Université de Montréal

Understanding the Shared Genetic Risk for Psychosis and Substance Use Disorders: a Study of Genetic Markers of Endogenous and Exogenous Cannabinoid-Related Risk

Par

Laurent Elkrief

Faculté de médecine – Programme de Sciences biomédicales

Mémoire présenté(e) en vue de l'obtention du grade de maitrise

en sciences biomédicales, option sciences psychiatriques

Février, 2021

© Laurent Elkrief, 2021

Université de Montréal

Département de psychiatrie et d'addictologie, Faculté de médecine

---

Ce mémoire (ou cette thèse) intitulé(e)

Understanding the Shared Genetic Risk for Psychosis and Substance Use Disorders: a Study of Genetic Markers of Endogenous and Exogenous Cannabinoid-Related Risk

Présenté par

Laurent Elkrief

A été évalué(e) par un jury composé des personnes suivantes

Dr. Linda Booij
Président-rapporteur

Dr. Patricia Conrod
Directeur de recherche

Dr. Clairélaine Ouellet-Plamondon
Membre du jury

Résumé

La consommation de cannabis durant l'adolescence est associée à des risques accrus de problèmes de santé mentale, y compris la toxicomanie et la psychose. Tout en considérant qu'une partie de l'étiologie de ces troubles est héréditaire, nous avons étudié le risque génétique de psychose et de troubles liés à l'usage de substances et leurs relations avec le cannabis et le système endocannabinoïde. Dans notre premier travail, nous avons étudié la relation entre les marqueurs génétiques endocannabinoïdes et les troubles d'usage d'alcool (TUA) pour deux cohortes d'adolescents. À l'aide d'approches de gènes candidats, nous avons démontré une relation significative entre ces gènes endocannabinoïdes et les TUA, mais ces résultats n'ont pas été répliqués chez une deuxième cohorte indépendante. Lors d'une seconde étude, nous avons examiné si la relation entre le score de risque polygénique pour la schizophrénie (PRS-Sz) et les expériences prépsychotiques (PLE) est médiée et/ou modérée par la consommation de cannabis, pour deux cohortes indépendantes. Des modèles de régression de médiation et de modération ont été utilisés pour examiner dans quelle mesure la relation prospective entre PRS-Sz et PLE est expliquée par la consommation de cannabis. Les résultats des analyses de médiation et de modération n'étaient pas significatifs, bien que le PRS-Sz et la consommation de cannabis aient tous deux prédit indépendamment les PLE. Ces résultats suggèrent que la consommation de cannabis reste un facteur de risque de psychose, au-delà de la vulnérabilité génétique connue pour la schizophrénie et qu'il n'y a pas de preuve que les individus génétiquement vulnérables étaient plus sensibles aux conséquences psychotiques de la consommation de cannabis. Le travail décrit démontre que les risques posés par la consommation de cannabis chez les adolescents pourraient ne pas être associés à une prédisposition génétique aux maladies psychiatriques, nonobstant l'implication du système endocannabinoïde dans la pathogenèse de ces mêmes maladies.

Mots clés: endocannabinoïde, cannabis, trouble d'usage d'alcool, psychose, gène candidat, score de risque polygénique

Abstract

Cannabis consumption during adolescence, increases the likelihood of adverse mental health outcomes, including substance abuse and psychosis. Considering that part of the etiology of these disorders are heritable, we aimed to elucidate the genetic risk for psychosis and substance use disorders and their relationships to cannabis and the endocannabinoid system. In our first work, we investigated the relationship between endocannabinoid genetic markers and alcohol use disorder in two adolescent cohorts. Through candidate gene approaches we demonstrated a significant relationship between these endocannabinoid genes and AUD, but the results were not replicated in the second cohort. In a second work, we examined if the relationship between polygenic risk score for schizophrenia (PRS-Sz) and psychotic like experiences (PLE) is mediated and/or moderated by cannabis use, in two cohorts. Mediation and moderation regression models were used to examine the extent to which the prospective relationship between PRS-Sz and PLE is accounted for by cannabis use. The results of both the mediation and moderation analyses were not significant, although PRS-Sz and cannabis use both independently predicted PLE. These results suggest that cannabis use remains a risk factor for psychotic-like experiences, over and above known genetic vulnerability for schizophrenia and there was no evidence that genetically vulnerable individuals were more susceptible to the psychosis-related outcomes of adolescent onset cannabis use. The work described demonstrates that the risks posed by adolescent cannabis consumption may be unrelated to one's genetic predisposition to psychiatric disease, notwithstanding the involvement of the endocannabinoid system in the pathogenesis of these same diseases.

Keywords: endocannabinoid, cannabis, alcohol use disorder, psychosis, candidate gene, polygenic risk score

Table of contents

Résumé 5

Abstract 7

Table of contents 9

List of tables 17

List of figures 19

Abbreviations 21

Acknowledgements 25

Preface 27

Introduction 29

Cannabis and the Endocannabinoid System 29

A biochemical overview of the endocannabinoid system 29

Cannabis, cannabinoids and the endocannabinoid system 30

Mental health risks associated to cannabis consumption 31

Addiction or Substance Use Disorders 31

Cannabinoids, endocannabinoids and addictive disorders 32

Psychiatric burden of cannabis consumption 35

Genetic markers of cannabinoid-related risk for mental health 38

Endocannabinoid genes as a marker for Substance Use Disorders 39

Polygenic risk scores: A reliable alternative to candidate gene methods 40

Objectives and hypotheses 44

Methods 46

Participants 46

Study 1 46
IMAGEN 46
Saguenay Youth Survey 47
Study 2 47
IMAGEN 47
Utrecht cannabis cohort 47
Genetic data 48
Quality control 48
STUDY 1 48
IMAGEN 48
SYS 48
STUDY 2 48
IMAGEN 48
Utrecht cannabis cohort 49
Phenotypes evaluated 49
STUDY 1 49
Alcohol misuse 49
IMAGEN 49
SYS 50
Covariates 50
STUDY 2 50
Cannabis use 50
IMAGEN 50
Utrecht cannabis cohort 51

Psychotic-like experiences 51
Polygenic risk score construction 52
Statistical analyses 53
Study 1 53
Set-based test 53
Case-control analysis and logistic regression 54
Generalized multifactor dimensionality reduction (GMDR) 54
Study 2 54
Regression analysis 54
Mediation and moderation analysis 55
Article 1: Endocannabinoid gene x gene interaction association to Alcohol Use Disorder in two adolescent cohorts 56
Author contributions 57
Abstract 58
Introduction 59
Materials and methods 62
Participants 62
Phenotype evaluated 63
Alcohol misuse 63
Covariates 63
Pipeline of SNP Selection 63
Statistical analysis 64
Generalized multifactor dimensionality reduction (GMDR) 65
SYS replication cohort 65

Results 66
Set-based tests: identifying candidate SNP 66
Case-control analysis and sensitivity analysis 67
GMDR: SNP x SNP interactions 67
Discussion 68
Endocannabinoid interactions in the brain and emotional regulation 70
Conflict of interest 72
Funding 72
Bibliography 73
Tables in article 74
Figures in Article 78
Article 2: Independent contribution of polygenic risk for schizophrenia and cannabis use in predicting psychotic-like experiences in young adulthood: Testing gene x environment moderation and mediation. 79
Author contributions 80
Abstract 81
Introduction 82
Methods 85
Participants 85
IMAGEN sample 85
Utrecht cannabis cohort 86
Phenotype measures 86
Cannabis use measures 86
Psychotic-like experience measures 87

Genetic Data 87
IMAGEN........ 87
Utrecht 87
Analysis 87
Polygenic risk scores (PRS) 87
Regression analysis 88
Mediation and moderation analysis 89
Results........ 90
Sample characteristics 90
Association of PRS-Sz and cannabis use........ 90
Association of PRS-Sz with PLE 90
Sensitivity analysis 91
Moderation and mediation analysis 91
Discussion 92
Limitations 94
Conclusion........ 95
Acknowledgements 95
Financial support 95
Bibliography........ 96
Tables in publication........ 97
Figures in publication........ 100
Conclusion 102
Interpretation and limitations of the presented studies 102
Current state of psychiatric genetics research 107

A field made accessible by the Psychiatric Genomics Consortium 107
A roadmap for future studies 112
Hypothesis-based methods in the era of modern psychiatric genomics: Biologically informed polygenic risk scores 112
Translating psychiatric genomics research to clinical settings 113
Bibliography 115
Annex A – Supplementary data article 1 138
Supplementary Methods 138
Phenotype Evaluated 138
Set-Based Test 138
Case-Control Analysis 139
Genetic Analysis of Population Stratification 139
Covariables 140
Seattle Seq Annotations of SNPs 140
Supplementary Results 141
Supplementary Tables 142
Supplementary Figures 157
Bibliography – Annex A 158
Annex B - Supplementary data Article 2 159
Supplemental material 159
Imagen Dataset 159
Utrecht Cannabis Cohort 160
Quality control 160
IMAGEN 160

Utrecht Cannabis Cohort 160

Base Data Set 161

CLOSUK + PGS (Pardiñas et al. 2018) 161

Cannabis Use (Pasman et al., 2018) 161

Polygenic Risk Score 161

Supplementary Figures and Tables 163

Bibliography – Annex B 178

List of tables

Tableau 1. – Table 1 (article 1). 74
Tableau 2. – Table 2 (article 1). 75
Tableau 3. – Table 3 (article 1). 76
Tableau 4. – Table 4 (article 1). 77
Tableau 5. – Table 1 (article 2). 97
Tableau 6. – Table 2 (article 2) .. 98
Tableau 7. – Table 3 (article 2). 99
Tableau 8. – Suppl. Table S1 (article 1). 142
Tableau 9. – Suppl. Table S2 (article 1). 143
Tableau 10. – Suppl. Table S3 (article 1). 143
Tableau 11. – Suppl. Table S4 (article 1). 144
Tableau 12. – Suppl. Table S5 (article 1). 145
Tableau 13. – Suppl. Table S6 (article 1). 146
Tableau 14. – Suppl. Table S7 (article 1). 147
Tableau 15. – Suppl. Table S8 (article 1). 148
Tableau 16. – Suppl. Table S9 (article 1). 149
Tableau 17. – Suppl. Table S10 (article 1). 152
Tableau 18. – Suppl. Table S11 (article 1). 153
Tableau 19. – Suppl. Table S12 (article 1). 154
Tableau 20. – Suppl. Table S13 (article 1). 155
Tableau 21. – Suppl. Table S14 (article 1). 156
Tableau 22. – Suppl. Table S15 (article 1). 156
Tableau 23. – Suppl. Table S16 (article 1). 157
Tableau 24. – Suppl. Table S1 (article 2). 164
Tableau 25. – Suppl. Table S2 (article 2). 165
Tableau 26. – Suppl. Table S3(article 2). 166
Tableau 27. – Suppl. Table S4(article 2). 168

Tableau 28. – Suppl. Table S5 (article 2). 169
Tableau 29. – Suppl. Table S6 (article 2). 170
Tableau 30. – Suppl. Table S7 (article 2). 171
Tableau 31. – Suppl. Table S8 (article 2). 173
Tableau 32. – Suppl. Table S9 (article 2). 173
Tableau 33. – Suppl. Table S10 (article 2). 174
Tableau 34. – Suppl. Table S11 (article 2). 175
Tableau 35. – Suppl. Table S12 (article 2). 175

List of figures

Figure 1. – Figure 1 (article 1) . 78

Figure 2. – Figure 1 (article 2). 100

Figure 3. – Figure 2 (article 2). 101

Figure 4. – Suppl. Figure S1 (article 1). 157

Figure 5. – Suppl. Figure S1 (article 2). 163

Abbreviations

AEA: anandamide

2-AG: 2-arachidonoylglycerol

BNST: basal nucleus of the stria terminalis

eCB: endocannnabinoid

CB1: cannabinoid receptor one

CB2: cannabinoid receptor two

CBD: cannabidiol

CNR1: cannabinoid receptor one gene

CNR2: cannabinoid receptor two gene

CRF: corticotropin-releasing factor

DAGL: diaglycerol lipase

DAGL: diaglycerol lipase gene

FAAH: fatty acid amide hydrolase

FAAH: fatty acid amide hydrolase gene

GWAS: genome-wide association studies

MAGL: monoacylglycerol lipase

MGLL: monoacylglycerol lipase gene

MR: mendelian randomization

NAc: Nucleus Accumbens

NAPE-PLD: N-acylphosphatidylethanolamine-specific phospholipase D

NAPEPLD: N-acylphosphatidylethanolamine-specific phospholipase D gene

PLE: psychotic-like experiences

PRS: polygenic risk score

PRS-Sz: polygenic risk score for schizophrenia

SNP: single nucleotide polymorphism

THC: Δ-9-tetrahydrocannabinol

VTA: Ventral Tegmental area

For mami Chelo. Your strength is my inspiration.

Acknowledgements

A special thank you to all of my teammates at the Conrod Venture lab, and my labmates on our floor at the CR-CHU Sainte Justine. I would like to especially thank Dr. Guillaume Huguet, who mentored me throughout my time at the Conrod lab. Without Dr. Huguet's dedication, time, and tutelage, I would have not learned the fundamental notions required to complete this work. I also would like to give special thanks to Dr. Patricia Conrod, my thesis supervisor. I would like to thank you, Dr. Conrod, not only for being an excellent academic supervisor, through whom I have learned how to formulate my curiosity into meaningful and testable scientific hypotheses, but also for continuously motivating me to strive for greatness.

