% des cas de leucémie (myéloïde et lymphoblastique confondues) avaient une activité accrue des facteurs NTRKs [194]. Combinée à son potentiel oncogénique, la

suractivation généralisée des NTRKs en fait des cibles thérapeutiques très intéressantes. Le développement d'inhibiteurs pour les NTRKs est d'ailleurs initié pour le traitement de la leucémie [195].

PIK3C3 avait été associé à la signature des B-LAL t(12;21) [196] sans pour autant avoir été lié à *ETV6*. Cette kinase phosphoryle le phosphatidylinositol 3 nécessaire au processus d'autophagie [197, 198]. Le rôle de l'autophagie dans le cancer, bien que reconnu, fait encore l'objet de débat. L'autophagie pourrait servir à prévenir l'apparition de cellules cancéreuses ou encore favoriser leur survie en les rendant plus robustes aux changements métaboliques et résistantes à certains signaux de mort cellulaire [199]. La forte expression de *PIK3C3* a été corrélée avec le mauvais pronostic des patients atteints de leucémie lymphoïde chronique [200], suggérant un mécanisme oncogénique dans ce cas. Des inhibiteurs visant à abolir l'activité kinase de *PIK3C3* réduisent l'autophagie [201] et possèdent un fort impact antitumoral sur des cellules leucémiques [202]. La surexpression de *PIK3C3* pourrait donc contribuer aux phénotypes de la B-LAL t(12;21) et éventuellement être ciblée par des agents thérapeutiques.

IRX3 est un facteur de transcription surexprimé dans plusieurs types de cancers comme ceux du sein [203], les cholangiocarcinomes [204], les adénomes colorectaux [205] et les carcinomes hépatocellulaires [206]. Bien que la surexpression de *IRX3* n'avait toujours pas été observée dans la leucémie, son impact dans les lymphocytes a été associé à une résistance à la lyse oncolytique virale induite par le Pexa-Vec [207], une nouvelle classe d'agents

immunothérapeutiques [208]. Les mécanismes moléculaires sous-jacents à son action demeurent cependant méconnus. En plus d'avoir une incidence sur un traitement potentiel, cet effet de *IRX3* rejoint le rôle d'*ETV6* et de *CLIC5* dans la réponse à l'infection (section précédente 5.5.1). Tout comme *CLIC5* et *IRX3*, la cible ***TSPAN9*** a également été associée à l'infection virale en rendant plus permissive la fusion des alphavirus avec les membranes des endosomes précoces sur lesquelles *TSPAN9* est localisée [209, 210].

DSC3 est la cible d'*ETV6* la plus fortement associée à la signature t(12;21). Ce gène encode une protéine de la famille des desmocollines impliquée dans l'adhésion cellulaire [211]. Dans le cancer de l'ovaire, la surexpression de ce gène a été décrite et promeut la prolifération des cellules cancéreuses induite par la FSH (*Follicle-stimulating hormone*) via l'activation de la voie de signalisation *EGFR/AKT* [212]. Les niveaux d'expression de *DSC3* servent à la classification des cancers du poumon [213, 214] et la surexpression des gènes appartenant à cette famille a été liée à une mortalité précoce chez les patients atteints de mélanomes et de cancers ovariens [215]. Sa fonction demeure cependant méconnue dans le processus leucémogénique.

Même si ***SLC51A*** (*OSTalpha*) semble être une cible très finement régulée par l'activité d'*ETV6* et de ses modulateurs, la fonction de ce transporteur transmembranaire de molécules stéroïdiennes [216] dans le développement du cancer est toujours inconnue. Comme ce qui a été fait pour *CLIC5*, des expériences visant à identifier un rôle pro-oncogénique pour *SLC51A* devraient être menées.

KCNA2 fait parti de la famille des canaux potassiques voltage-dépendant impliqués dans une multitude de processus majoritairement au niveau des cellules nerveuses du cerveau et du système nerveux central [217]. Des altérations dans ce gène ont été associées à certaines neuropathies notamment l'épilepsie [217]. Son impact oncogénique n'a toujours pas été élucidé. Finalement, aucune étude ne fait état du gène **C6orf222**.

5.6 La régulation de la fonction d'*ETV6*

L'identification à large échelle des modulateurs d'*ETV6* est une étape cruciale pour comprendre les mécanismes régissant son activité transcriptionnelle. L'élaboration de son réseau de régulation en amont permettra de mieux comprendre l'impact que pourraient avoir certains gènes dans les maladies hématologiques via la régulation d'*ETV6*. Jusqu'à présent, aucune étude de ce type n'avait été faite sur *ETV6*.

5.6.1. Le criblage fonctionnel pangénomique par ARN interférant

La méthodologie de criblage par interférence de petits ARN en épingle à cheveux (shRNAs) est utilisée pour identifier des gènes ayant une fonction dans une voie précise [218-221] ou bien la cible de nouvelles petites molécules chimiques (agent chimiothérapeutique, médicaments, etc.) [222, 223]. Le criblage dit "positif" est conçu de manière à ce que seules les cellules ayant intégré un shRNA spécifique pour un gène particulier puissent être sélectionnées. Comme nous avons maintenant accès à des couches

d'informations plus subtiles grâce à un séquençage en profondeur, il est possible d'avoir un portrait global plus complet et complexe d'une voie moléculaire. Ici, le criblage visait l'identification des gènes impliqués dans l'activité répressive d'*ETV6* sur un promoteur artificiel reconnu par ce dernier (EBS₃tk).

En plus des limitations classiques inhérentes à l'utilisation d'ARN interférant (shRNAs létaux, trop peu représentés dans les *pools* ou tout simplement non efficaces), la nature de ce promoteur, tout comme son site d'intégration, a un impact important sur les résultats obtenus. Les analyses de données d'expression, de ChIP-seq et de spectrométrie de masse ont toutes révélé l'étroite relation entre *ETV6* et la protéine de fusion ETV6-RUNX1 (voir section 5.4). La répression de certains gènes par *ETV6* semble dépendante de ETV6-RUNX1. Pourtant, aucun shRNA contre *RUNX1* (qui aurait pu inhiber ETV6-RUNX1) n'a été retrouvé. Comme le promoteur artificiel est directement ciblé par *ETV6* (aucun site RUNX), l'inhibition de ETV6-RUNX1 n'a aucun impact sur l'expression du gène de résistance (BlastR). Nous avons donc possiblement omis certains modulateurs de la fonction d'*ETV6* inutiles dans ce contexte précis (faux négatifs).

Le criblage primaire a pu identifier 1156 gènes ayant une implication potentielle dans la répression médiée par *ETV6*. Afin de minimiser les effets *off-targets* des ARN interférant (faux positifs), nous avons sélectionné 81 gènes couverts par plusieurs shRNAs [224]. Comme les effets *off-targets* sont aléatoires, la chance que plusieurs shRNAs pour un même gène engendrent le phénotype recherché (ici une résistance à la Blastidicin) via d'autres gènes est beaucoup moins

probable. Comme pour les cibles d'*ETV6*, il ne semble pas y avoir d'enrichissement pour une fonction ou voie de signalisation particulière au sein des modulateurs potentiels (résultats non publiés) [170, 171]. Bien que le taux de faux positifs soit vraisemblablement plus élevé pour les gènes couverts par un seul shRNA, il n'en demeure pas moins que certains d'entre eux doivent être de vrais modulateurs. Neuf de ces gènes ont été inclus sur une base prédictive d'interaction [225], mais les validations révéleront cependant qu'aucun d'eux ne semble moduler la fonction d'*ETV6*. Il aurait été éventuellement préférable d'inclure ceux ayant une fonction connue sur la transcription. À titre d'exemples, les histones déacétylases *HDAC7*, *SIRT3* et *SIRT4* ont été récupérées et pourraient être impliquées dans la répression médiée par *ETV6* [226]. Les facteurs *E2F6* et KLFs associés aux motifs récurrents identifiés dans le profil de liaison à l'ADN d'*ETV6* et présents dans le criblage auraient également pu être inclus. La fonction d'*E2F6* comme répresseur transcriptionnel [227, 228] associé aux protéines du groupe *polycomb* [229] fait de lui un modulateur potentiel intéressant. De plus, la présence conjointe d'un site de liaison à *E2F6* et au facteur ETS GABP a été associée à une répression du promoteur de *BRCA1* [230], suggérant un mécanisme similaire avec *ETV6*.