Thank you to my family, for their love and encouragement, as well as their input and feedback on my work (special mention to my brilliant mother). Finally, thank you to my incredible fiancée who continues to support me through thick and thin. Laurie you are my source of inspiration and motivation. I love you.

Preface

Although the first recorded use of cannabis dates back to 2500 BCE, archeological and palaeontological evidence suggests that cannabis has been used by humans for almost 10,000 years (1). Today, cannabis is among the most commonly used substances in the world, with the latest Canadian estimates reporting that close to 50% of people aged 15 or older have consumed cannabis in their lifetime(2). Over the last year, close to 30% of Quebec residents aged 18-24 have consumed cannabis at least once in the last 3 months (3). In view of the socially accepted, and widespread, use of cannabis in the general population, an understanding of the effects of the substance on the body, and in particular the brain, is of public health importance.

The field of cannabinoid science was born in the 1940s, through the isolation of the first cannabinoid molecules (4). Since then, researchers have discovered a slew of receptors, enzymes and signalling molecules that are grouped into what is now known as the endogenous cannabinoid, or endocannabinoid (eCB), system. This system interacts with and is modulated by the hundreds of cannabinoid molecules that are found in varying concentrations in the cannabis plant. Moreover, the proteins of the eCB system - which are among the most widely expressed proteins in the mammalian nervous system(5)(6) – are involved in everything from fetal development, pain signalling, digestion, and higher order cognitive processes. Bearing the complexity of cannabinoids and endocannabinoid signalling in mind, a brief primer on certain fundamental biochemical notions of the endocannabinoid system is necessary. Thus, I will begin this work with a brief description of the functions of the various proteins of the "canonical" endocannabinoid system and the interactions of phytocannabinoids (cannabinoids produced by the cannabis plant) with the mammalian endocannabinoid system. Next, I will outline certain consequences of cannabis consumption on mental health and psychiatric disease, with a particular focus on substance use disorders and psychosis. Following this introductory chapter, the subsequent chapter will serve as an overview of genetic markers of cannabinoid-related risk for mental health. Particularly, I will examine the transition from candidate gene studies to a hypothesis free framework through the examples of research on endocannabinoid genetic

markers and their role in substance use disorders, and an introduction of polygenic risk scoring in cannabis-psychosis research. In the context of this literature review, I will advance the objectives and hypotheses of the research completed throughout this Master's degree. A description of the methodology used in my studies thesis will follow. Then, the two manuscripts completed during this degree will be presented. Finally, I will describe my work in the context of current psychiatric genomics literature, outline the strengths and limitations of my studies, and elaborate on the potential for future works.

Introduction

Cannabis and the Endocannabinoid System

A biochemical overview of the endocannabinoid system

The first scientific publication studying cannabis use for medical purposes dates back to the 1800(4). A century later, Δ-9-tetrahydrocannabinol (THC) – the main psychoactive molecule in the cannabis plant – was first isolated by Gaoni and Mechoulam (7). While this discovery and the isolation of other cannabinoids were tantamount to the field cannabinoid science, it was only two decades later that the binding sites for THC, the cannabinoid receptors one (CB1) and two (CB2), were discovered (8)(9). Thereafter, endogenous ligands and related enzymes responsible for the synthesis and degradation of these ligands have been identified, laying the foundations for the field of endocannabinoid research.

CB1, regarded as one of the most highly expressed G-protein coupled receptors in the human nervous system (5)(6), is expressed in many sub-cortical (amygdala, nucleus accumbens (NAc), ventral tegmental area (VTA)) and cortical brain regions and has been shown to be involved in the processes of brain development, cognition, reward and addiction(10)(11). While the neuronal nature of CB2 has been subject of controversy(12), it is generally accepted that CB2 is widely expressed by immune and glial cells, and is involved in the general neuronal "protective system"(13). While other endocannabinoid receptors have been proposed (GPR55 (14) and TRPV1(15)), most of the research focusing on the neurodevelopmental functions of the endocannabinoid system have focused on CB1 and CB2. The two main endocannabinoid signalling molecules, which act on CB1 and CB2, are anandamide (AEA) and 2-arachidonoylglycerol (2-AG). These non-classical neurotransmitters are not stored in vesicles(16). Rather they are produced, in an on demand fashion, by a calcium dependent enzymatic cascade and are rapidly degraded thereafter by enzymes such as fatty acid amide hydrolase (FAAH) and monoacylglycerol lipase (MAGL) (16)(17). N-acylphosphatidylethanolamine-specific phospholipase D (NAPE-PLD), plays a crucial role in the synthesis of AEA, while 2-AG is synthesized by diaglycerol lipase (DAGL). While a complete review of the biochemical functions of the endocannabinoid system can be found

here (17), this goes beyond the scope of this work. Succinctly, at the neuronal level, endocannabinoids can be seen as important mediators of the balance between excitatory and inhibitory neurotransmission(18), and are required for several forms of synaptic plasticity(19) and neurodevelopmental mechanisms(17).

Cannabis, cannabinoids and the endocannabinoid system

Cannabis is the taxonomic genus that includes three species of plants: *Cannabis Sativa, Cannabis Indica, Cannabis Ruderalis*. Through our consumption for religious, medicinal, and recreational purposes, humans have created an enduring and complicated relationship with the cannabis plant. Historically, this plant has been used as material for the fabrication of various materials including clothing, rope, and sails, in the form of hemp(20). Above its material uses, the cannabis plant has been consumed – in a variety of ways – for its euphoric effects. These euphoric effects are caused by the pharmacological effects of the multiple phytocannabinoid molecules. While researchers have identified over 140 phytocannabinoid molecules in the cannabis plant(21), the best studied and most significant molecules remain THC and cannabidiol (CBD). In short, THC can act as an agonist and antagonist of both CB1 and CB2, while CBD is more typically an antagonist of these receptors(22). Moreover, the action of THC and CBD on cannabinoid receptors depend on their complex pharmacodynamics and pharmacokinetics properties: specifically the mode of consumption, circulating concentration, and regularity of use (see Lucas et al. for review (23)). Circulating levels of THC and CBD have been shown to disrupt neural plasticity mechanisms such as short-term depression, long term depression and long-term potentiation (24)(25)(26). Moreover, THC acts at both excitatory and inhibitory synapses(17), as well as interacting with various neurotransmitter and neuromodulators including the serotoninergic, opioid and dopaminergic systems(27). While the acute intoxicating effects of phytocannabinoids are usually short lived (2-12 hours depending on mode of consumption)(23), cannabis use has various acute and chronic consequences on the brain; particularly, on the developing adolescent brain. Moreover, considering the inordinate increases in THC concentrations over the last decades(28), an increase in the mental health consequences of cannabis consumption, at the populational level, is of legitimate concern.

Mental health risks associated to cannabis consumption

From its roles in fetal axonal growth and synaptic formation(29) to its role in the maintenance of cognitive, behavioral, emotional and developmental processes(18) (30), the endocannabinoid system plays a vital role in the various stages of brain development. Thus, considering the innate vulnerability of the developing brain, some have hypothesized that disrupting the endocannabinoid system through the introduction of exogenous cannabinoids, particularly during neurodevelopmental processes, could increase the likelihood of deleterious consequences(31). Epidemiological evidence has linked adolescent cannabis consumption to serious adverse mental health outcomes, such as depressive disorders and increased suicidality(32). This risk is often shown to be dose dependent and increased in those who consume chronically(33). Correspondingly, cannabis consumption is an addictive process, with the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) (34), including cannabis use disorder (CUD) among the many other substance use disorders. At this point, I do feel it necessary to define "addiction" and outline the conceptualization of addiction I will refer to throughout this text. Please also note that although I recognize that they are not truly equivalent terms, for the sake of simplicity, I have chosen to use the terms "addiction", "substance abuse" and "substance use disorder" interchangeably throughout.

Addiction or Substance Use Disorders

I have elected to use the American Society for Addiction Medicine's short definition of addiction, for this work:

> *Addiction is a primary, chronic disease of brain reward, motivation, memory and related circuitry. Dysfunction in these circuits leads to characteristic biological, psychological, social and spiritual manifestations. This is reflected in an individual pathologically pursuing reward and/or relief by substance use and other behaviors.*
>
> *Addiction is characterized by inability to consistently abstain, impairment in behavioral control, craving, diminished recognition of significant problems with one's behaviors and interpersonal relationships, and a dysfunctional emotional response. Like other chronic diseases, addiction often involves cycles of relapse and remission. Without treatment or engagement in recovery activities, addiction is progressive and can result in disability or premature death*

This definition is based on notions proposed in the brain disease model of addiction, first published by Leshner in 1997 (35). Through this definition of addiction, one first recognizes that addiction, at its core, is a disease of the brain. In its most simplified version, the brain disease model of addiction posits that addictive substances or behaviours, will permanently modify the mesolimbic circuitry – or reward circuits – of the brain(35). It is also important to note that we cannot simply reduce addiction and addictive behaviour to only a "brain disease". As Leshner writes, "it is a brain disease for which the social contexts in which it has both developed and is expressed are critically important" (35). Thus, while some research (such as my own) cannot consider the social and cultural contexts of drug use and abuse, it is important to note that these considerations must be included in any complete model of addiction. Building upon Leshner's original model, Nora Volkow and Allan Koob – among others – have proposed a more complete neurobiological model of addiction(36)(37)(38). Their model conceptualizes drug addiction as a chronic cyclical process comprising of three stages: binge/intoxication, withdrawal/negative affect and preoccupation/anticipation. These stages are influenced by multiple neuroadaptation in three domains; 1) increased incentive salience, 2) decreased brain reward and increased stress, and 3) compromised executive function in three neuro-biological brain circuits (basal ganglia, extended amygdala, and the prefrontal cortex (PFC) (37)(38). As such, it is within the context of this extension of the brain disease model of addiction that we will understand the potential effects of cannabis consumption on the brain.

Cannabinoids, endocannabinoids and addictive disorders

Cannabis is among the most widely used addictive substances, with approximately 1 in 11 (9%) users developing a cannabis use disorder in Canada (39). Moreover, this estimate increases to 1 in 6 individuals, in those who report using cannabis during their adolescence(39). Cannabis use disorder is theorized to develop from the neuroadaptations that occur following repeated chronic exposure to THC(40). THC directly modifies reward circuits of the brain(41), with ample evidence suggesting that repeated exposure to THC reduces dopaminergic function in chronic cannabis users(42). Additionally, in rodent models, THC withdrawal is associated to increases in corticotropin-releasing factor (CRF) in the amygdala; similar increases of CRF have been observed in the withdrawal of other drugs of abuse such as nicotine, psychostimulants and opiates(43).

Chronic cannabis use has also been associated to a collection of morphological brain changes, identified in brain imaging studies. For example, CUD is associated with reductions in amygdala gray matter volume, in a dose dependent manner: higher CUD is inversely correlated to gray matter volume(40)(44)(45). This finding is particularly interesting considering the role of the amygdala in craving and drug seeking behaviours described in Volkow and Koob's model of addiction(37). Finally, chronic THC exposure leads to increased tolerance for the drug, in a mechanism associated to downregulations of CB1(46). While the consumption of phytocannabinoids, such as THC, are associated with the development of addictive processes, there is also evidence that the endocannabinoid system, in it of itself, plays an etiological role in the development of many substance use disorders (see (47) for review).

Through its anatomic distribution, and influence of neurotransmission and synaptic plasticity, the endocannabinoid system has an influence on many rewarding behaviours such as feeding, sexual activity, social interactions, and drug use(41)(48)(49)(50). As alluded to above, endocannabinoid receptors and related enzymes are expressed in many of the reward centers of the brain involved in the three stages of the addiction development, including the NAc, VTA, amygdala, and basal nucleus of the stria terminalis (BNST), in animal models and humans(6)(41). Expectantly, in rodent models, AEA and 2-AG have been shown to increase extracellular levels of dopamine in the NAc, in a CB1 dependent manner (51), by acting as a retrograde feedback system on presynaptic glutamatergic and GABAergic nerve terminals(42). The effect of endocannabinoids on dopamine transmission can also be blocked by CB1 antagonists(42). Correspondingly, levels of circulating endocannabinoids are modified by drugs of abuse such as ethanol(52). For example, Basavarajappa and colleagues (53) demonstrated, in rodents, that that acute ethanol use is associated with an increase in endocannabinoid signaling, while others have reported that alcohol use decreases endocannabinoid signaling(54)(55). Moreover, as is the case with other drugs of abuse, endocannabinoids mediate the reward signals associated with alcohol use(56). The endocannabinoid system has also been linked to the enhanced stress reactivity seen in people suffering from substance use disorders (37). This involvement is crucial, considering that the overreactive stress systems, and prolonged negative emotional states seen in individuals with addictive disorders, are theorized to be central to the development and maintenance of the

chronically relapsing compulsive drug taking behaviours seen in individuals with addictive disorders (37).