5.6.2. Validation et complexité de la régulation médiée par d'*ETV6*

L'impact des gènes identifiés par le criblage (n= 90) a été testé individuellement sur l'ensemble du réseau de régulation d'*ETV6* préalablement élaboré à partir des études d'expression (Chapitre 2: Article 1) et de CHIP-seq (Chapitre 3: Article

2). En plus de valider le rôle de plusieurs gènes comme modulateurs d'*ETV6*, l'inclusion de plusieurs cibles a permis de mettre en évidence toute la complexité de la régulation médiée par *ETV6*. En effet, hormis quelques modulateurs clés notamment *AKIRIN1*, *DYRK4* et *SRP72*, la majorité semblait avoir un effet marqué mais restreint à un petit nombre de gènes cibles, suggérant qu'*ETV6* fait appel à plusieurs mécanismes de régulation différents selon le contexte. Il est possible que l'intégration de signaux (p. ex. les cibles transcriptionnelles devant être modulées en réponse à un stress cellulaire précis) ou le contexte chromatinien des promoteurs ciblés soient des facteurs importants permettant de définir le choix de ces mécanismes. Il est intéressant de constater que l'expression des gènes cibles d'*ETV6* *SEMA3A* et *SEMA3D* était dépendante des mêmes modulateurs, suggérant une régulation par les mêmes mécanismes pour permettre leur co-expression. Les cibles étant régulées de manière similaire par les modulateurs d'*ETV6* (p. ex. *SLC51A* et *KIAA0226L*; *CLIC5* et *RAP1GAP*) pourraient donc intervenir dans une même voie biologique.

Étonnamment, de tous les gènes ayant été identifiés comme modulateurs de la fonction d'*ETV6*, seul *CBX3* avait été démontré comme interagissant physiquement avec *ETV6*. Comme 3 shRNAs différents pour *CBX3* avaient initialement été identifiés dans le criblage et que 8 des 22 cibles d'*ETV6* étaient significativement surexprimées suite à son inhibition, il est fort probable qu'*ETV6* puisse lier l'ADN en complexe avec *CBX3*. Ce dernier participe à la compaction de la chromatine et l'inhibition de la transcription [231]. Ce genre de mécanisme ne semble pas s'appliquer aux autres modulateurs qui *a priori* ne seraient pas

impliqués dans la régulation médiée par *ETV6* via la formation d'un complexe répresseur. Il est à noter que des complexes formés exclusivement sur l'ADN par la liaison d'*ETV6* suivie du recrutement subséquent d'autres protéines sont difficilement identifiables en spectrométrie de masse car leur présence est plus rare et sporadique et qu'il est difficile d'en faire une immunoprécipitation efficace [232, 233]. Des approches alternatives utilisant la fixation (p. ex. ChIP-MS et RIME) existent pour identifier ce genre d'interactions [234-236] mais n'ont pas été faites dans le cadre de ma thèse. Hormis la formation de complexes répresseurs, certains modulateurs seraient impliqués dans la mise en place d'une structure chromatinienne nécessaire à la liaison d'*ETV6*, par exemple *AKIRIN1* [237, 238]. *ETV6* ne dépendrait donc pas de ces modulateurs pour son activité répressive mais plutôt pour sa liaison. Aussi, la sumoylation [126, 161-163] et la phosphorylation [125, 142, 164] d'*ETV6* modulent de manière importante sa fonction transcriptionnelle (section 1.4.7.2). Certaines kinases comme *DYRK4* ont un effet drastique sur la fonction d'*ETV6*. Comme l'interaction entre une kinase et son substrat (p. ex. *ETV6*) peut être courte et sporadique [239], il est probable que l'analyse par spectrométrie de masse n'ait pu identifier ce type de partenaire. La distance (partenaires directs ou indirects) qui sépare les protéines d'un complexe de celle ciblée par l'immunoprécipitation modifie également les résultats de spectrométrie de masse. À titre d'exemple, *SRP72* fait partie des modulateurs d'*ETV6* extensivement validés sans toutefois avoir été identifié en complexe avec *ETV6*. *SRP9*, un membre de cette même famille a quant à lui été identifié par spectrométrie de masse. Il est probable qu'il existe un

complexe entre *ETV6*, *SRP9* (partenaire direct d'*ETV6*) et *SRP72* (partenaire indirect d'*ETV6* via *SRP9*) impliqué dans la régulation transcriptionnelle. L'intégration de ces approches complémentaires permet de mettre en évidence l'éventail de paramètres pouvant moduler l'activité d'*ETV6*.

Pour faire ressortir les modulateurs ayant le meilleur potentiel, les filtres instaurés dans la validation ont été très rigoureux et visaient à cibler les gènes ayant un effet fort, élargi et spécifique à *ETV6*. L'aspect "spécificité" a été inclus en évaluant l'effet de ces modulateurs sur des gènes non corrélés à l'expression d'*ETV6*. Or, il est possible qu'un vrai modulateur d'*ETV6* soit en plus impliqué dans la régulation transcriptionnelle indépendamment d'*ETV6*. Ceci pourrait être le cas, par exemple, de facteurs communs à la répression transcriptionnelle comme *CBX3* (ou *HP1 γ* ; *heterochromatin protein 1*) [231] qui auraient pu être écartés à tort par cette approche. *MCM3AP*, à l'inverse, aurait potentiellement un rôle général sur la transcription sans être associé à *ETV6*. En effet, ce gène avait un effet fort sur bon nombre de cibles d'*ETV6* (13/22), mais également sur plusieurs gènes contrôles (non ciblés par *ETV6*) de manière à ce que la comparaison ne soit pas statistiquement significative. Or, ce gène (ou plus probablement le gène *GANP* qui contient la séquence complète de *MCM3AP* [240]) aurait une implication dans la disposition de la chromatine à l'intérieur du noyau via le complexe TREX-2 [241, 242]. La disposition de la chromatine a un effet très important sur la transcription et donc les effets observés pourraient y être attribuée sans qu'*ETV6* soit impliqué. Ce genre de gènes aurait facilement pu se retrouver dans le criblage initial en favorisant l'expression du gène BlastR

sans faire intervenir *ETV6*. Ce cas de figure pouvait difficilement être évité, mais les gènes pouvant être associés à ce mécanisme ont potentiellement été exclus par ces filtres. Il est à noter que l'inhibition individuelle des 5 modulateurs les plus forts (*AKIRIN1*, *COMMD9*, *DYRK4*, *JUNB* et *SRP72*) avait un effet fortement dépendant de la présence d'*ETV6*, corroborant l'efficacité de notre approche pour retirer les gènes agissant de manière indépendante.

Certains shRNAs retrouvés dans le criblage avaient un effet notable et significatif sur le réseau transcriptionnel d'*ETV6* sans toutefois diminuer l'expression du gène ciblé. Les gènes *L3MBTL3*, *DAPK2*, *CRK* et *ZRANB1* en sont de bons exemples. Cet effet serait attribuable à un gène inconnu (*off-target*) ou à des amplicons inadéquats pour évaluer leur expression. Parmi ces quatre gènes, *L3MBTL3* avait été ciblé par un seul shRNA et inclus dans la validation sur la base prédictive de son interaction avec d'autres modulateurs. Comme une interaction avait en plus été observée entre *ETV6* et un autre membre de la famille de corépresseur *polycomb* (*L3MBTL1*) [150], il aurait été intéressant de s'y attarder plus en profondeur car il y aurait potentiellement un complexe faisant intervenir ces protéines.

5.6.3. Pathologies associées à *ETV6* et ses modulateurs

ETV6 est connu comme étant impliqué dans plusieurs réarrangements associés à des désordres hématologiques (majoritairement des leucémies). Depuis peu, des mutations germinales au niveau d'*ETV6* ont également été associées à un historique familial de maladies hématologiques [25-28], suggérant que la perte de fonction d'*ETV6* puisse causer ces maladies. Il est donc possible que

l'altération des modulateurs d'*ETV6* puisse également y participer. Or, la majorité des modulateurs d'*ETV6* identifiés et validés ne semblent être ni mutés ni réarrangés dans ces maladies. L'inhibition partielle et spécifique de la fonction d'*ETV6* par l'altération de ses modulateurs ne serait pas suffisante pour initier ces maladies. Il est important de mentionner que les désordres hématologiques comme les leucémies sont pour la plupart des maladies complexes nécessitant plusieurs éléments combinés avant de se manifester. Il est possible que plusieurs gènes cibles d'*ETV6* (ou une combinaison précise) doivent être dérégulés et souvent même complétés par d'autres altérations [103] pour enclencher le processus oncogénique. Les modulateurs impliqués dans la régulation de seulement une fraction des cibles d'*ETV6* auraient donc moins de chance de contribuer à la transformation cellulaire advenant une perte de fonction. Malgré tout, il demeure envisageable qu'un modulateur d'*ETV6* ayant un effet restreint puisse néanmoins affecter un gène cible important responsable d'un phénotype délétère.

Il est probable que seuls les modulateurs ayant un impact généralisé sur *ETV6* et sur son réseau de régulation puissent être associés à certaines maladies hématologiques. À cet égard, il a été démontré que des mutations dans *SRP72* étaient liées avec des cas d'aplasies et de syndromes myélodysplasiques familiaux [243]. Ces pathologies sont également associées à des mutations germinales d'*ETV6* [26]. La fonction canonique de la protéine *SRP72* dans la translocation des peptides naissants vers le réticulum endoplasmique explique difficilement son implication dans cette maladie. À la lumière de nos résultats, il

est possible que la perte de fonction de *SRP72* observée dans ces cas perturbe le réseau transcriptionnel d'*ETV6* de manière à favoriser l'apparition de la maladie.

Hormis les cas de fibrosarcomes congénitaux [244] et de carcinomes sécrétoires du sein et des glandes salivaires [165] où la translocation t(12;15) (fusion *ETV6-NTRK3*) est récurrente, l'implication d'*ETV6* et de ses modulateurs dans d'autres cancers (autres que les leucémies) demeure cependant très peu connue. Néanmoins, l'analyse de données d'expression provenant de différents tissus tumoraux a pu révéler que les gènes corrélant avec l'expression d'*ETV6* étaient également corrélés avec l'expression de certains modulateurs, suggérant que les modulateurs identifiés dans le contexte de la leucémie sont également valables dans d'autres conditions. Ces associations entre l'expression des modulateurs et ces gènes laissent sous-entendre que certains d'entre eux seraient régulés par *ETV6* en empruntant des mécanismes similaires à ceux employés dans les cellules pré-B leucémiques.

5.7 Conclusions générales et perspectives

À l'aide d'une combinaison d'approches à large échelle, nous avons élaboré le réseau de régulation d'*ETV6*, du moins partiellement, dans les cellules de leucémie lymphoblastique aiguë. Nous avons fait une première identification des gènes cibles d'*ETV6*, de son profil de liaison et de ses modulateurs. Bien que ces travaux soient d'une nature plutôt fondamentale, certains éléments

pourraient ultimement être transposés au milieu clinique. Étant donné l'importance de la classification des patients leucémiques pour leur prise en charge, l'intégration des niveaux d'expression de *CLIC5* pourrait compléter les informations obtenues des caryotypes et des hybridations *in situ* pour identifier les patients porteurs de la translocation t(12;21)(p13;q22).

En plus de son utilisation potentielle comme biomarqueur, l'impact de *CLIC5* dans la résistance au stress oxydatif le relie à l'hypothèse virale associée à la B-LAL. Bien que le dérèglement d'*ETV6* et de ses cibles transcriptionnelles puisse avoir une incidence sur la leucémogénèse par le biais de plusieurs voies moléculaires, il serait intéressant de s'attarder à leurs implications en lien avec l'infection. En ce sens, l'identification des impacts fonctionnels attribuables à la dérégulation des cibles d'*ETV6* pourrait aboutir à des bénéfices pour le patient. Des gènes comme *CLIC5* qui sont drastiquement surexprimés dans les cellules des patients t(12;21)(p13;q22) et qui, en plus, peuvent être facilement inhibés, sont d'excellentes cibles thérapeutiques potentielles.

Finalement, la caractérisation en profondeur des génomes et des profils d'expression a permis de mettre en évidence de nouvelles relations entre les gènes et les maladies. Les mécanismes pathophysiologiques sous-jacents sont cependant plus complexes à élaborer et nécessitent des études fonctionnelles plus approfondies. Suite à l'identification de *SRP72* comme modulateur d'*ETV6*, il serait intéressant d'éclaircir l'impact de son nouveau rôle dans les cas de syndromes myélodysplasiques associés à ses mutations. Ces nouveaux

éléments pourraient contribuer à la prise en charge personnalisée de ces patients.

Beaucoup de travail reste encore à faire pour élucider les implications pathophysiologiques liées à la dérégulation d'*ETV6* et de son réseau. Les travaux effectués dans le cadre de ce projet serviront de point de départ pour de futures études visant à mieux comprendre la fonction moléculaire d'*ETV6* et son implication dans les pathologies humaines.

Références

1. Coldman A, Phillips N, Warren L, Kan L. Breast cancer mortality after screening mammography in British Columbia women. *Int J Cancer*. 2007;120:1076-1080
2. Kurzrock R, Gutterman JU, Talpaz M. The molecular genetics of Philadelphia chromosome-positive leukemias. *N Engl J Med*. 1988;319:990-998
3. Childhood Acute Myeloid Leukemia/Other Myeloid Malignancies Treatment (PDQ(R)): Health Professional Version. In PDQ Cancer Information Summaries. Bethesda (MD); 2002
4. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatr Clin North Am*. 2015;62:47-60
5. Robb L. Cytokine receptors and hematopoietic differentiation. *Oncogene*. 2007;26:6715-6723
6. Zhu J, Emerson SG. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene*. 2002;21:3295-3313
7. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell Stem Cell*. 2012;10:120-136
8. Saussele S, Silver RT. Management of chronic myeloid leukemia in blast crisis. *Ann Hematol*. 2015;94 Suppl 2:S159-165
9. Tasian SK, Loh ML, Hunger SP. Childhood acute lymphoblastic leukemia: Integrating genomics into therapy. *Cancer*. 2015;121:3577-3590
10. Mandel EM, Grosschedl R. Transcription control of early B cell differentiation. *Curr Opin Immunol*. 2010;22:161-167
11. Bryder D, Sigvardsson M. Shaping up a lineage--lessons from B lymphopoiesis. *Curr Opin Immunol*. 2010;22:148-153
12. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446:758-764
13. Hu Y, Yoshida T, Georgopoulos K. Transcriptional circuits in B cell transformation. *Curr Opin Hematol*. 2017;24:345-352

14. Belver L, Ferrando A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer*. 2016;16:494-507
15. Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell*. 2012;22:209-221
16. Belson M, Kingsley B, Holmes A. Risk factors for acute leukemia in children: a review. *Environ Health Perspect*. 2007;115:138-145
17. Schuz J, Erdmann F. Environmental Exposure and Risk of Childhood Leukemia: An Overview. *Arch Med Res*. 2016;47:607-614
18. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer*. 2006;6:193-203
19. Alter BP. Cancer in Fanconi anemia, 1927-2001. *Cancer*. 2003;97:425-440
20. Hasle H. Pattern of malignant disorders in individuals with Down's syndrome. *Lancet Oncol*. 2001;2:429-436
21. Whitlock JA. Down syndrome and acute lymphoblastic leukaemia. *Br J Haematol*. 2006;135:595-602
22. Moriyama T, Relling MV, Yang JJ. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood*. 2015;125:3988-3995
23. Shah S, Schrader KA, Waanders E, Timms AE, Vijai J, Miething C, et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat Genet*. 2013;45:1226-1231
24. Auer F, Ruschendorf F, Gombert M, Husemann P, Ginzel S, Izraeli S, et al. Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. *Leukemia*. 2014;28:1136-1138
25. Zhang MY, Churpek JE, Keel SB, Walsh T, Lee MK, Loeb KR, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet*. 2015;47:180-185
26. Topka S, Vijai J, Walsh MF, Jacobs L, Maria A, Villano D, et al. Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet*. 2015;11:e1005262

27. Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet.* 2015;47:535-538
28. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M, et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol.* 2015;
29. Colby-Graham MF, Chordas C. The childhood leukemias. *J Pediatr Nurs.* 2003;18:87-95
30. Alexander S. Clinically defining and managing high-risk pediatric patients with acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program.* 2014;2014:181-189
31. Rabin KR, Poplack DG. Management strategies in acute lymphoblastic leukemia. *Oncology (Williston Park).* 2011;25:328-335
32. Schultz KR, Pullen DJ, Sather HN, Shuster JJ, Devidas M, Borowitz MJ, et al. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood.* 2007;109:926-935
33. Hunger SP, Mullighan CG. Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood.* 2015;125:3977-3987
34. Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. *N Engl J Med.* 2015;373:1541-1552
35. Moricke A, Zimmermann M, Reiter A, Henze G, Schrauder A, Gadner H, et al. Long-term results of five consecutive trials in childhood acute lymphoblastic leukemia performed by the ALL-BFM study group from 1981 to 2000. *Leukemia.* 2010;24:265-284
36. Pui CH, Pei D, Sandlund JT, Ribeiro RC, Rubnitz JE, Raimondi SC, et al. Long-term results of St Jude Total Therapy Studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia. *Leukemia.* 2010;24:371-382

37. Silverman LB, Stevenson KE, O'Brien JE, Asselin BL, Barr RD, Clavell L, 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;24:320-334
38. Schrappe M, Hunger SP, Pui CH, Saha V, Gaynon PS, Baruchel A, et al. Outcomes after induction failure in childhood acute lymphoblastic leukemia. *N Engl J Med*. 2012;366:1371-1381
39. Winters AC, Bernt KM. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr*. 2017;5:4
40. Meyer C, Hofmann J, Burmeister T, Groger D, Park TS, Emerenciano M, et al. The MLL recombinome of acute leukemias in 2013. *Leukemia*. 2013;27:2165-2176
41. Ford AM, Ridge SA, Cabrera ME, Mahmoud H, Steel CM, Chan LC, et al. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature*. 1993;363:358-360
42. Gill Super HJ, Rothberg PG, Kobayashi H, Freeman AI, Diaz MO, Rowley JD. Clonal, nonconstitutional rearrangements of the MLL gene in infant twins with acute lymphoblastic leukemia: in utero chromosome rearrangement of 11q23. *Blood*. 1994;83:641-644
43. Gale KB, Ford AM, Repp R, Borkhardt A, Keller C, Eden OB, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A*. 1997;94:13950-13954
44. Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood*. 1999;94:1057-1062
45. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A*. 2002;99:8242-8247
46. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet Cytogenet*. 2001;130:93-104

47. Bernardin F, Yang Y, Cleaves R, Zahurak M, Cheng L, Civin CI, et al. TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. *Cancer Res.* 2002;62:3904-3908
48. van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J, et al. Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood.* 2011;118:1041-1051
49. Greaves M. Childhood leukaemia. *BMJ.* 2002;324:283-287
50. Gruhn B, Taub JW, Ge Y, Beck JF, Zell R, Hafer R, et al. Prenatal origin of childhood acute lymphoblastic leukemia, association with birth weight and hyperdiploidy. *Leukemia.* 2008;22:1692-1697
51. Paulsson K, Johansson B. High hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer.* 2009;48:637-660
52. Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood.* 2003;102:2951-2959
53. Andersson A, Olofsson T, Lindgren D, Nilsson B, Ritz C, Eden P, et al. Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. *Proc Natl Acad Sci U S A.* 2005;102:19069-19074
54. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell.* 2002;1:133-143
55. Safavi S, Paulsson K. Near-haploid and low-hypodiploid acute lymphoblastic leukemia: two distinct subtypes with consistently poor prognosis. *Blood.* 2017;129:420-423
56. Muhlbacher V, Zenger M, Schnittger S, Weissmann S, Kunze F, Kohlmann A, et al. Acute lymphoblastic leukemia with low hypodiploid/near triploid karyotype is a specific clinical entity and exhibits a very high TP53 mutation frequency of 93%. *Genes Chromosomes Cancer.* 2014;53:524-536

57. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet.* 2013;45:242-252
58. Begley CG, Aplan PD, Davey MP, Nakahara K, Tchorz K, Kurtzberg J, et al. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A.* 1989;86:2031-2035
59. Chen Q, Yang CY, Tsan JT, Xia Y, Ragab AH, Peiper SC, et al. Coding sequences of the tal-1 gene are disrupted by chromosome translocation in human T cell leukemia. *J Exp Med.* 1990;172:1403-1408
60. Ziegler SF, Liu YJ. Thymic stromal lymphopoietin in normal and pathogenic T cell development and function. *Nat Immunol.* 2006;7:709-714
61. Russell LJ, Capasso M, Vater I, Akasaka T, Bernard OA, Calasanz MJ, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood.* 2009;114:2688-2698
62. Dyer MJ, Akasaka T, Capasso M, Dusanj P, Lee YF, Karran EL, et al. Immunoglobulin heavy chain locus chromosomal translocations in B-cell precursor acute lymphoblastic leukemia: rare clinical curios or potent genetic drivers? *Blood.* 2010;115:1490-1499
63. Nelson KN, Peiris MN, Meyer AN, Siari A, Donoghue DJ. Receptor Tyrosine Kinases: Translocation Partners in Hematopoietic Disorders. *Trends Mol Med.* 2017;23:59-79
64. Score J, Calasanz MJ, Ottman O, Pane F, Yeh RF, Sobrinho-Simoes MA, et al. Analysis of genomic breakpoints in p190 and p210 BCR-ABL indicate distinct mechanisms of formation. *Leukemia.* 2010;24:1742-1750
65. Melo JV. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood.* 1996;88:2375-2384
66. Zhang X, Subrahmanyam R, Wong R, Gross AW, Ren R. The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl. *Mol Cell Biol.* 2001;21:840-853

67. Beissert T, Hundertmark A, Kaburova V, Travaglini L, Mian AA, Nervi C, et al. Targeting of the N-terminal coiled coil oligomerization interface by a helix-2 peptide inhibits unmutated and imatinib-resistant BCR/ABL. *Int J Cancer*. 2008;122:2744-2752
68. Kang ZJ, Liu YF, Xu LZ, Long ZJ, Huang D, Yang Y, et al. The Philadelphia chromosome in leukemogenesis. *Chin J Cancer*. 2016;35:48
69. Biondi A, Schrappe M, De Lorenzo P, Castor A, Lucchini G, Gandemer V, et al. Imatinib after induction for treatment of children and adolescents with Philadelphia-chromosome-positive acute lymphoblastic leukaemia (EsPhALL): a randomised, open-label, intergroup study. *Lancet Oncol*. 2012;13:936-945
70. Bleckmann K, Schrappe M. Advances in therapy for Philadelphia-positive acute lymphoblastic leukaemia of childhood and adolescence. *Br J Haematol*. 2016;172:855-869
71. Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JG, Peters ST, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol*. 2009;10:125-134
72. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell*. 2012;22:153-166
73. Hunger SP, Galili N, Carroll AJ, Crist WM, Link MP, Cleary ML. The t(1;19)(q23;p13) results in consistent fusion of E2A and PBX1 coding sequences in acute lymphoblastic leukemias. *Blood*. 1991;77:687-693
74. Lu Q, Kamps MP. Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specificities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. *Oncogene*. 1997;14:75-83
75. McWhirter JR, Neuteboom ST, Wancewicz EV, Monia BP, Downing JR, Murre C. Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia. *Proc Natl Acad Sci U S A*. 1999;96:11464-11469
76. Aspland SE, Bendall HH, Murre C. The role of E2A-PBX1 in leukemogenesis. *Oncogene*. 2001;20:5708-5717

77. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol.* 2001;21:2249-2258
78. Mishra BP, Zaffuto KM, Artinger EL, Org T, Mikkola HK, Cheng C, et al. The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. *Cell Rep.* 2014;7:1239-1247
79. Bitoun E, Oliver PL, Davies KE. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum Mol Genet.* 2007;16:92-106
80. Okuda H, Kanai A, Ito S, Matsui H, Yokoyama A. AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-dependent transcription. *Nat Commun.* 2015;6:8869
81. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet.* 2002;30:41-47
82. Sanjuan-Pla A, Bueno C, Prieto C, Acha P, Stam RW, Marschalek R, et al. Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. *Blood.* 2015;126:2676-2685
83. Bardini M, Spinelli R, Bungaro S, Mangano E, Corral L, Cifola I, et al. DNA copy-number abnormalities do not occur in infant ALL with t(4;11)/MLL-AF4. *Leukemia.* 2010;24:169-176
84. Romana SP, Mauchauffe M, Le Coniat M, Chumakov I, Le Paslier D, Berger R, et al. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood.* 1995;85:3662-3670
85. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell.* 1996;84:321-330
86. Wang LC, Swat W, Fujiwara Y, Davidson L, Visvader J, Kuo F, et al. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev.* 1998;12:2392-2402
87. de Bruijn M, Dzierzak E. Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood.* 2017;129:2061-2069

88. Cai Z, de Bruijn M, Ma X, Dortland B, Luteijn T, Downing RJ, et al. Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*. 2000;13:423-431
89. Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*. 2009;457:887-891
90. Lancrin C, Mazan M, Stefanska M, Patel R, Lichtinger M, Costa G, et al. GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*. 2012;120:314-322
91. Sun W, Downing JR. Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors. *Blood*. 2004;104:3565-3572
92. Ng KP, Hu Z, Ebrahim Q, Negrotto S, Lausen J, Sauntharajah Y. Runx1 deficiency permits granulocyte lineage commitment but impairs subsequent maturation. *Oncogenesis*. 2013;2:e78
93. Egawa T, Tillman RE, Naoe Y, Taniuchi I, Littman DR. The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. *J Exp Med*. 2007;204:1945-1957
94. Lichtinger M, Ingram R, Hannah R, Muller D, Clarke D, Assi SA, et al. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J*. 2012;31:4318-4333
95. Lichtinger M, Hoogenkamp M, Krysinska H, Ingram R, Bonifer C. Chromatin regulation by RUNX1. *Blood Cells Mol Dis*. 2010;44:287-290
96. Zelent A, Greaves M, Enver T. Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. *Oncogene*. 2004;23:4275-4283
97. Hiebert SW, Sun W, Davis JN, Golub T, Shurtleff S, Buijs A, et al. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol*. 1996;16:1349-1355
98. Gunji H, Waga K, Nakamura F, Maki K, Sasaki K, Nakamura Y, et al. TEL/AML1 shows dominant-negative effects over TEL as well as AML1. *Biochem Biophys Res Commun*. 2004;322:623-630

99. Linka Y, Ginzel S, Kruger M, Novosel A, Gombert M, Kremmer E, et al. The impact of TEL-AML1 (ETV6-RUNX1) expression in precursor B cells and implications for leukaemia using three different genome-wide screening methods. *Blood Cancer J.* 2013;3:e151
100. Agape P, Gerard B, Cave H, Devaux I, Vilmer E, Lecomte MC, et al. Analysis of ETV6 and ETV6-AML1 proteins in acute lymphoblastic leukaemia. *Br J Haematol.* 1997;98:234-239
101. Poirel H, Lacronique V, Mauchauffe M, Le Coniat M, Raffoux E, Daniel MT, et al. Analysis of TEL proteins in human leukemias. *Oncogene.* 1998;16:2895-2903
102. Morrow M, Horton S, Kioussis D, Brady HJ, Williams O. TEL-AML1 promotes development of specific hematopoietic lineages consistent with preleukemic activity. *Blood.* 2004;103:3890-3896
103. Sun C, Chang L, Zhu X. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget.* 2017;8:35445-35459
104. Mangolini M, de Boer J, Walf-Vorderwulbecke V, Pieters R, den Boer ML, Williams O. STAT3 mediates oncogenic addiction to TEL-AML1 in t(12;21) acute lymphoblastic leukemia. *Blood.* 2013;122:542-549
105. Raptis L, Arulanandam R, Geletu M, Turkson J. The R(h)oads to Stat3: Stat3 activation by the Rho GTPases. *Exp Cell Res.* 2011;317:1787-1795
106. Ford AM, Palmi C, Bueno C, Hong D, Cardus P, Knight D, et al. The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. *J Clin Invest.* 2009;119:826-836
107. Fuka G, Kauer M, Kofler R, Haas OA, Panzer-Grumayer R. The leukemia-specific fusion gene ETV6/RUNX1 perturbs distinct key biological functions primarily by gene repression. *PLoS One.* 2011;6:e26348
108. Kaindl U, Morak M, Portsmouth C, Mecklenbrauker A, Kauer M, Zeginigg M, et al. Blocking ETV6/RUNX1-induced MDM2 overexpression by Nutlin-3 reactivates p53 signaling in childhood leukemia. *Leukemia.* 2014;28:600-608

109. Fuka G, Kantner HP, Grausenburger R, Inthal A, Bauer E, Krapf G, et al. Silencing of ETV6/RUNX1 abrogates PI3K/AKT/mTOR signaling and impairs reconstitution of leukemia in xenografts. *Leukemia*. 2012;26:927-933
110. Inthal A, Krapf G, Beck D, Joas R, Kauer MO, Orel L, et al. Role of the erythropoietin receptor in ETV6/RUNX1-positive acute lymphoblastic leukemia. *Clin Cancer Res*. 2008;14:7196-7204
111. Torrano V, Procter J, Cardus P, Greaves M, Ford AM. ETV6-RUNX1 promotes survival of early B lineage progenitor cells via a dysregulated erythropoietin receptor. *Blood*. 2011;118:4910-4918
112. Mullighan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest*. 2012;122:3407-3415
113. Raynaud S, Cave H, Baens M, Bastard C, Cacheux V, Grosgeorge J, et al. The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood*. 1996;87:2891-2899
114. Cave H, Cacheux V, Raynaud S, Brunie G, Bakkus M, Cochaux P, et al. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia*. 1997;11:1459-1464
115. Patel N, Goff LK, Clark T, Ford AM, Foot N, Lillington D, et al. Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br J Haematol*. 2003;122:94-98
116. Kar A, Gutierrez-Hartmann A. Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. *Crit Rev Biochem Mol Biol*. 2013;48:522-543
117. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev*. 2007;21:1882-1894
118. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem*. 2011;80:437-471
119. Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer*. 2005;41:2462-2478