The addictive consequences of cannabis use are particularly important for adolescent users. First, when controlling for other factors, age of onset of use is predictive with future cannabis use disorder, i.e. those who begin using at a younger age have a higher risk of developing cannabis use disorder(40). Moreover, the peak of vulnerability for future CUD development is in late adolescence or young adulthood(57). Adolescent cannabis use also increases ones risk for other use of other substance of abuse(58). A recent meta-analysis reported a close to 8 fold increase in risk of other illicit substance use among those who reported using cannabis daily before the age of 17 (adjusted OR = 7.80; 95% CI, 4.46–13.63) (59). Interestingly, some adolescents may have a biological predisposition to cannabis consumption, with research demonstrating that reduced brain volume in the orbitofrontal cortex(65) and reduced grey matter in the hippocampus(66) are significantly associated to cannabis consumption in young teens. While the consequences of these morphological changes are not yet clear, it is evident that cannabis consumption during the vulnerable period of adolescence has a particular impact on substance use behaviours.

Although endocannabinoid signalling seems to play a role in the development of various addictive behaviours in animal models, the interactions between the endocannabinoid system and alcohol use disorder are of particular interest considering the widespread and concomitant use of both alcohol and cannabis among Canadians. In fact, according to the Canadian Tobacco, Alcohol and Drugs Survey (CTADS) in 2017, alcohol use and cannabis use were among the top three most used psychoactive substances in Canada(60). Alcohol use disorders, and alcohol use in general are associated with very high levels of morbidity, mortality and societal costs (61). A recent study demonstrated that 10,556 deaths (95% UI, 8285-13,609), and 440,709 disability adjusted life years (DALYs) (95% UI, 388,853, 527,260) were attributable to alcohol use in Canada in 2016, with 99,501 DALYs were attributable to AUD itself in Canada in 2016(62), and close to 21.5 million DALYs being attributed to AUD worldwide(61). Moreover, individuals with AUD are at increased risk of death (HR = 2.98; 95% CI, 2.96-3.00) and die 12.2 years younger on average(63). Epidemiological surveys of alcohol and cannabis use suggest that the combined use

of these substances is related to greater harms than alcohol use alone(64). In a study by Subbaraman and Kerr(64), the simultaneous use of alcohol and cannabis compared to alcohol use alone, was associated to an increased risk of monthly heavy drinking (OR = 5.56 ; 95% CI, 3.43-9.02), drunk driving (OR = 2.30; 95% CI, 1.61-3.30), and self-harm (OR = 2.22; 95% CI, 1.49-3.32). Cannabis use is also associated with the development of alcohol use disorder(65). For example, one study reported that cannabis use in individuals was associated to an increased incidence of alcohol use disorders three years later, compared to individuals who did not use cannabis (OR = 5.43; 95% CI, 4.54-6.49) (65). Taken together these results highlight the importance of understanding the shared etiology of cannabis use and alcohol use disorders, as well as the important influences that use of each drug has on other.

Overall, the underlying evidence suggests that cannabis use is addictive through an interaction with the endocannabinoid system and the brain circuits that mediate reward, and that the endocannabinoid system, which is modulated by other drugs of abuse, plays an independent role in the pathophysiology of various substance use disorders, including alcohol use disorder. Nevertheless, a gap in the literature remains: notably there is a lack of understanding of the underlying mechanisms driving the common vulnerability to substance use disorders and the role of the eCB system in determining an individual's vulnerability to these disorders.

Psychiatric burden of cannabis consumption

Although cannabis use does not lead to fatal overdose, cannabis use has been associated with an increased risk for a variety of psychiatric disorders and other substance use disorders(66). In a recent study of a nationally representative sample of US adults, Blanco and colleagues reported that cannabis use was significantly associated with increased odds of any psychiatric or substance use disorders (adjusted OR = 2.1; 95% CI, 1.8-2.6), when adjusting for a variety of confounders including history of divorce, history of SUD, history of social deviance, age, and sex among others(67). However, while cannabis use at time one significantly increased risk for any substance use disorder at time two (adjusted OR = 2.8; 95% CI, 2.4-3.4), it was not significantly associated to an increased risk for mood or anxiety disorders in this cohort(67). Among the best studied consequences of cannabis use, is the increased risk for psychosis and schizophrenia spectrum disorders among cannabis users. Multiple prospective, and longitudinal studies have

demonstrated that cannabis use often precedes psychosis(68), independent of alcohol and other drug use(69)(70)(71). In one study, following 6534 youths from the Northern Findland Birth Cohort, teenage cannabis use was significantly associated to developing a psychotic disorder, irrespective of baseline prodromal symptoms or other drug use (HR = 3.0; 95% CI, 1.1–8.0) (72). On the other hand, not using cannabis is a protective factor for patients, with non-users reporting less positive (OR = 0.42; 95% CI, 0.34-0.51), negative (OR =0.18; 95% CI, 0.15-21), and disorganization symptoms (OR = 0.33; 95% CI, 0.27-0.40)(73). These results remain consistent across systematic-review and meta-analysis (74)(75). This relationship is also dose-dependent, with the heaviest users having close to a four-fold increase in psychosis risk(76). The cannabis-psychosis association has also been demonstrated experimentally, with systematic review and meta-analyses reporting that even single doses of THC can induce temporary positive psychotic symptoms in healthy individuals, with large effect sizes (77). Thus, a recent international multi-site study, calculating the populational attributable fraction of cannabis use on psychosis, argued that if high-potency cannabis were no longer in circulation, at least 12.2% (95% CI, 3.0-16.1) of cases of first-episode psychosis could be prevented in the areas studied(78). This cannabis-psychosis relationship is highlighted by the particular vulnerability of adolescent cannabis consumers. Adolescent cannabis use is strongly associated to earlier onset of psychotic symptoms(79) and worsened prognosis(80) for those who do transition to psychosis.

The relationship between cannabis use and psychosis development is particularly interesting in the adolescent “clinical high-risk for psychosis”(81) population. These individuals are at a high risk for psychosis in the context of sub-clinical psychotic symptoms, functional decline and/or genetic risk(81). While cannabis use rates have systematically been shown to be higher in cohorts of clinical high-risk patients(82), a recent meta-analysis of enriched samples did not find that cannabis use was a risk factor for transition to psychosis in this population(83). As the authors note, this is probably due to the enrichment of the studies examined, and binary nature of the cannabis use measure in many studies(83). In fact, previous meta-analyses only found a significant association between cannabis use and psychosis transition in clinical high risk populations, among the highest users, or those with cannabis use disorder(84), further emphasizing the dose-dependent nature of the cannabis-psychosis relationship.

To understand the development of psychosis and the clinical-high risk state, researchers study the emergence of psychotic-like experiences (PLE). PLE are highly prevalent sub-clinical psychotic symptoms (85), reported in about 7% of individuals(86)(87). Meta-analysis has shown the PLE are associated with increased odds for any mental disorder (OR = 3.08; 95% CI, 2.26-4.21) (88). Self-reported PLE have also been shown to be risk factors for subsequent psychiatric hospitalization (89). In their study, Werbeloff and colleagues demonstrated that self-reported PLE predicted the risk for later hospitalization for nonaffective psychotic disorders 5 years after baseline (adjusted OR = 4.31; 95% CI, 2.21-8.41) and the risk of later hospitalization for other psychiatric disorders (adjusted OR = 2.21; 95% CI, 1.02-4.82) (89). PLE are also predictive of increased risk for suicidality later in adolescents; in a cohort of 16 and 17 year old individuals with suicidal ideation, co-occuring psychotic experiences predicted a 6-fold increase of persistence of suicidal ideation at ages 19 and 20(OR = 5.53; 95% CI, 1.33-23.00) (90). While these sub-clinical experiences are transitory in about 80% of individuals, PLE are persistent in 20% of individuals(91). The identification of PLE is particularly important in those considered at clinical-high risk for psychosis(81). These individuals have a higher incidence of clinical disorders along the psychotic spectrum, with 10-40% converting to psychotic disorders within 2 years(92)(93). With this in mind, along with more recent genetic, neuropsychological, social and environmental studies, authors have argued for a dimensional model of psychosis(94)(95)(96). Accordingly, there would exist a distribution of psychotic-like and psychotic experiences across the entire population, ranging from “normative” experiences to clinical psychotic symptoms(95). Thus, while psychotic-like experiences may not be exclusive to psychotic disorders, “these experiences can endure over time in some individuals, and may be followed by a psychotic disorders” (91). As such, diagnostic tools which examine psychotic-like experiences such as the Comprehensive Assessment of the At-Risk Mental State (97) and the Structured Interview for Psychosis Risk Syndromes (SIPS)(98), have been developed to identify those at a clinical-high risk for psychosis. Moreover, self-report questionnaires such as the Community Assessment of Psychic Experiences - 42 (CAPE-42) questionnaire (99), have been used as screening tools for high-risk states for psychosis(100)(101)(102)(103). Factor analysis of these clinical interviews, and self-report questionnaires have allowed researchers to subdivide PLE into replicable domains, such as

positive, negative, depressive and anxious symptoms(85)(104)(105). Taken together, the above literature would suggest that although the presence of PLE does not mean presence of psychotic disorder, their presence should be treated as part of the continuous psychotic spectrum, and measures of PLE could be used in the study of risk factors for future psychotic or other psychopathological disorders.

Considering the close relationship of psychotic-like experiences to psychotic disorders, many have tested the hypothesis that cannabis use also increases one's risk for PLE (see (106) for systematic-review). One study found that cannabis use is significantly associated with the positive PLE sub-clinical symptoms (β=0.061, $p<1X10^{-4}$), even after controlling for numerous confounding factors such as age, gender, socio-economic status, social support, alcohol use, cigarette smoking, and urbanicity, among others(105). Another study found that the relationship between PLE and cannabis use is increased in the heaviest of cannabis consumers(104); in those who spend >25 €/week on cannabis (i.e heaviest users), there was an increased odds for various domains of PLE such as, negative symptoms (OR = 3.4; 95% CI, 2.9-4.1), positive symptoms (OR = 3.0; 95% CI, 2.4-3.6), and depressive symptoms (OR = 2.8; 95% CI, 2.3-3.3)(104). Furthermore, cannabis use has also been shown to temporally precede PLE in adolescent cohorts(107), but psychotic-like experiences in childhood do not predict cannabis use(108). Overall, the study of PLE in cohorts of cannabis users, therefore, may be an interesting avenue to understand the nature and potential directionality of the cannabis-psychosis relationship.

Genetic markers of cannabinoid-related risk for mental health

Part of the etiology of substance use disorders, psychosis and psychopathology in general can be explained through genetics. In fact, the observation that psychiatric disorders are heritable can be traced to antiquity(109). These early observations were then confirmed by the first major twin, family and adoption studies in the early 20th century(110). By the end of the 20th century, with the introduction of new genotyping methods, researchers set out to explore the relationship between genomics and psychopathology by studying the relationship of polymorphisms in biologically plausible candidate genes and related psychopathologies. These "candidate gene studies", led to a boom in the field of psychiatric genetics. Within ten years, hundreds of studies

were published in the field, and there was excitement among researchers interested in understanding the basis of psychiatric disease. However, critics quickly noted that most candidate gene studies are riddled with bias, and major meta-analyses of large samples demonstrated that we cannot draw any conclusions from most candidate gene studies in psychiatry(111)(112). Furthermore, the introduction of large-scale genome wide association studies (GWAS) to the field further pushed the classical candidate gene studies aside. The main advantage of the these GWAS publications was the replicability of the results, and the ability to use the data to make valid predictions in other datasets(113), using tools like polygenic risk scores and mendelian randomization analyses. Although psychiatric genetic publications employing candidate gene methodologies have dramatically dropped in recent years, the statistical and genetic methods employed by these studies are foundational to the more advanced methods used in today's genomic research. It should be noted that the psychiatric candidate gene studies that have been the most successful to date are those studying substance use disorders – specifically in the field of alcohol abuse ((114) for review). These studies are among the only candidate gene studies in psychiatric genomics to have identified genetic markers that have been replicated by GWAS. Therefore, an outline of some of the literature surrounding candidate gene studies in psychiatry – in particular those studying the cannabinoid system and alcohol – remains relevant.

Endocannabinoid genes as a marker for Substance Use Disorders

The first studies examining the link between the genetics of the endocannabinoid system to substance use demonstrated that the *CNR1* gene, the gene coding for CB1 is associated with a range of diseases, psychiatric disorders and substance use (115)(116)(117). Although, many studies have looked at various aspects of the endocannabinoid genes and their relationship with substance abuse, and risk-taking behaviour, results vary between studies, and most studies have yet to be replicated, as is the case with most of the candidate gene literature. For example, one of the first studied single nucleotide polymorphism (SNP) in the *CNR1* gene, rs1049353;G1359A, was associated with severe alcoholism(118). In this cohort, the minor allele (A) was associated with the drug abuse behaviour, as well as impulsivity in another study (119) but in other studies the major allele (G) was associated with heroin addiction(120). Moreover, several haplotype blocks within the *CNR1* gene, have been associated to addiction and addictive behaviour

(121)(122)(123)(124). In a cohort of mostly alcohol dependent individuals, Agarwal and colleagues found that a haplotype block including SNPs rs806380, rs806368 and rs754387, was associated with cannabis dependence (123), however other studies have reported conflicting results in similar or different cohorts (121)(122). A meta-analysis of studies focusing on the *CNR1* gene and its association with illicit drug dependence found that among the variety of genetic changes in the *CNR1* gene studied at the time, only the AAT repeat (>=16) had a modest association to abuse(125). Among the other endocannabinoid genes, additional genetic loci have been implicated in endocannabinoid regulation of addiction. For example, polymorphisms in the *CNR2 (*coding for CB2) and *FAAH* (coding for FAAH) genes have been associated with problem drug use and addiction (126)(127)(128)(129). Other works have looked at the *MGLL* and *DAGL* genes, the gene coding for the MAGL and DAGL enzymes (130)(131)(132). Recently, Carey and colleagues (132), found that one SNP in the *MGLL* gene (rs604300) interacted with childhood adversity to predict cannabis dependence symptoms, while Muldoon et al. (131), found that rs549662 was significantly associated with less subjective symptoms of nicotine withdrawal. Ishiguro and colleagues found that the missense polymorphism Pro899Leu in *DAGLA*(one of the genes coding for DAGL) was associated to alcoholism in a Japanese population(133). Considering the issues with candidate gene studies outlined above, all of the above results must be interpreted with caution, considering they use candidate gene methods and have yet to be consistently replicated.

Polygenic risk scores: A reliable alternative to candidate gene methods

With the rapidly decreasing cost of genome sequencing, the feasibility of studying genetic markers as they relate to psychiatric phenotypes – through GWAS – has increased substantially. Though modern GWAS studies employ a wide-variety of genomic methods (including gene-set analysis, natural selection analysis, and fine mapping approaches) to analyze and understand the relationships between genetics and complex diseases, it is the use and manipulation of the GWAS results for other predictive analyzes that are of interest to this work. Due to the open-source nature of psychiatric GWAS results, thanks to the collaboration of researchers through the Psychiatric Genetic Consortium(134), researchers around the world could download the summary statistics of GWAS and apply higher order statistical techniques to better understand how the

genome interacts with the environment in the development of psychiatric disease; one such technique being the polygenic risk score.

While the results of GWAS studies are robust, and replicated across studies, the variants identified typically have small effects, and represent only a small fraction of the truly associated variants(135)(136)(137). Bearing this in mind, Yang and colleagues demonstrated that by aggregating the effects of SNPs identified in GWAS, and combining them into a score – the polygenic risk score (PRS) – one can explain larger proportions of the heritability of complex traits, such as height (138). Accordingly, a variety of polygenic risk calculation techniques were developed to estimate genetic liability to a trait at the individual level (100).

In general, polygenic risk scores are created by taking the sum of the number of risk alleles of an individual, weighted by the effect size of the risk allele as estimated by the GWAS of the phenotype(135). Moreover, by varying the number of SNPs included in the PRS (i.e. using a priori p-value thresholds) one can increase the predictive ability of the PRS (135). Although polygenic risk scores for psychiatric diseases can currently only account for small portions of the variance of said disease (approximately <10% (140)), they remain flexible proxies of an individual's vulnerability to a trait, and can be used to identify shared genetic etiology of various phenotypes (i.e PRS for depression predicts suicide attempts(141)). As such, the study of the summary statistics from the GWAS studies of schizophrenia(142), psychotic-like experiences(143), and cannabis use(144), have been used to elucidate the relationship between psychosis and cannabis use.

In a recent article focusing on genetically informed methods, Gillespie and Kendler outlined three different hypothesis which could explain the cannabis-psychosis association: (1) the relationship is fully causal, i.e cannabis use causes schizophrenia, (2) the relationship may be partially confounded by shared genetic and environmental confounders and/or reverse causation (i.e. that genetic risk for schizophrenia may increase risk for cannabis use rather than cannabis use causing schizophrenia), (3) this link is entirely non-causal (145). Thus, researchers have used the PRS for schizophrenia (PRS-Sz), which was derived to summarize the contribution of variants consistently related to schizophrenia risk(142), to understand these associations, with most

studies reporting a positive significant correlation between PRS-Sz and cannabis use (146)(147)(148)(149). Moreover, recent work has shown that there is strong genetic overlap between schizophrenia and cannabis use (144) and between psychotic experiences and risk for schizophrenia (150). From these studies, some researchers have concluded that the PRS-Sz and cannabis association represents a pathway from genetic risk for schizophrenia to cannabis use (149), while others have suggested that the sensitivity to exposure to cannabis use is moderated by this genetic risk for schizophrenia(151). Contrary to these reports, one highly powered study demonstrated that PRS-Sz was not associated to cannabis use disorder in healthy controls, or patients with psychiatric disorders other than schizophrenia(152). Additionally, the authors show that the association between prior cannabis use disorder and later development for schizophrenia was not altered after adjustment for PRS-Sz and PRS of other psychiatric disorders(152). Thus Hjorthøj and colleagues argue that the association between cannabis use and development of schizophrenia is not explained by common genetic vulnerability (152). Nevertheless, the majority of the evidence studying polygenic risk scores, and other similar methods such as linkage disequilibrium score regression (153)(154), suggest that significant genetic confounding exists, i.e. that common genetic risk factors increase the probability of using cannabis and schizophrenia, which is inconsistent with the fully causal, first hypothesis (145).

The PRS-Sz is also related to PLE, but the results of these studies are less consistent. Initial reports did not demonstrate an association between PRS-Sz and the positive symptoms (hallucinations, paranoia, thought disturbance) of PLE in the general population (155)(156). Yet, more recent work has demonstrated an association between PLE and PRS-Sz. Specifically studies have demonstrated a significant association between the negative (157)(158), positive symptom domains (159)(160).

Other genetic modalities have also been used to study the cannabis-psychosis association. For example, discordant relative designs allow for strong, yet indirect, statistical control of genetic confounding for specific outcomes. Discordant relative designs are studies in which one sibling or family member is exposed to the variable of interest, while the other is not(161). This design allows the researchers to study the exposure-outcome association, all the while controlling for genetic and shared environmental factors(161). Some studies are considered “co-relative control

designs", which include relatives of varying degrees (siblings, half-siblings, first-cousin pairs), which in turn would allow for varying amounts of shared genetic and environmental relatedness across pairs (161). In the only study(161) examining the relationship between cannabis use and schizophrenia using a discordant co-relative control design, the authors reported that although the association between cannabis use and schizophrenia was clinically significant, the relationship decreased as genetic relationship increased: general population OR=10.44, discordant-monozygotic twins OR=3.38. This result supports the second hypothesis outlined by Gillespie and Kendler (145) (i.e. the cannabis-psychosis relationship is partially confounded by genetic and/or environmental confounders and/or reverse causation). This result, along with the results of other studies employing instrumental-variables to infer causal relationships – such as mendelian-randomization studies (144)(162) – support the notion that the epidemiological relationship between schizophrenia and cannabis use may actually be confounded by shared genetic markers and reverse causation (145) .

The strong epidemiological evidence, along with the results from polygenic-risk score studies and from other genomic designs suggest that the while the cannabis-psychosis association may partially causal, it is also partially confounded by shared genetic and environmental factors, as well as possibly being confounded by some reverse causation processes (145). Considering that the purported risks for adverse effects of cannabis are largely increased during adolescence, and that PLE are strong markers for psychosis risk in adolescent populations(100), examining the link between all three risk factors (polygenic risk, cannabis use, and PLE), is an interesting avenue of study. Specifically, using polygenic risk scores, and a longitudinal design, one could study if cannabis use partially mediates or moderates the relationship between polygenic risk for schizophrenia and psychotic-like experiences.

Objectives and hypotheses

Previous candidate gene studies were limited by cohort size (typically N<500), and the high cost of genotyping a wide array of genomic markers. As such, past candidate gene efforts were often limited by the study of a constrained number of polymorphisms within biological systems of interest. Due to the inability to replicate findings of said candidate gene studies through meta-analysis, some authors have called for an end to hypothesis-testing approaches in psychiatric genetics(163). Nonetheless, others have argued that the failure to replicate candidate gene findings do not necessarily suggest the findings are false, rather significant candidate gene findings, in adequately powered cohorts, could represent particular phenotypes of sub-groups of individuals, which could account for a small portion of the genetic effect on the phenotype in question(164). In addition, the increased affordability of genome-wide arrays, larger collaborative cohort studies, and the efficiency of new data analysis platforms, have enabled researchers to study the aggregate effects genetic changes across entire genes, biological systems and the genome as a whole(165). This is important because the study of the effects of multiple genes not only reduces the likelihood of false negative results, by reducing the number of statistical tests performed, but is also more aligned with the polygenic nature of most complex behaviours (132)(166). One recent example of the utilization of a biological systems level genomic approach (rather than a genome-wide approach) in the field of addiction psychiatry, is the set-based analysis used by Carey and colleagues(132). In this study, the group employed a set-based test to examine whether genetic variation in multiple endocannabinoid genes and childhood sexual abuse interact to predict cannabis dependence symptoms(132). At the gene level, they reported a significant interaction between the *MGLL* gene and childhood sexual abuse to predict cannabis dependence (p=0.009). This effect seems to have been driven by one SNP (rs604300) within the *MGLL* gene(132). This study, like others utilizing similar gene-set approaches applied to the study of addictive disorders(167), support the use of candidate gene studies to help identify putative mechanisms, or biological markers, involved in the development of psychiatric and substance use disorders.

Thus, I completed two studies examining different genomic approaches (i.e., biological systems approach and genome-wide approaches) in the aim of understanding the polygenic nature of the development of psychopathologies in the context of substance use in adolescent and young adult populations. While the substance use outcomes vary greatly in the two studies, these works remain related considering the prevalence of alcohol use, and cannabis use across ages, and the relationships between the endocannabinoid system, alcohol use, cannabis use and the development of psychopathologies (i.e. role of endocannabinoid system in development of substance use disorders, relationships between cannabis use and increased burden of alcohol use and alcohol use disorder, relationship between cannabis use and development of other psychopathologies such as psychotic disorders).

Study one: Considering the previous literature associating the endocannabinoid genes to substance use disorder, and the relationship between the endocannabinoid system and the neurobiological systems associated to substance use, I hypothesize that genetic markers within the endocannabinoid related genes are predictive of alcohol use disorder in adolescents. This study is novel in that it is the first study, to our knowledge, to use gene-set analyses to examine the relationship between problematic alcohol use in adolescent populations and the genomics of the endocannabinoid system as a whole.

Study two: Due to the fact that it has not yet been determined if the commonly reported cannabis-schizophrenia relationship is determined by genetic risk for psychotic disorders or through other environmental mediators, I examined if the relationship between polygenic risk score for schizophrenia (PRS-Sz) and psychotic like experiences (PLE) is at least partially mediated by cannabis use, in two independent cohorts of European individuals. Additionally, we also test a moderation hypothesis, in which cannabis might exacerbate genetic vulnerability to psychosis. This is contrasted with a null hypothesis that suggests that despite any potential shared genetic vulnerability between cannabis use and schizophrenia, the relationship between cannabis use and psychotic-like experiences is independent of (or in addition to) the relationship between polygenic risk for schizophrenia and psychotic-like experiences.

Methods

Participants

Study 1

IMAGEN

Data from the IMAGEN cohort, which is a longitudinal imaging genetics study of over healthy adolescents, mostly of European descent, was used for primary analysis in this study(168). The IMAGEN cohort has been repeatedly assessed on various cognitive, psychosocial, and substance use outcomes at 14, 16, 18 and 21 years of age. Moreover, participants contributed genetic data at 14, and underwent anatomic and functional neuroimaging batteries at the various follow ups. The IMAGEN project had obtained ethical approval by the local ethics committees (at their respective sites) and written informed consent from all participants and their legal guardians. All datasets were de-identified by using codes for individuals. See Schumann et al.(168), for a more detailed description of the IMAGEN cohort.

Study 1 used data for all 2087 individuals who completed the IMAGEN battery at 14, 16 or 18 years of age, and who contributed genetic data at 14. Considering the use of sex as a covariate, 3 individuals were removed because of an unassigned sex after sex determination analysis in PLINK1.9(169). Moreover, 11 the individuals who did not complete the Alcohol Use Disorder Identification Test (AUDIT) were removed from the study. 11 pairs of siblings are found among the IMAGEN participants. As such, one sibling from each pair (random assignation) was removed, to reduce potential bias in estimation of standard errors of SNP effect sizes(170). We asses for European ancestry using Admixture, using HapMap III (171) as a reference population. Eleven individuals, with non-European ancestry were removed prior to analysis. Ancestry is controlled because allele frequencies differ between subpopulations, and the presence of multiple subpopulations can therefore lead to false positive or false negative results(170)(172). Overall, in this study there was a total of 1043 female and 1008 males.

Saguenay Youth Survey

Genetic and alcohol-use data from the Saguenay Youth Study (SYS) were used as an independent replication cohort in study 1. The SYS is a two-generational study comprised of 1029 adolescents and 962 parents(173). Participants in this study were recruited over a 10-year period. After recruitment, adolescents provided genetic material and underwent a detailed neurocognitive and psychosocial assessment. Data for the 772 adolescents, aged 14 or older, who had contributed genetic data and completed SUD assessment were included in this study.