120. Turner DP, Findlay VJ, Moussa O, Watson DK. Defining ETS transcription regulatory networks and their contribution to breast cancer progression. *J Cell Biochem.* 2007;102:549-559
121. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005;310:644-648
122. Sun C, Dobi A, Mohamed A, Li H, Thangapazham RL, Furusato B, et al. TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation. *Oncogene.* 2008;27:5348-5353
123. Baens M, Peeters P, Guo C, Aerssens J, Marynen P. Genomic organization of TEL: the human ETS-variant gene 6. *Genome Res.* 1996;6:404-413
124. Sasaki K, Nakamura Y, Maki K, Waga K, Nakamura F, Arai H, et al. Functional analysis of a dominant-negative DeltaETS TEL/ETV6 isoform. *Biochem Biophys Res Commun.* 2004;317:1128-1137
125. Lopez RG, Carron C, Ghysdael J. v-SRC specifically regulates the nucleocytoplasmic delocalization of the major isoform of TEL (ETV6). *J Biol Chem.* 2003;278:41316-41325
126. Roukens MG, Alloul-Ramdhani M, Vertegaal AC, Anvarian Z, Balog CI, Deelder AM, et al. Identification of a new site of sumoylation on Tel (ETV6) uncovers a PIAS-dependent mode of regulating Tel function. *Mol Cell Biol.* 2008;28:2342-2357
127. Mavrothalassitis G, Ghysdael J. Proteins of the ETS family with transcriptional repressor activity. *Oncogene.* 2000;19:6524-6532
128. Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem.* 1999;274:30132-30138
129. Park H, Seo Y, Kim JI, Kim WJ, Choe SY. Identification of the nuclear localization motif in the ETV6 (TEL) protein. *Cancer Genet Cytogenet.* 2006;167:117-121
130. Green SM, Coyne HJ, 3rd, McIntosh LP, Graves BJ. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem.* 2010;285:18496-18504

131. Coyne HJ, 3rd, De S, Okon M, Green SM, Bhachech N, Graves BJ, et al. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol.* 2012;421:67-84
132. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: a review. *Leuk Res.* 2012;36:945-961
133. Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, Orkin SH. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J.* 1997;16:4374-4383
134. Hock H, Meade E, Medeiros S, Schindler JW, Valk PJ, Fujiwara Y, et al. Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* 2004;18:2336-2341
135. Fenrick R, Wang L, Nip J, Amann JM, Rooney RJ, Walker-Daniels J, et al. TEL, a putative tumor suppressor, modulates cell growth and cell morphology of ras-transformed cells while repressing the transcription of stromelysin-1. *Mol Cell Biol.* 2000;20:5828-5839
136. Rompaey LV, Potter M, Adams C, Grosveld G. Tel induces a G1 arrest and suppresses Ras-induced transformation. *Oncogene.* 2000;19:5244-5250
137. Yamagata T, Maki K, Waga K, Mitani K. TEL/ETV6 induces apoptosis in 32D cells through p53-dependent pathways. *Biochem Biophys Res Commun.* 2006;347:517-526
138. Irvin BJ, Wood LD, Wang L, Fenrick R, Sansam CG, Packham G, et al. TEL, a putative tumor suppressor, induces apoptosis and represses transcription of Bcl-XL. *J Biol Chem.* 2003;278:46378-46386
139. Waga K, Nakamura Y, Maki K, Arai H, Yamagata T, Sasaki K, et al. Leukemia-related transcription factor TEL accelerates differentiation of Friend erythroleukemia cells. *Oncogene.* 2003;22:59-68
140. Takahashi W, Sasaki K, Kvomatsu N, Mitani K. TEL/ETV6 accelerates erythroid differentiation and inhibits megakaryocytic maturation in a human leukemia cell line UT-7/GM. *Cancer Sci.* 2005;96:340-348
141. Golub TR, Barker GF, Lovett M, Gilliland DG. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell.* 1994;77:307-316

142. Poirel H, Oury C, Carron C, Duprez E, Laabi Y, Tsapis A, et al. The TEL gene products: nuclear phosphoproteins with DNA binding properties. *Oncogene*. 1997;14:349-357
143. Wang Q, Dong S, Yao H, Wen L, Qiu H, Qin L, et al. ETV6 mutation in a cohort of 970 patients with hematologic malignancies. *Haematologica*. 2014;99:e176-178
144. Salomon-Nguyen F, Della-Valle V, Mauchauffe M, Busson-Le Coniat M, Ghysdael J, Berger R, et al. The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion. *Proc Natl Acad Sci U S A*. 2000;97:6757-6762
145. Otsubo K, Kanegane H, Eguchi M, Eguchi-Ishimae M, Tamura K, Nomura K, et al. ETV6-ARNT fusion in a patient with childhood T lymphoblastic leukemia. *Cancer Genet Cytogenet*. 2010;202:22-26
146. Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A, et al. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood*. 2000;96:2557-2561
147. Chakrabarti SR, Nucifora G. The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun*. 1999;264:871-877
148. Wang L, Hiebert SW. TEL contacts multiple co-repressors and specifically associates with histone deacetylase-3. *Oncogene*. 2001;20:3716-3725
149. Petrie K, Guidez F, Howell L, Healy L, Waxman S, Greaves M, et al. The histone deacetylase 9 gene encodes multiple protein isoforms. *J Biol Chem*. 2003;278:16059-16072
150. Boccuni P, MacGrogan D, Scandura JM, Nimer SD. The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). *J Biol Chem*. 2003;278:15412-15420
151. Schick N, Oakeley EJ, Hynes NE, Badache A. TEL/ETV6 is a signal transducer and activator of transcription 3 (Stat3)-induced repressor of Stat3 activity. *J Biol Chem*. 2004;279:38787-38796

152. Kuwata T, Gongora C, Kanno Y, Sakaguchi K, Tamura T, Kanno T, et al. Gamma interferon triggers interaction between ICSBP (IRF-8) and TEL, recruiting the histone deacetylase HDAC3 to the interferon-responsive element. *Mol Cell Biol.* 2002;22:7439-7448
153. Bastian LS, Kwiatkowski BA, Breininger J, Danner S, Roth G. Regulation of the megakaryocytic glycoprotein IX promoter by the oncogenic Ets transcription factor Fli-1. *Blood.* 1999;93:2637-2644
154. Kwiatkowski BA, Bastian LS, Bauer TR, Jr., Tsai S, Zielinska-Kwiatkowska AG, Hickstein DD. The ets family member Tel binds to the Fli-1 oncoprotein and inhibits its transcriptional activity. *J Biol Chem.* 1998;273:17525-17530
155. Fears S, Gavin M, Zhang DE, Hetherington C, Ben-David Y, Rowley JD, et al. Functional characterization of ETV6 and ETV6/CBFA2 in the regulation of the MCSFR proximal promoter. *Proc Natl Acad Sci U S A.* 1997;94:1949-1954
156. Fears S, Chakrabarti SR, Nucifora G, Rowley JD. Differential expression of TCL1 during pre-B-cell acute lymphoblastic leukemia progression. *Cancer Genet Cytogenet.* 2002;135:110-119
157. Xue HH, Bollenbacher-Reilly J, Wu Z, Spolski R, Jing X, Zhang YC, et al. The transcription factor GABP is a critical regulator of B lymphocyte development. *Immunity.* 2007;26:421-431
158. Xue HH, Bollenbacher J, Rovella V, Tripuraneni R, Du YB, Liu CY, et al. GA binding protein regulates interleukin 7 receptor alpha-chain gene expression in T cells. *Nat Immunol.* 2004;5:1036-1044
159. Yu S, Cui K, Jothi R, Zhao DM, Jing X, Zhao K, et al. GABP controls a critical transcription regulatory module that is essential for maintenance and differentiation of hematopoietic stem/progenitor cells. *Blood.* 2011;117:2166-2178
160. Chauhan BK, Reed NA, Zhang W, Duncan MK, Kilimann MW, Cvekl A. Identification of genes downstream of Pax6 in the mouse lens using cDNA microarrays. *J Biol Chem.* 2002;277:11539-11548
161. Chakrabarti SR, Sood R, Ganguly S, Bohlander S, Shen Z, Nucifora G. Modulation of TEL transcription activity by interaction with the ubiquitin-conjugating enzyme UBC9. *Proc Natl Acad Sci U S A.* 1999;96:7467-7472