Study 2

IMAGEN

Data from the IMAGEN cohort, described above, was used for primary analysis in study two. Data from all 2087 individuals who completed the IMAGEN study and provided genetic data at 14 years of age were included in this study. After genetic quality control, described below, 1740 individuals remained for statistical and genetic analysis.

Utrecht cannabis cohort

Data from the Utrecht cannabis cohort, used in study 2, comes from a subset (N=1223) of data from a cohort (N=17,698) of young Dutch participants. This subset is comprised of individuals who had contributed genetic, cannabis use and PLE data was available. Participants gave online informed consent, and the data collection methods of the Utrecht cannabis cohort received approval by the University Medical Centre Utrecht medical ethical commission. This is a particular cohort due to the enriched nature of PLE and cannabis use data. To increase power for gene x environment interactions in previous studies(174), data from individuals from the general population was combined with data of participants selected from the top or bottom quintile of total PLE scores, who are either non-users(<2 lifetime exposures to cannabis) or heavy users (i.e. current expenditure for personal cannabis use exceeded €10 weekly).

Genetic data

Quality control

STUDY 1

IMAGEN

The genotyping for the IMAGEN cohort was run using the Illumina Quad 610 chip and 660Wq at the "Centre National de Genotypage" (Paris, France). First, quality control (QC) of individuals is performed, by removing individuals with ambiguous sex, call rates of >98%, related (identity by state analysis), and non-European individuals. Non-Europeans are removed following methods published by Huguet et al.(175). Next, SNPs within ±10kb of the genes of interest (*CNR1, NAPE, FAAH, MGLL, DAGLA),* with a minor allele frequency (MAF) of > 5%, a call SNP rate of 90%, and respecting Hardy Weinberg Equilibrium (HWE) ($<1x10^{-6}$), were kept for the study. Following quality control steps, the first study analyzed data from 69 SNPs and 2051 individuals. All QC methods are run in PLINK1.9(169).

SYS

All individuals in this dataset were genotyped using whole blood samples. Genotyping was executed at either the “Centre Nationale de Génotypage” (610Kq array; 599 arrays) or at the Genome Analysis Centre of Helmholtz Zentrum München (HOE-V12 array; 1395 arrays). We use genetic data that was imputed using previously published methods(173), after which data for the 69 SNPs identified in the IMAGEN study was extracted. Detailed descriptions of the cohort, genotyping and data collection have previously been published(176)(177).

STUDY 2

IMAGEN

In study two we exclude SNPs with a minor allele frequency (MAF) of less than 2%, a genotyping rate of 2% or SNPs that did not respect Hardy Weinberg Equilibrium (HWE) ($<1x10^{-6}$). Individuals with disproportionate levels of individual missingness (<2%), ambiguous sex, evidence of cryptic relatedness (>0.125), excessive heterozygosity were removed. After the first steps of

quality control, 1950 individuals of the IMAGEN cohort remained. The data of the IMAGEN cohort were then combined with data from HapMap III, and principal component analysis was performed in PLINK1.9(169) to determine ancestry information. We removed individuals who did not fall within 3SD of the mean of the first 2 principal components of the CEU + TSI populations from HapMapIII(171). In all, 1740 individuals and 488426 SNPs remained for polygenic risk score and regression analysis.

Utrecht cannabis cohort

The genotyping in this cohort was run using either the Illumina® HumanOmniExpress (733,202 SNPs; 576 individuals) or the Illumina® Human610-Quad Beadchip (620,901 SNPs; 768 individuals). Detailed descriptions of the quality control and imputation methods for this cohort were previously published(178).

Phenotypes evaluated

STUDY 1

Alcohol misuse

IMAGEN

The Alcohol Use Disorder Identification Test (AUDIT), is used as a measure for alcohol use disorder. The AUDIT is a validated self-report questionnaire, developed by the World Health Organization to screen for heavy drinking and current alcohol dependence(179). Individuals were considered to screen positive for an AUD, if they scored 8 or more on the AUDIT. While other studies focusing on adolescent alcohol abuse report using a less stringent cut-off (180),(181),(182), or differing cut-off depending on sex(183), a score of 8 or more was used as this is the cut-off with the strongest sensitivity, and a favorable specificity across all studies(184). Four different “case/control” statuses were derived: “ALL” representing having screened positive for AUD at any timepoint from 14-18 years of age, then a score for each of the reassessments, 14, 16 and 18 years.

SYS

The alcohol-use data used for study 1, were obtained via a self-report questionnaire developed specifically for the SYS cohort (185). The items from this questionnaire that were deemed to overlap sufficiently with AUDIT questions were used for this study (see Annex 1 – Supplementary Table 4, for overlapping questions).

Covariates

In this study, for logistic regressions, we include sex, the first 6 principal components, parental alcohol abuse, and parental education as co-variables (see *Annex 1 - Figure S1 for results of principal component analysis)*. Data for parental educational is taken from the relevant questions in the European School Survey Project on Alcohol and Other Drugs (ESPAD) (186). AUD in parents was measured using the AUDIT for the first 2 follow ups. If ESPAD or AUDIT data was missing for parents at 18 years old, the most complete and recent information was used. In post-hoc analysis, for SNPs that significantly predicted AUDIT flags, after controlling for covariates, we control for potential confounding of interaction(187), and include the interaction of the covariate of no interest by SNP. The set-based test, Fisher exact test, and logistic analyses were all carried out using PLINK program(169).

STUDY 2

Cannabis use

IMAGEN

Participants in the IMAGEN cohort were assessed for cannabis use at 14, 16, 18 and 21 years of age using the ESPAD questionnaire. ESPAD is a widely used self-report questionnaire that measures use of various drugs of abuse (186) (188). In this study cannabis use data from the 16-year-old follow up is taken from responses to the question, “On how many occasions in your whole lifetime have you used marijuana (grass, pot) or hashish (hash, hash oil)?”. Answers are scored on a scale from 0-6: ‘0’ = 0, ‘1-2 times’=1, ‘3-5 times’=2, ‘6-9 times’=3, ‘10-19 times’=4, ’20-39 times’=5, ‘40 or more times’=6. To compare data directly to the replication cohort,

cannabis use is then dichotomized into “case/control” status, where >=10 lifetime uses is considered a case.

Utrecht cannabis cohort

Cannabis use data for this cohort is taken from a self-report questionnaire, developed for this cohort. Data for lifetime cannabis use is categorized as never = 0, ‘1 time’ = 1, ‘2 times’ = 2, ‘5-9 times’ = 3, ‘>=10 times’ =4. Data is then dichotomized, to allow direct comparison to the IMAGEN data set, with >= 10 lifetime uses considered as cases.

Psychotic-like experiences

PLE data for both IMAGEN and Utrecht cannabis cohorts is taken from the Community Assessment of Psychic Experiences - 42 (CAPE-42) questionnaire (99). CAPE-42 is a reliable, and validated, self-report questionnaire that measures lifetime PLE (189). Moreover, this questionnaire was designed to have three subscales, which capture three dimensions of PLE: positive, negative and depressive symptoms. The CAPE-42 questionnaire is mainly derived from the Peters et al. Delusions Inventory (PDI-21)(190), which is a dimensional questionnaire developed to measure delusional ideation in the general population(99). Thus, the CAPE-42 was developed through the addition of questions on auditory hallucinations, negative symptoms and depressive symptoms. Negative symptoms were taken from the highly cited SANS questionnaire(191). Due to the difficulty to discriminate between depressive and negative symptoms, only questions that allowed for more reliable discrimination between depressive and negative symptoms (i.e. questions of sadness, pessimism, hopelessness, feeling a failure, and feeling guilt) were included into the CAPE-42(192).

The CAPE-42 questionnaire seems to be both internally reliable and factorially valid. Meta-analysis of the psychometric properties of the CAPE-42 questionnaire demonstrated high internal reliability(189). The meta-analytic mean of the Cronbach’s alpha scores reported in the 5 studies analyzed was 0.91 (SD=0.05) for the CAPE-42 questionnaire as a whole(189). Moreover, it should be mentioned the positive and negative subscales are more reliable in adolescent populations than adult population, while the depressive subscale is equally reliable across studies (189). A review of studies using confirmatory factor analysis has demonstrated that the purported three-

factor model of the CAPE-42 (positive, negative and depressive subscales) is acceptable in the original Greek version(189). However, results of studies which conducted confirmatory factor analysis in translated versions of the CAPE-42 demonstrated that the 3-model fit was not "optimal"(189). For example, in a validation study of the English and French translation of CAPE-42, the three-factor model did not meet optimal goodness of fit indices on all fit indexes (GFI, AGFI, CFI <0.9, and RMR <0.05)(193). Nevertheless, the authors note the since the goodness of fit index was near 0.9 (GFI=0.862) it was within an acceptable range (193). Thus, while it is generally accepted that the CAPE-42 questionnaire has a factor structure which contains three dimensions (positive, negative and depressive symptoms), potential cultural and language differences should be considered when interpreting results from studies using the CAPE-42(189).

In this study, we use only the frequency of symptom measured, as distress caused by symptoms is highly correlated to frequency of symptoms in the IMAGEN cohort (Pearson r=0.86). For primary analysis of this study, the logarithm of sum total frequency scores are used, while the log of the sum of the various-sub domains are analyzed in post-hoc analyses. In IMAGEN data from the 18-year-old follow up is used.

Polygenic risk score construction

Polygenic risk scores for schizophrenia (PRS-Sz) are constructed for each of the IMAGEN and Utrecht cannabis cohort individuals, who passed genetic QC, as described above. The base dataset is taken from the most recent schizophrenia GWAS based on 40 675 cases and 64 643 controls (139) as a base set. Base and target (IMAGEN + Utrecht cannabis cohort) datasets underwent stringent quality control prior to PRS calculation. Target quality control has been described above. SNPs included in the construction of the PRS are those that overlap with the 1000 Genome reference set, target and base sets. Then SNPs with a MAF of <0.01 and SNP impute quality of <0.8 are excluded from base data set. Next, SNPS in complex LD-regions are removed (see Annex 2 – table S1). Using standard clumping procedures, PRS are then calculated using PRSice2(194). For each cohort, two PRS are retained for primary analyses. First, for each individual, a PRS-Sz is built using the p-value threshold (p_t) of <0.05, as this is the PRS-SZ that optimally captures phenotypic variance in schizophrenia(142). Next, we construct 12 PRS-Sz for

each individual (P_t= $5x10^{-8}$,$5X10^{-7}$, $5X10^{-6}$, $5X10^{-5}$, $5X10^{-4}$, $5X10^{-3}$, $5X10^{-2}$,0.1,0.2,0.3,0.4,0.5), to capture the PRS which explains the most variance (Nagelkerke's R^2) in the logistic regression between PRS-Sz and cannabis use. To control for type-1 error, we obtain an empirical p-value 10000. Furthermore, the replication of the PRS result in an independent and similarly powered cohort would further confirm the validity of the result. Therefore, the same protocol is used to create PRS, and dichotomize phenotype measures, to ensure comparability between cohorts. Finally, we control for age (in years) in regressions using the Utrecht cannabis cohort data, as this study is cross sectional and includes participants with a wider age range. After PRS calculation, PRS-Sz are center scaled using the scale function in R(195) to ease interpretability of the results.

Statistical analyses

Study 1

After quality control, 69 SNPs located within or near 5 endocannabinoid related genes are studied in relation to AUDIT scores. Set based test, Fisher exact test, and logistic regression are all run in PLINK1.9(169).

Set-based test

69 SNPs within the *CNR1, NAPEPLE, FAAH, MGLL, DAGLA* genes are included for gene-set based tests, which are run in the PLINK1.9 program. Considering that no "standard" parameters have been accepted in the literature, at the time of this study, we conduct three set-based tests while varying parameters across each test. The parameters that are varied across tests are p-value for significant variants, r^2 of variant pairs, and maximum set size. Results for each test underwent 10,000 label swapped permutations. The parameters for the first test, for which results are considered in subsequent analyzes, include a p-value of 0.05, r^2 of 0.5, and set-max of 5 SNPs. The second and third are more stringent with p-value of 0.05 and 0.01, r^2 of 0.3 and 0.1, and set max of 3 and 2 respectively. The second and third test were run as a form of sensitivity analyses. Statistical significance for set-based test was corrected for multiple testing (5 gene sets) and maintained at an empirical p-value of <0.01. SNPs that were independent (based on r^2) and

nominally significant ($p<0.05$), were then analyzed in subsequent case-control and logistic analyzes.

Case-control analysis and logistic regression

Significant and independent SNPs were considering for Fisher's exact case-control analysis. False-Discovery rate (FDR) was used to correct for multiple tests in case-control analysis. SNPs that survived correction were included in logistic regression models as independent variables. In logistic regression we include all co-variates described above. As a final post-hoc analysis, we control for potential confounding of interaction of covariates of no interest by SNP of interest(187).

Generalized multifactor dimensionality reduction (GMDR)

We utilize a generalized multifactor dimensionality reduction analysis to study how SNPxSNP interactions across the endocannabinoid system relate to AUDIT scores. This analysis is run using the GMDR(v1.0) tool, which is an open source tool readily available online(196). We screen for the best interactions across the 69 SNPs and AUDIT flag, and analyze one SNP, two SNP and three SNP models. 10,000 label swapped permutations are used to obtain an empirical corrected p-value. For more information on the GMDR method see Lou et al, 2007(196).

Study 2

Regression analysis

Multimonial linear and logistic regression are used to study the association between PRS-Sz, cannabis use and PLEs. In our primary analysis, the PRS-Sz that best predicts cannabis use, and well of the PRS-Sz that best predicts schizophrenia ($P_t<0.05$) are used as independent variables to predict PLE. We use the log-transformed sum of CAPE-42 score as the dependent variable in primary analyses. For analyses the first six principal components and sex are used as covariates. We replicate these results using data from the Utrecht cannabis cohort cohort, where we add age as a covariable. A p-value of 0.025 is considered as significant, to account for multiple testing (two PRS-Sz). For models using dichotomous dependent variables Nagelkerke's R^2 is used to determine variance explained. As a post-hoc analysis, we conduct similar regression analyses where the

various sub-domains of the CAPE-42 (positive, negative and depressive symptoms) are considered as dependent variables. Finally, as a form of sensitivity analyses, we use PRS-Sz with more stringent P_t ($5X10^{-8}$, $5X10^{-5}$). We also control for genetic risk for cannabis use, by including a PRS for lifetime cannabis use – derived from a recent lifetime cannabis use GWAS(144) – as a covariate in sensitivity analyses. All regression analyses are run in R(195).

Mediation and moderation analysis

In our model, we hypothesize that cannabis use (M), mediates the relationship between PRS-Sz (independent variable; IV) and PLE (dependent variable; DV). We run all mediation analyses using the '*mediation'* package in R(197). Considering that the Utrecht data is cross-sectional, we only run mediation analyses with the IMAGEN data, to infer directionality of the result. Thus, in mediation analyses, sex and the first 6 principal genetic components are considered as covariables. A non-parametric bootstapping (n=10000) procedure is used to estimate sampling distribution of the indirect effect (mediation effect). If the confidence interval of the indirect effect does not cross zero, we consider the average causal mediation effect (ACME; or indirect effect) to be significant.

To examine if cannabis use increases risk for PLE in those with high genetic risk for schizophrenia, a moderation analysis, using the '*MeMoBootR'* package in R. This package "a complete two-way moderation analysis with one moderator, similar to model 1 in PROCESS by A. Hayes (2013)". In this model cannabis use is hypothesised to moderate the effect of PRS-Sz on PLE. Again, the model is run for both IMAGEN and Utrecht cannabis cohorts. Covariates for these analyses are sex, the first 6 principal components and age (in the Utrecht cannabis cohort analysis).

Article 1: Endocannabinoid gene x gene interaction association to Alcohol Use Disorder in two adolescent cohorts

Manuscript status: **Published, Frontiers in Psychiatry (April, 2021)**
Manuscript has been modified to fit stylistically with the present work.

Laurent Elkrief [1,2], Sean Spinney [2, 3], Daniel E. Vosberg [4], Tobias Banaschewski M.D., Ph.D.[5]; Arun L.W. Bokde Ph.D.[6]; Erin Burke Quinlan PhD[7]; Sylvane Desrivières Ph.D.[7]; Herta Flor Ph.D.[8,9]; Hugh Garavan Ph.D.[10]; Penny Gowland Ph.D.[11]; Andreas Heinz M.D., Ph.D.[12,13]; Rüdiger Brühl Ph.D.[14]; Jean-Luc Martinot M.D., Ph.D.[15], Marie-Laure Paillère Martinot M.D., Ph.D.[16]; Frauke Nees Ph.D.[5,8]; Dimitri Papadopoulos Orfanos Ph.D.[17]; Luise Poustka M.D.[18]; Sarah Hohmann M.D.[5]; Sabina Millenet Dipl.-Psych.[5]; Juliane H. Fröhner MSc[19]; Michael N. Smolka M.D.[19]; Henrik Walter M.D., Ph.D.[12]; Robert Whelan Ph.D.[20]; Gunter Schumann M.D.[7, 21]; Zdenka Pausova[22], Tomáš Paus [4, 23], Guillaume Huguet [2,3]†, Patricia Conrod [2,3,24*]†, and the IMAGEN consortium.

Author Affiliations:
1. Department of Medicine, Université de Montréal, Montreal, Quebec, Canada.
2. Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Quebec, Canada.
3. Department of Pediatrics, Université de Montréal, Montreal, Quebec, Canada
4. Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada.
5. Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany;
6. Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland;
7. Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, United Kingdom;
8. Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;
9. Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany;
10. Departments of Psychiatry and Psychology, University of Vermont, 05405 Burlington, Vermont, USA;
11. Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom;
12. Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charitéplatz 1, Berlin, Germany;
13. Department of Psychiatry and Psychotherapy, Berlin Institute of Health, Campus Charité Mitte, Berlin, Germany
14.Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2 - 12, Berlin, Germany;
15 Institut National de la Santé et de la Recherche Médicale,nINSERM U1299 "Trajectoires développementales en psychiatrie," Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette, France.
16 Institut National de la Santé et de la Recherche Médicale, INSERM U A10 "Trajectoires développementales en psychiatrie," Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli and AP-HP. Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France.
17. Neurospin, Commissariat à l'Energie Atomique, CEA-Saclay Center,Paris,France
18. Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075, Göttingen, Germany;
19. Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany;
20. School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland;
21. PONS Research Group, Dept of Psychiatry and Psychotherapy, Campus Charite Mitte, Humboldt University, Berlin and Leibniz Institute for Neurobiology, Magdeburg, Germany,
22. Hospital for Sick Children, Departments of Physiology and Nutritional Science
23. Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada
24. Department of Psychiatry, Université de Montréal, Montréal, QC, Canada.

Corresponding author contact information:
Patricia Conrod, Université de Montréal and CHU Ste Justine, Centre de Recherche, 3175 Ch Cote Ste-Catherine, Montreal, QC H3T1C5, patricia.conrod@umontreal.ca

Key Words: AUD. CNR1. DAGL. Endocannabinoid System. MGLL.

Author contributions

Laurent Elkrief conceptualized the analysis, ran the analysis using PLINK1.9, wrote and edited the manuscript. Sean Spinney helped conceptualize the analysis and ran analyses using the PLINK1.9. Daniel E. Vosberg compiled the SYS cohort data and edited the manuscript. Tomáš Paus helped secure access to the SYS and IMAGEN data, edited the manuscript and supervised the work. Patricia Conrod and Guillaume Huguet supervised the project, helped conceptualize the project, secured access to the data and edited the manuscript.

Abstract

Genetic markers of endocannabinoid system have been linked to a variety of addiction-related behaviours that extend beyond cannabis use. In the current study we investigate the relationship between endocannabinoid (eCB) genetic markers and alcohol use disorder (AUD) in European adolescents (14-18 years old), followed in the IMAGEN study (n=2051) and explore replication in a cohort of North American adolescents from Canadian Saguenay Youth Study (SYS) (n=772). Case-Control status is represented by a score of more than 7 on the Alcohol Use Disorder Identification Test (AUDIT). First a set-based test method was used to examine if a relationship between the eCB system and AUDIT case/control status exists, at the gene level. Using only SNPs that are both independent and significantly associated to case-control status, we perform Fisher's exact test to determine SNP level odds ratios in relation to case-control status and then perform logistic regressions as post-hoc analysis, while considering various covariates. Generalized multifactor dimensionality reduction (GMDR) was used to analyze the most robust SNP x SNP interaction of the 5 eCB genes with positive AUDIT screen. While no gene-sets were significantly associated to AUDIT scores after correction for multiple tests, in the case/control analysis, 7 SNPs were significantly associated with AUDIT scores of > 7 ($p<0.05$; OR<1). Two SNPs remain significant after correction by false discovery rate (FDR); rs9343525 in *CNR1* ($p_{corrected}=0.042$, OR=0.73) and rs507961 in *MGLL* ($p_{corrected}=0.043$, OR=0.78). Logistic regression showed that both rs9353525 (*CNR1)* and rs507961 (*MGLL*) remained significantly associated with positive AUDIT screens ($p<0.01$; OR<1), after correction for multiple covariables and interaction of covariable x SNP. This result was not replicated in the SYS cohort. The GMDR model revealed a significant three-SNP interaction (p= 0.006) involving rs484061 (*MGLL*), rs4963307 (*DAGLA*) and rs7766029 (*CNR1*) predicted case-control status, after correcting for multiple covariables in the IMAGEN sample. A binomial logistic regression of the combination of these three SNPs by phenotype in the SYS cohort showed a result in the same direction as seen in the IMAGEN cohort (BETA=0.501, p=0.06). While preliminary, the present study suggests that the eCB system may play a role in the development of AUD in adolescents.

Introduction

Substance use disorders are a growing concern across the world, with an estimated 31 million users worldwide suffering from drug use disorders. After alcohol and tobacco, cannabis ranks as the most used drug worldwide(198). Moreover, those who use cannabis are more than five times more likely to have an alcohol use disorder (AUD) (65). Considering that the endocannabinoid (eCB) system is responsible for the physiological consequence and subjective "high" of cannabis, much attention has been paid to the eCBs role in the development of various substance use disorders. Cannabinoid receptors and related enzymes are expressed in many of the reward centers of the brain; nucleus accumbens (NAc), ventral tegmental area (VTA), amygdala, and basal nucleus of the stria terminalis (BNST)(6)(41). These eCB levels are affected by ethanol(52), and the eCB system plays a role in the development of AUD and other substance use disorders in humans(41). Basavarajappa and colleagues (53) demonstrated that that acute ethanol use has been associated with an increase in eCB signaling, while others have reported that alcohol use decreases eCB signaling(54)(55). Moreover, as is the case with other drugs of abuse, eCBs mediate the reward signals associated with alcohol use(56). Overall, the underlying evidence shows that the eCB system is modulated by ethanol use, and this same system may play an independent role in AUD(199).

The first eCB receptor isolated, of which tetrahydrocannabinol (THC) is also a ligand, is the cannabinoid receptor one (CB1)(8)(200). Binding to this receptor and a second cannabinoid receptor (CB2) are the two main eCB agonists anandamide (AEA) and 2-arachidonoylglycerol (2-AG). These agonists - which are not stored in vesicles - are produced through an enzymatic cascade in a Ca^{22++} dependent manner, and then are rapidly degraded by specific enzymes (fatty acid amide hydrolase (FAAH), and monoacylglycerol lipase (MAGL)). N-acylphosphatidylethanolamine-specific phospholipase D (NAPE-PLD), plays a crucial role in the synthesis of AEA, which then binds to CB1. 2-AG is synthesized by diaglycerol lipase (DAGL).

It has been shown that polymorphisms in the *CNR1* gene, the gene coding for the CB1 receptor protein, are associated with a range of diseases, psychiatric disorders and substance use (115)(116)(117). Many studies have assessed the various aspects of the eCB genes and their

relationship with substance use disorders and risk-taking behaviour. The single nucleotide polymorphism (SNP) rs1049353, in the *CNR1* gene has been associated with severe alcoholism (minor A allele)(118), heroin addiction (major G allele)(120), and impulsivity(119). Furthermore, haplotype blocks within the *CNR1* gene have been associated with addiction and addictive behaviour(121)(123). Polymorphisms in the *FAAH* gene have also been associated with problem drug use and addiction(126)(129). In contrast, there have been relatively few studies examining the *MGLL* gene, the gene coding for the MAGL enzyme, and the *DAGL* in association to drug dependence (130)(131)(132). Among these, only one study has found a positive association between SNPs of the *MGLL* gene and drug dependence(132), while no studies have reported a significant association between *DAGL* and any form of drug abuse. Moreover, many of the original findings reporting an association between SNPs located in genes of the eCB and various drug abuse behaviours have not been replicated(41),(201), suggesting the possibility of false positive results in these candidate gene approaches. Nevertheless, while there are conflicting results among studies, the candidate gene literature suggests that genes related eCB proteins may play a role in the development of substance use problems.

While candidate gene findings in psychiatric genetics have been widely criticized for replication failure, particularly with respect to GWAS and meta-analysis(163), candidate gene approaches in addiction research have identified genetic markers that have been confirmed in GWAS and meta-analysis(114). This is perhaps related to particularly heritable nature of addictive behaviours compared to other psychiatric conditions, or to the fact that candidate gene approaches can be directly informed by pharmacogenetic studies on how drugs of abuse interact with the brain's neurochemistry. Others have argued(164), the failure to replicate candidate gene findings through GWAS and meta-analysis do not necessarily suggest the findings are false. The candidate gene findings may represent particular endophenotypes of sub-populations, which may account for a portion, albeit small, of genetic influence on the phenotype in question. Thus, other groups have utilized novel methodologies, such as gene-set approaches, to analyze hypothesis-based questions in psychiatric genetics and addiction medicine. Recently, one group – utilizing said gene-set approaches found that *MGLL* and the SNP rs604300 interact with childhood sexual abuse to predict cannabis dependence symptoms(132). Considering our

relatively limited understanding of the roles of the various endocannabinoid genes in the pathogenesis of addictive behaviours, and the lack of robust findings at the individual SNP, or GWAS levels, gene-set and system-based approaches remain of interest (132),(202). Thus, the current study employs a gene-set based approach in an attempt to shed light on the role of the eCB system in the pathogenesis of addictive behaviours.