162. Chakrabarti SR, Sood R, Nandi S, Nucifora G. Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent assembly into nuclear bodies. *Proc Natl Acad Sci U S A*. 2000;97:13281-13285
163. Wood LD, Irvin BJ, Nucifora G, Luce KS, Hiebert SW. Small ubiquitin-like modifier conjugation regulates nuclear export of TEL, a putative tumor suppressor. *Proc Natl Acad Sci U S A*. 2003;100:3257-3262
164. Arai H, Maki K, Waga K, Sasaki K, Nakamura Y, Imai Y, et al. Functional regulation of TEL by p38-induced phosphorylation. *Biochem Biophys Res Commun*. 2002;299:116-125
165. Stevens TM, Parekh V. Mammary Analogue Secretory Carcinoma. *Arch Pathol Lab Med*. 2016;140:997-1001
166. Neveu B, Spinella JF, Richer C, Lagace K, Cassart P, Lajoie M, et al. CLIC5: a novel ETV6 target gene in childhood acute lymphoblastic leukemia. *Haematologica*. 2016;101:1534-1543
167. Cooper CD, Newman JA, Gileadi O. Recent advances in the structural molecular biology of Ets transcription factors: interactions, interfaces and inhibition. *Biochem Soc Trans*. 2014;42:130-138
168. Ghazavi F, De Moerloose B, Van Loocke W, Wallaert A, Helmoortel HH, Ferster A, et al. Unique long non-coding RNA expression signature in ETV6/RUNX1-driven B-cell precursor acute lymphoblastic leukemia. *Oncotarget*. 2016;
169. Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH, et al. TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell*. 2009;5:43-53
170. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57
171. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13
172. Pierchala BA, Munoz MR, Tsui CC. Proteomic analysis of the slit diaphragm complex: CLIC5 is a protein critical for podocyte morphology and function. *Kidney Int*. 2010;78:868-882

173. Wegner B, Al-Momany A, Kulak SC, Kozlowski K, Obeidat M, Jahroudi N, et al. CLIC5A, a component of the ezrin-podocalyxin complex in glomeruli, is a determinant of podocyte integrity. *Am J Physiol Renal Physiol*. 2010;298:F1492-1503
174. Gagnon LH, Longo-Guess CM, Berryman M, Shin JB, Saylor KW, Yu H, et al. The chloride intracellular channel protein CLIC5 is expressed at high levels in hair cell stereocilia and is essential for normal inner ear function. *J Neurosci*. 2006;26:10188-10198
175. Salles FT, Andrade LR, Tanda S, Grati M, Plona KL, Gagnon LH, et al. CLIC5 stabilizes membrane-actin filament linkages at the base of hair cell stereocilia in a molecular complex with radixin, taperin, and myosin VI. *Cytoskeleton (Hoboken)*. 2014;71:61-78
176. Al-Momany A, Li L, Alexander RT, Ballermann BJ. Clustered PI(4,5)P(2) accumulation and ezrin phosphorylation in response to CLIC5A. *J Cell Sci*. 2014;127:5164-5178
177. Tavasoli M, Al-Momany A, Wang X, Li L, Edwards JC, Ballermann BJ. Both CLIC4 and CLIC5A activate ERM proteins in glomerular endothelium. *Am J Physiol Renal Physiol*. 2016;311:F945-F957
178. Berryman M, Bretscher A. Identification of a novel member of the chloride intracellular channel gene family (CLIC5) that associates with the actin cytoskeleton of placental microvilli. *Mol Biol Cell*. 2000;11:1509-1521
179. Berryman M, Bruno J, Price J, Edwards JC. CLIC-5A functions as a chloride channel in vitro and associates with the cortical actin cytoskeleton in vitro and in vivo. *J Biol Chem*. 2004;279:34794-34801
180. Irigoyen M, Garcia-Ruiz JC, Berra E. The hypoxia signalling pathway in haematological malignancies. *Oncotarget*. 2017;8:36832-36844
181. Deynoux M, Sunter N, Herault O, Mazurier F. Hypoxia and Hypoxia-Inducible Factors in Leukemias. *Front Oncol*. 2016;6:41
182. Kantner HP, Warsch W, Delogu A, Bauer E, Esterbauer H, Casanova E, et al. ETV6/RUNX1 induces reactive oxygen species and drives the accumulation of DNA damage in B cells. *Neoplasia*. 2013;15:1292-1300
183. Takizawa H, Boettcher S, Manz MG. Demand-adapted regulation of early hematopoiesis in infection and inflammation. *Blood*. 2012;119:2991-3002

184. Baldrige MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*. 2010;465:793-797
185. Essers MA, Offner S, Blanco-Bose WE, Waibler Z, Kalinke U, Duchosal MA, et al. IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*. 2009;458:904-908
186. Takizawa H, Manz MG. Impact of inflammation on early hematopoiesis and the microenvironment. *Int J Hematol*. 2017;106:27-33
187. Ren G, Zhang L, Zhao X, Xu G, Zhang Y, Roberts AI, et al. Mesenchymal stem cell-mediated immunosuppression occurs via concerted action of chemokines and nitric oxide. *Cell Stem Cell*. 2008;2:141-150
188. Kim JJ, Lee SB, Park JK, Yoo YD. TNF-alpha-induced ROS production triggering apoptosis is directly linked to Romo1 and Bcl-X(L). *Cell Death Differ*. 2010;17:1420-1434
189. Khotetskaya YB, Holla VR, Farago AF, Mills Shaw KR, Meric-Bernstam F, Hong DS. Targeting TRK family proteins in cancer. *Pharmacol Ther*. 2017;173:58-66
190. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:pl1
191. Bollig-Fischer A, Michelhaugh SK, Wijesinghe P, Dyson G, Kruger A, Palanisamy N, et al. Cytogenomic profiling of breast cancer brain metastases reveals potential for repurposing targeted therapeutics. *Oncotarget*. 2015;6:14614-14624
192. Tacconelli A, Farina AR, Cappabianca L, Desantis G, Tessitore A, Vetuschi A, et al. TrkA alternative splicing: a regulated tumor-promoting switch in human neuroblastoma. *Cancer Cell*. 2004;6:347-360
193. Tomasson MH, Xiang Z, Walgren R, Zhao Y, Kasai Y, Miner T, et al. Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. *Blood*. 2008;111:4797-4808
194. Li Z, Beutel G, Rhein M, Meyer J, Koenecke C, Neumann T, et al. High-affinity neurotrophin receptors and ligands promote leukemogenesis. *Blood*. 2009;113:2028-2037

195. Ter Elst A, Diks SH, Kampen KR, Hoogerbrugge PM, Ruijtenbeek R, Boender PJ, et al. Identification of new possible targets for leukemia treatment by kinase activity profiling. *Leuk Lymphoma*. 2011;52:122-130
196. Gandemer V, Rio AG, de Tayrac M, Sibut V, Mottier S, Ly Sunnaram B, et al. Five distinct biological processes and 14 differentially expressed genes characterize TEL/AML1-positive leukemia. *BMC Genomics*. 2007;8:385
197. Yu X, Long YC, Shen HM. Differential regulatory functions of three classes of phosphatidylinositol and phosphoinositide 3-kinases in autophagy. *Autophagy*. 2015;11:1711-1728
198. Munson MJ, Ganley IG. MTOR, PIK3C3, and autophagy: Signaling the beginning from the end. *Autophagy*. 2015;11:2375-2376
199. Pasquier B. SAR405, a PIK3C3/Vps34 inhibitor that prevents autophagy and synergizes with MTOR inhibition in tumor cells. *Autophagy*. 2015;11:725-726
200. Kristensen L, Kristensen T, Abildgaard N, Thomassen M, Frederiksen M, Mourits-Andersen T, et al. High expression of PI3K core complex genes is associated with poor prognosis in chronic lymphocytic leukemia. *Leuk Res*. 2015;39:555-560
201. Ronan B, Flamand O, Vescovi L, Dureuil C, Durand L, Fassy F, et al. A highly potent and selective Vps34 inhibitor alters vesicle trafficking and autophagy. *Nat Chem Biol*. 2014;10:1013-1019
202. Liu X, Wang A, Liang X, Liu J, Zou F, Chen C, et al. Simultaneous inhibition of Vps34 kinase would enhance PI3Kdelta inhibitor cytotoxicity in the B-cell malignancies. *Oncotarget*. 2016;7:53515-53525
203. Yang ZQ, Liu G, Bollig-Fischer A, Giroux CN, Ethier SP. Transforming properties of 8p11-12 amplified genes in human breast cancer. *Cancer Res*. 2010;70:8487-8497
204. Seol MA, Chu IS, Lee MJ, Yu GR, Cui XD, Cho BH, et al. Genome-wide expression patterns associated with oncogenesis and sarcomatous transdifferentiation of cholangiocarcinoma. *BMC Cancer*. 2011;11:78
205. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*. 2007;5:1263-1275