Given the effect of alcohol on the eCB system(52) and the purported relationship between eCB SNPs and the risk for substance use disorder, we assessed the association between eCB genetics and alcohol abuse behaviours in the IMAGEN cohort(168). The IMAGEN cohort is a European cohort of 2087 adolescents, recruited in France, UK, Ireland and Germany. Endocannabinoid genetic influence was studied through a candidate gene approach. Multiple SNPs in eCB genes that have been previously examined (*CNR1, FAAH, MAGL, DAGLA*) as well as genes that have not yet been investigated (*NAPEPLD*) were analyzed in the context of alcohol use disorder (AUD). To understand this relationship a three-tiered approach was used. First, a set-based test (203) is utilized to study, at the gene-level, the link between the eCB system and alcohol abuse behaviour. Through this approach we also identify SNPs that are significantly and independently associated to positive AUD screening, and these SNPs are selected for further study using a case/control analysis and subsequent logistic regression. Finally, while some studies have investigated the interaction between two eCB genes and addictive behaviour (127),(204), none have examined the eCB system as a whole. Considering the complex interplay between the multiple eCB ligands (AEA and 2-AG among others) and various receptors (CB1, CB2, etc.) in their relationship to addictions (41), we hypothesize that a single genetic marker association study could not account adequately for the multifaceted role the eCB system plays in risk for AUD. A new wave of candidate gene studies have explored more complex gene-gene interactions, using various methods of multifactor dimensionality reductions analyses to yield promising results such as predicting outcomes in breast cancer treatment(205), in determining genetic biomarkers to predict antidepressant response (206), and further understanding the genetic influences of nicotine addiction(207). Here, we utilized Generalized Multifactor Dimensionality Reduction (GMDR) to understand the effects that the multiple eCB genes may have on each other and their combined influence on alcoholic behavior in adolescence. To replicate the results, genetic and

alcohol use data were used from the Saguenay Youth Study (SYS), a two-generational study comprised of 1029 French-Canadian adolescents and their parents.

Materials and methods

Participants

The IMAGEN study is a longitudinal imaging genetics study of 2087 healthy adolescents, mostly of European descent. Detailed descriptions of this study, genotyping procedures, and data collection have previously been published(168). The IMAGEN cohort has been repeatedly assessed on substance use outcomes at 14, 16 and 18 years of age. The multicentric IMAGEN project had obtained ethical approval by the local ethics committees (at their respective sites) and written informed consent from all participants and their legal guardians. The parents and adolescents provided written informed consent and assent, respectively. All datasets were de-identified by using codes for individuals. See Schumann et al.(168), for a more detailed description of the IMAGEN cohort.

The current study used data for all 2087 individuals who completed the IMAGEN assessment battery at 14, 16 and 18 years of age, and who contributed their genetic data at 14 years of age. Of those followed at 16 or 18 years of age, 3 individuals had unassigned sex according to sex determination analysis in PLINK1.9(169) and were thus excluded from the genetic analyses. Moreover, eleven individuals did not answer the Alcohol Use Disorder Identification Test (AUDIT) at any time point and were thus removed from the genetic analyses. Eleven pairs of siblings were a part of the IMAGEN database, and thus one sibling from each pair was removed from the study, according to the methods published (175). European ancestry was determined using Admixture (http://www.genetics.ucla.edu/software/admixture) using HapMap III (171) as a reference population. Eleven individuals, with non-European ancestry were removed prior to analysis. Thus, in this study there was a total of 1043 female and 1008 males. A summary of the individuals can be seen in Tables 1 and *S1*.

Phenotype evaluated

Alcohol misuse

AUDIT, is a self-report questionnaire, developed by the World Health Organization, and validated(179), to screen for heavy drinking, and current alcohol dependence. Individuals were considered to screen positive for risk for AUD and were included in the case group, if they scored 8 or more on the AUDIT (case-control status). While other studies focusing on adolescent alcohol abuse used a less stringent cut-off (180)(181)(182), the more stringent cut-off of 8 was chosen as this is the cut-off with the strongest sensitivity, and a favorable specificity across all studies(184)**.** Four AUD scores were derived: “Any AUD” representing having screened positive for AUD at any timepoint from 14-18 years of age, and then individual dichotomized scores for each of the time points, 14, 16 and 18 years. For details about choice of cut-off, see *supplementary methods*.

Covariates

Covariables include sex, the first six genetic principal components, parental alcohol abuse, and parental education as co-variables. Parental education was taken from parent-report questionnaire using the educational categories specified in the European School Survey Project on Alcohol and Other Drugs (ESPAD+) questionnaire. Risk for AUD in parents was measured using the AUDIT obtained at the first two time points in IMAGEN. If ESPAD+ and AUDIT information were missing at the 18-year-old time point, the most complete and recent information was used at this time point. If a parent had signaled a positive AUDIT at any time, they were flagged as such. Moreover, if parental information was missing, individuals were not included in the logistic regression.

Pipeline of SNP Selection

The genotyping was run using the Illumina Quad 610 chip and 660Wq at the "Centre National de Genotypage" (Paris, France). Only autosomal SNPs were kept for this study. SNPs with a minor allele frequency (MAF) of less than 5%, a missing SNP rate of 10% or SNPs that did not respect Hardy Weinberg Equilibrium (HWE) ($<1x10^{-6}$) were also removed from this study. All available SNPs in the genes of interest (*CNR1, NAPE, FAAH, MGLL, DAGLA)* within ±10kb (to

include promoter and flanker regions) were then selected. Gene length, and location were obtained using the UCSC Genome Browser). The SNP coordinates were updated from hg18 to hg19 using Illumina information and the liftover tool from the genome browser. (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Nevertheless, SNP information was scarce on the *CNR2*, as such, the gene was not included in this study. A summary of the locations and details of each SNP (gene, chromosome, base pair, function and etc …) can be seen in *Table S2*.

Statistical analysis

Sixty-nine SNPs appearing across five cannabinoid-related genes were analyzed for their relation to problematic alcohol consumption. As a primary analysis, we first conduct three set-based tests using parameters of varying stringencies, to study the relationship between 5 endocannabinoid gene-sets (*CNR1, NAPEPLE, FAAH, MGLL, DAGLA*). The parameters that were adjusted between the tests were p-value for significant variants between tests, r^2 of variant pairs, and maximum set size. Data in all three set-based tests underwent 10,000 label-swapped permutation as well, using the --perm function in PLINK1.9. The first test was the default test in PLINK1.9, with a p-value of 0.05, r^2 of 0.5, and a set-max of 5, the second test had a p-value of 0.05, r^2 of 0.3, and set-max of 3, while test 3 had a p of 0.01, r^2 of 0.1 and set-max of 2. Tests 2 and 3 were more stringent, and were run to challenge the data, to ensure robustness of our results. Statistical significance for set-based test was determined using a Bonferroni corrected empirical p-values of $p<0.01$ (0.05/5 genes). Burden and optimized sequence kernel association tests (SKAT, R package) (208) were used to analyze the joint effects of SNPs (in gene sets). These analyses were performed on three groups of variants: a) Set 1, b) Set 2, c) Set 3 defined with gene set PLINK analyses. We resampled 10 000 times to compute empirical p-values (p-values were adjusted controlling for family-wise error rate) for the analyses (with “bootstrap” option).

Next, to determine SNP level odds ratios (OR) case-control analysis was run on the SNPs that were nominally significant and independent after set-based analysis, using Fisher’s exact test. In the case-control analysis false discovery rate (FDR) was used to correct for multiple tests. To test the robustness of these findings after controlling for various relevant covariates, a logistic regression was performed which included only the SNPs that remained significant after correction for multiple tests, sex, the first six ancestry components, parental AUDIT flag, and parental

education were included in the logistic model. In post-hoc analysis, for SNPs that significantly predicted case-control status, after controlling for covariates, we control for potential confounding of interaction(187), and include the interaction of the covariate of no interest by SNP (see *supplementary methods* for descriptions of the covariables and *Figure S1 for results of principal component analysis).* The set-based test, Fisher exact test, and logistic analyses were all carried out using PLINK program(169).

Generalized multifactor dimensionality reduction (GMDR)

In order to test the replicability of these findings across a different analytic strategy, GMDR was employed to analyze the SNP x SNP interaction with phenotype. GMDR (v1.0) is a free open source tool for identification of interactions, developed by Guo-Bo Chen (196). This program was used to screen for the best interaction combinations among the 69 SNPs and the phenotype of interest. Permutation with 10 000 shuffles providing empirical p-values to measure the significance of an identified model was used. For these analyses, logistic regression with the same covariables as described above were performed. For more information on the GMDR method see Lou et al, 2007 (196).

SYS replication cohort

Genetic and alcohol-use data from the Saguenay Youth Study (SYS) were used to replicate the findings. The SYS is a two-generational study comprised of 1029 adolescents and 962 parents(173). For descriptive characteristics of the participants included in the replication see *Tables S3, and Table 1.* All individuals were genotyped using whole blood samples from which DNA was extracted. The genotyping was performed at "Centre Nationale de Génotypage" for 610Kq (No. arrays=599) and at the Genome Analysis Centre of Helmholtz Zentrum München (Munich,Germany) for HOE-V12 (No.arrays=1,395). Genetic information was imputed following previously published methods(173) and after that, the 69 SNPs studied were extracted. Detailed descriptions of the cohort, genotyping and data collection have previously been published(176)(177).

Participants were recruited over a 10-year period. Once recruited, adolescents provided genetic material and underwent a detailed assessment in several domains. Alcohol-use data for

the SYS cohort were obtained via a self-report questionnaire developed specifically for the SYS to assess mental health and substance use based on validated protocols (185). The items from this questionnaire that were deemed to overlap sufficiently with AUDIT questions are listed in *Table S4*. Of 1029 adolescents in the SYS cohort, 772 adolescents aged 14 years and older had completed both the SUD assessment and provided genetic information and were therefore included in this study.

In the replication of the case-control study, we studied the 7 SNPs found in the set-based test. Description of SNPs can be found in *Table S5*. Two statistical models were used to study the replication group. To study the native continuous phenotype, a model based on the quasi-poisson distribution was used. The participants were also separated into 4 different drinking groups, based on scoring distribution. A binomial logistic model was then used separating the participants into controls (groups 0 - 1; low alcohol use) and cases (groups 2-4; high alcohol use). Both models considered sex, age as covariables and family ID as random effect. Statistical analyses were performed using R, with the glmmTMB library, version 3.5.3 (https://www.R-project.org/).

Results

Set-based tests: identifying candidate SNP

The three set-based tests were run, with varying results (*Table S6*). In the first set-based test, 9 SNPS returned with nominal p values of <0.05, of which 7 also passed linkage disequilibrium (LD) criterion. Through the first set-test criterion, only the CNR1 gene-set had a significant empirical p-value (p=0.022), but this was not significant after correction for multiple tests. Within this set, only rs9353525 was significantly and independently related to dichotomized AUDIT scores. In the second set-based test, the same 9 SNPS returned with nominal p values of <0.05, of which 5 SNPs passed the LD criterion. Nonetheless, no gene sets were significantly associated to case control status (p>0.05). Finally, 4 SNPs returned with a p value <0.01 in the third test, with 2 SNPs passing LD criterion. No genes remained significant after correction for multiple testing (p_{FDR}>0.05). As mentioned above, the 7 SNPs that had marginal p values of <0.01 in the first set-based test, and that passed LD criterion (r2<0.5) were extracted and only these were analyzed in

the case-control analysis and logistic regression analysis. SKAT demonstrated similar results for the CNR1 gene (*Table S7*)

Case-control analysis and sensitivity analysis

In the case-control analysis of the IMAGEN cohort, which considered cases as individuals who scored eight or more on AUDIT at any time point (ALL), all 7 SNPs analyzed were significant ($p<0.05$) [Table 2]. All of the minor alleles were protective against having a case control status (OR<1). Two SNPs remained significant after correction by FDR; rs9343525 in *CNR1* (p_{FDR}=0.043, OR=0.73) and rs507961 in *MGLL* (p_{FDR}=0.043, OR=0.78). A multivariate logistic regression analysis was done for the two SNPs that were significant after FDR correction in the Fisher test [Table 3]. As a first post-hoc analysis logistic models were done for significant SNPs, at each time point (14, 16 and 18), as well as for any positive screen (ALL) for case-control status. After controlling for the effects of the first six principal components, sex, parental AUDIT scores (at any time) and parental education, both rs9353525 and rs507961 were still significantly associated with positive AUDIT screen in the ALL analysis [Table 3] ($p<0.01$), with both SNPs minor allele acting as protective factors (OR<1). In our post-hoc analysis, we find a significant interaction between rs9353525 and PC1 and PC6, as well as a significant interaction of rs507961 and PC3, suggesting that the genetic background, captured by the principal components, may modify the genetic effects of the SNPs on AUDIT scores. For complete results of logistic regression see *Table S9,* and see supplementary table S10 for results of post-hoc interaction analyses. Finally, we conducted post hoc analyses to study the association between AUDIT scores and SNPs of interest at each IMAGEN time point (14, 16, and 18 alone). After correction for multiple testing, none of the post-hoc analysis demonstrated significant results (see supplementary results for detailed results).

In the replication cohort, rs484061 was significantly associated with problematic alcohol use (p=7.47*10-6) in the binomial model. None of the other SNPs in the replication analysis had a significant result, after correction for multiple tests [*Table S8*].