206. Wang P, Zhuang C, Huang D, Xu K. Downregulation of miR-377 contributes to IRX3 deregulation in hepatocellular carcinoma. *Oncol Rep.* 2016;36:247-252
207. Lee NH, Kim M, Oh SY, Kim SG, Kwon HC, Hwang TH. Gene expression profiling of hematologic malignant cell lines resistant to oncolytic virus treatment. *Oncotarget.* 2017;8:1213-1225
208. Breitbach CJ, Parato K, Burke J, Hwang TH, Bell JC, Kim DH. Pexa-Vec double agent engineered vaccinia: oncolytic and active immunotherapeutic. *Curr Opin Virol.* 2015;13:49-54
209. Ooi YS, Stiles KM, Liu CY, Taylor GM, Kielian M. Genome-wide RNAi screen identifies novel host proteins required for alphavirus entry. *PLoS Pathog.* 2013;9:e1003835
210. Stiles KM, Kielian M. Role of TSPAN9 in Alphavirus Entry and Early Endosomes. *J Virol.* 2016;90:4289-4297
211. Delva E, Tucker DK, Kowalczyk AP. The desmosome. *Cold Spring Harb Perspect Biol.* 2009;1:a002543
212. Yang X, Wang J, Li WP, Jin ZJ, Liu XJ. Desmocollin 3 mediates follicle stimulating hormone-induced ovarian epithelial cancer cell proliferation by activating the EGFR/Akt signaling pathway. *Int J Clin Exp Pathol.* 2015;8:6716-6723
213. Ezzat Nel S, Tahoun N. The role of Napsin-A and Desmocollin-3 in classifying poorly differentiating non-small cell lung carcinoma. *J Egypt Natl Canc Inst.* 2016;28:13-22
214. Zhan C, Yan L, Wang L, Sun Y, Wang X, Lin Z, et al. Identification of immunohistochemical markers for distinguishing lung adenocarcinoma from squamous cell carcinoma. *J Thorac Dis.* 2015;7:1398-1405
215. Salerno EP, Bedognetti D, Mauldin IS, Deacon DH, Shea SM, Pinczewski J, et al. Human melanomas and ovarian cancers overexpressing mechanical barrier molecule genes lack immune signatures and have increased patient mortality risk. *Oncoimmunology.* 2016;5:e1240857
216. Ballatori N, Christian WV, Wheeler SG, Hammond CL. The heteromeric organic solute transporter, OSTalpha-OSTbeta/SLC51: a transporter for steroid-derived molecules. *Mol Aspects Med.* 2013;34:683-692

217. Robbins CA, Tempel BL. Kv1.1 and Kv1.2: similar channels, different seizure models. *Epilepsia*. 2012;53 Suppl 1:134-141
218. Nagaoka-Yasuda R, Matsuo N, Perkins B, Limbaeck-Stokin K, Mayford M. An RNAi-based genetic screen for oxidative stress resistance reveals retinol saturase as a mediator of stress resistance. *Free Radic Biol Med*. 2007;43:781-788
219. Tschop K, Conery AR, Litovchick L, Decaprio JA, Settleman J, Harlow E, et al. A kinase shRNA screen links LATS2 and the pRB tumor suppressor. *Genes Dev*. 2011;25:814-830
220. Huang X, Wang JY, Lu X. Systems analysis of quantitative shRNA-library screens identifies regulators of cell adhesion. *BMC Syst Biol*. 2008;2:49
221. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*. 2009;137:835-848
222. Kassner PD. Discovery of novel targets with high throughput RNA interference screening. *Comb Chem High Throughput Screen*. 2008;11:175-184
223. Perrimon N, Friedman A, Mathey-Prevoit B, Eggert US. Drug-target identification in *Drosophila* cells: combining high-throughout RNAi and small-molecule screens. *Drug Discov Today*. 2007;12:28-33
224. Mohr SE, Smith JA, Shamu CE, Neumuller RA, Perrimon N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol*. 2014;15:591-600
225. Singh R, Park D, Xu J, Hosur R, Berger B. Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. *Nucleic Acids Res*. 2010;38:W508-515
226. Haery L, Thompson RC, Gilmore TD. Histone acetyltransferases and histone deacetylases in B- and T-cell development, physiology and malignancy. *Genes Cancer*. 2015;6:184-213
227. Cartwright P, Muller H, Wagener C, Holm K, Helin K. E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene*. 1998;17:611-623

228. Trimarchi JM, Fairchild B, Verona R, Moberg K, Andon N, Lees JA. E2F-6, a member of the E2F family that can behave as a transcriptional repressor. *Proc Natl Acad Sci U S A.* 1998;95:2850-2855
229. Trimarchi JM, Fairchild B, Wen J, Lees JA. The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. *Proc Natl Acad Sci U S A.* 2001;98:1519-1524
230. MacDonald G, Stramwasser M, Mueller CR. Characterization of a negative transcriptional element in the BRCA1 promoter. *Breast Cancer Res.* 2007;9:R49
231. Mishima Y, Jayasinghe CD, Lu K, Otani J, Shirakawa M, Kawakami T, et al. Nucleosome compaction facilitates HP1 γ binding to methylated H3K9. *Nucleic Acids Res.* 2015;43:10200-10212
232. Lambert JP, Baetz K, Figeys D. Of proteins and DNA--proteomic role in the field of chromatin research. *Mol Biosyst.* 2010;6:30-37
233. Rusconi F, Guillonneau F, Praseuth D. Contributions of mass spectrometry in the study of nucleic acid-binding proteins and of nucleic acid-protein interactions. *Mass Spectrom Rev.* 2002;21:305-348
234. Guillen-Ahlers H, Shortreed MR, Smith LM, Olivier M. Advanced methods for the analysis of chromatin-associated proteins. *Physiol Genomics.* 2014;46:441-447
235. Tacheny A, Dieu M, Arnould T, Renard P. Mass spectrometry-based identification of proteins interacting with nucleic acids. *J Proteomics.* 2013;94:89-109
236. Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc.* 2016;11:316-326
237. Bonnay F, Nguyen XH, Cohen-Berros E, Troxler L, Batsche E, Camonis J, et al. Akirin specifies NF-kappaB selectivity of Drosophila innate immune response via chromatin remodeling. *EMBO J.* 2014;33:2349-2362
238. Nowak SJ, Aihara H, Gonzalez K, Nibu Y, Baylies MK. Akirin links twist-regulated transcription with the Brahma chromatin remodeling complex during embryogenesis. *PLoS Genet.* 2012;8:e1002547

239. Waas WF, Dalby KN. Transient protein-protein interactions and a random-ordered kinetic mechanism for the phosphorylation of a transcription factor by extracellular-regulated protein kinase 2. *J Biol Chem.* 2002;277:12532-12540
240. Wickramasinghe VO, McMurtrie PI, Marr J, Amagase Y, Main S, Mills AD, et al. MCM3AP is transcribed from a promoter within an intron of the overlapping gene for GANP. *J Mol Biol.* 2011;406:355-361
241. Schneider M, Hellerschmied D, Schubert T, Amlacher S, Vinayachandran V, Reja R, et al. The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression. *Cell.* 2015;162:1016-1028
242. Jani D, Lutz S, Hurt E, Laskey RA, Stewart M, Wickramasinghe VO. Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acids Res.* 2012;40:4562-4573
243. Kirwan M, Walne AJ, Plagnol V, Velangi M, Ho A, Hossain U, et al. Exome sequencing identifies autosomal-dominant SRP72 mutations associated with familial aplasia and myelodysplasia. *Am J Hum Genet.* 2012;90:888-892
244. Knezevich SR, McFadden DE, Tao W, Lim JF, Sorensen PH. A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nat Genet.* 1998;18:184-187