GMDR: SNP x SNP interactions

A GMDR model was used to screen for the most robust interaction of combinations for the 69 SNPs in the candidate genes and case control status. For the one and two-SNP models, no

significance was found $p>0.05$. However, we found a significant three-SNP model ($p = 0.006$) involving rs484061 (*MGLL*, intron), rs4963307 (*DAGLA*, intron) and rs7766029 (*CNR1*, downstream-gene) with AUDIT positive screens. An interaction between rs484061, rs4963307 and rs7766029 was significantly associated with case-control status, with a combination of G/A;G/A;C/C or G/G;G/G;C/C conferring protection against problem drinking in the cohort ($p=0.004$ and $p=0.02$ respectively *Table S11*). The cross-validation consistency of this three-locus model was 19/20. The testing accuracy of the three SNP model (54%) was greater than the testing accuracy of either the one SNP(49%) or two SNP models (50%) (Table 4);(Figure 1). This result was verified by re-analyzing the model using 10 different random seeds and this model remained significant for each seed. An analysis of the same three SNP combination in the SYS cohort, binomial logistic model, showed a result in the same direction as seen in the IMAGEN cohort ($\beta=0.50$, $p=0.06$), and the distribution of at risk and protective combinations of SNP with phenotype is comparable to that of the IMAGEN population [*Tables S11 and S12]*.

Discussion

Although no gene-sets were significantly predictive of binary AUDIT scores, after correction for multiple tests, our case/control analysis suggest that two SNPs, rs507961 (*MGLL*) and rs9343525 (*CNR1*) are associated with problem drinking and remained significantly associated after correction for multiple tests. The SNPs remained significantly associated to case-control status in logistic regression, while considering multiple covariables, and the interaction of these covariables and the SNPs in question. The results of our logistic regression were not replicated in the replication cohort. To our knowledge, one study (132) had investigated rs507961 in *MGLL* in relation to substance use disorders, however the association did not remain significant after correction for multiple tests. While rs507961 is intronic in *MGLL*, this SNP plays a role in histone regulation of this gene in the brain *(Table S13)*. The robustness of our result confers evidence that carrying the minor T allele may in fact confer protection against problem drinking. Moreover, no study has investigated the relationship between rs484061, another *MGLL* SNP and substance use disorders. The recurrence of rs484061 in both the GMDR model and case-control analyses suggests that being a carrier of this SNP protects against risk for AUD. While rs484061 was significantly associated to positive AUDIT screens in the case-control analysis of the IMAGEN

cohort (p=0.009, pFDR=0.055) and replicated in SYS($p=7.47*10^{-6}$), it was significantly associated to lower alcohol use. Our results suggest a role for *MGLL* in AUD but work in larger cohorts is needed to confirm this result.

The second SNP that remained significant after correction for multiple tests in our case-control analysis was rs9353525. It is localized in an intergenic region less than 10Kb of the 3' region of *CNR1*. In an attempt to understand the biological role that this SNP plays in the regulation of *CNR1* expression, we scanned the various available databases for potential roles, however this SNP is relatively understudied. While this SNP was not associated with higher rates of alcohol use in the SYS cohort, this SNP is in strong linkage disequilibrium with rs806368 (at 78% for allele T with G respectively for rs806368 and rs9353525). The rs806368 has been associated to alcohol, dependence in other studies(209). We also investigated rs806368 in our cohort, using the same case-control analysis as for our other SNPs, and the major allele is associated with likelihood of reporting a clinically significant AUDIT score at any timepoint in the IMAGEN cohort (p=0.007 OR=1.28). Moreover, this result remains significant after controlling for the various covariates described above in the IMAGEN cohort (p=0.007; see *Tables S14 and S15*). Taken together, these results suggest that the haplotype block containing both of the major alleles of rs9353525 and rs806368 plays a role in the development of AUD in adolescents.

A GMDR model was used to screen for the gene x gene interaction that would be most associated to problem drinking, across genes showing a signal in previous analyses. We found a significant interaction involving rs484061 (*MGLL*), rs4963307 (*DAGLA*) and rs7766029 (*CNR1*), that predicted clinically significant AUDIT scores after correction for covariates. Each of these three SNPs are associated to loci, which are key regulators of gene expression (*Tables S13 and S16*). This observation was supported by the consistency of the result in the GMDR, across IMAGEN and SYS GMDR results (p=0.06) *(Table S8)*. The similar distribution pattern of problem drinkers within the SYS cohort suggests that the marginal result in the SYS cohort is probably due to a lack of statistical power. The SYS cohort comprises a relatively young sample (mean age = 15 years old), as compared to the IMAGEN cohort which includes data from individuals when they are 14, 16 and 18 years of age. As such, many of the participants in the SYS cohort have not had their first contact with alcohol, and therefore might not have developed heavy patterns of drinking. This marginal

effect should be investigated using data from this sample as it ages, to explore whether the effect becomes larger and more significant when substance use behaviours are assessed during the typical age when substance use disorders have their onset

Endocannabinoid interactions in the brain and emotional regulation

The GMDR analysis suggests that a certain combination of SNPs along the *CNR1-DAGLA-MGLL* genes protect against or pose a risk for alcoholism, by presumably modulating DAGLA and or MGLL expression and subsequently 2-AG levels. The DAGLA protein (encoded by *DAGLA)* catalyzes the formation of 2-AG, which then acts as an agonist of CB1. 2-AG is then promptly degraded by MAGL (encoded by *MGLL)*. 2-AG has been shown to play a key role in the regulation of the hypothalamic-pituitary-adrenal (HPA) stress response axis (210), which is altered in alcohol addiction (211). In response to increased corticosterone, 2-AG levels increase in the medial prefrontal cortex and paraventricular nucleus of the hypothalamus, and acts as a negative feedback signal to inhibit the HPA axis and terminate the acute stress response (210). While 2-AG levels increase in situations of chronic stress, it is theorized to play a role in stress habituation (210). Along the same line, 2-AG has also been shown to play a role in the reduction of stress induced-anxiety in a role mediated through the actions of MAGL and DAGLA (210). MAGL antagonists have been shown to have a strong anxiolytic effect in rodents (212),(213). Knockout studies have shown that *DAGLA (-/-)* mice, which have large reductions in brain 2-AG levels, have increased anxiety-like symptoms (214),(215). Moreover, the anxiety-like state seen in animal models of alcohol dependence and withdrawal symptoms are mediated by corticosterone-releasing factor release in the central nucleus of the amygdala (CeA)(36). A recent study in alcohol dependent rodents found that 2-AG levels were decreased in the CeA of these animal models, and that inhibition of MAGL – increasing 2-AG levels – ameliorated abstinence-related anxiety and excessive alcohol intake (216). Mice exposed to chronic mild stress have reduced levels of *DAGLA* expression and reduced *DAGLA* expression in this same study was significantly associated to increased preference for alcohol(133). The study by Ishiguro and colleagues, was also the first to link SNPs in the *DAGLA* gene and alcoholism in humans(133). Our study supports the hypothesis that suggests that the eCB system plays a role in the development and/or maintenance of AUD in adolescents. Previous findings suggest that this vulnerability might be achieved by affecting

sensitivity to anxiety-like symptoms and influencing reward sensitivity to alcohol intake and warrants further study.

While the results of this study suggest a relationship between eCB genes and AUD, we must acknowledge that the results of this study are preliminary and modest. First, many researchers have called hypothesis-based candidate gene approaches into question (163),(217),(218). This is due to the fact that, while very large GWAS studies consistently report that individual SNPs exert very small effects on complex phenotypes such as addiction, most published studies in the field report significant results, even with relatively small sample sizes(163). Considering that these small candidate gene studies may be underpowered (219),(220) (including ours), the significant results reported in the past are most likely false positive(163). It is also possible that this might be the case in the current study, however, the use of a replication sample provides a context in which to interpret the findings and make conclusions about generalisability of the findings. According to the results of the SYS replication analysis, there is a 85% chance that the findings reported herein will be replicated in another dataset of similar size.

Moreover, we are were unable to replicate many of the previously reported findings in relation to substance use and eCB genes. This is because our set-based test eliminated many of the previously reported SNPs as they were non-independent according to our criteria. Moreover, some SNPs that are previously reported, mainly rs2023239 (56),(204), and rs6454674 (209),(221),(222) are not assessed in the assay chips used in the present study or were too infrequent in our cohort for analysis. This was also the case for SNPs within *CNR2* that have been previously evaluated for their relationships with substance use. Considering that our findings were most robust within the analysis considering all timepoints, we cannot be certain what role these SNPs play in the development of AUD (initiation of drinking, susceptibility to binge drinking, proneness towards harmful alcohol use or maintenance of abuse habits etc.). Our findings suggest a more robust relationship at later time points, potentially related to the power that increased prevalence of AUD at the older age affords in a statistical analysis. However, it will also be important to investigate whether these genetic markers are linked to maintenance of drinking in your adults, relative to early initiation behaviours, using larger longitudinal cohorts, when they become available. Finally, there are limitations with the cohort used for this study. Considering

that our cohort is population-based sample of adolescents, the number of problem drinkers is relatively low. Moreover, as the cohorts aged, they reduced in size due to participants leaving the study, diminishing the power of the analyses. Finally, while the results of our replication study were in line with the results of the IMAGEN analysis, our main findings were not significant according to classic standards (p=0.05).

Nevertheless, the present suggest an interaction amongst various candidate genes relevant to the eCB system in predicting AUD - specifically the *CNR1-MGLL-DAGL* loop and their relationship to 2-AG. Further studies are required to further explore the generalisability of these findings and to understand the psychiatric implications of the results.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

Dr Huguet is supported by the Sainte-Justine Foundation, the Merit scholarship program for foreign students, and the Network of Applied Genetic Medicine fellowships.

Dr. Conrod is supported by a Tier 1 Canada Research Chair and this research was supported by a CIHR catalyst grant FRN155406 and the Canadian Cannabis and Psychosis Team Grant (FRN CA7170130)

This work received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313), ERANID (Understanding the Interplay between Cultural, Biological and Subjective Factors in Drug Use Pathways) (PR-ST-0416-10004), BRIDGET (JPND: BRain Imaging, cognition Dementia and next generation GEnomics) (MR/N027558/1), Human Brain Project (HBP SGA 2, 785907), the FP7 project MATRICS (603016), the Medical Research Council Grant 'c-VEDA' (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institute

for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940/2), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1), the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: ANR (project AF12-NEUR0008-01 - WM2NA, and ANR-12-SAMA-0004), the Eranet Neuron (ANR-18-NEUR00002-01), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the fondation de l'Avenir (grant AP-RM-17-013); the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

The Saguenay Youth Study project was funded by the Canadian Institutes of Health Research, Heart and Stroke Foundation of Quebec, and the Canadian Foundation for Innovation.

Bibliography

See Bibliography section of thesis

Tables in article

Tableau 1. – Table 1 (article 1).

Cohort		IMAGEN	SYS
N (female %)		2051 (50.8%)	772 (52.07%)
N Family		2051	401
Age (SD)		14 to 18 years[c]	15 years (1.85)
AUDIT[a]	Control	1476	-
	Case	575	-
GRIP[b]	Control	-	724
	Case	-	48

Table 1. Description of subjects in IMAGEN and SYS. **a**: IMAGEN subjects are classified by status with AUDIT score, case is > or = to 8 and control < 8 ; **a**: SYS subjects are classified by status with GRIP score, case > or = to 2 and control < 2.**c**: IMAGEN cohort is a longitudinal cohort, so it's not possible to calculate the standard deviation (SD).

Tableau 2. – Table 2 (article 1).

SNP	A1	A2	Freq AC	Freq AU	OR	Pvalue	FDR Pvalue
rs782446	C	A	0.224	0.258	0.830	0.0242	0.0811
rs484061	G	A	0.464	0.509	0.833	0.00912	0.0552
rs604300	A	G	0.091	0.114	0.774	0.0327	0.0847
rs507961	T	C	0.197	0.238	0.784	0.00471	0.0427
rs9353525	A	G	0.103	0.136	0.729	0.00400	0.0427
rs4729873	G	A	0.330	0.367	0.849	0.0268	0.0811
rs10488693	T	C	0.058	0.077	0.727	0.0262	0.0811

Table 2. Table of results for Case/Control analysis ALL. A1= minor allele. A2= major allele. Freq AC= Frequency of minor allele in cases. Freq AU = frequency of minor allele in controls. OR= Odds ratio. FDR Pvalue = p value after False Discovery Rate correction

Tableau 3. – Table 3 (article 1).

Phenotype	SNP	A1	NMISS	BETA	OR	STAT	p
AUDIT ALL	rs507961	T	2030	-0.270	0.764	-3.064	**0.002**
	rs9353525	A	2026	-0.301	0.740	-2.605	**0.009**
AUDIT for 14	rs507961	T	2024	-0.243	0.784	-1.098	0.27
	rs9353525	A	2020	-0.216	0.806	-0.776	0.44
AUDIT for 16	rs507961	T	1535	-0.190	0.827	-1.486	0.14
	rs9353525	A	1532	-0.461	0.630	-2.489	**0.01**
AUDIT for 18	rs507961	T	1243	-0.304	0.738	-2.588	**0.01**
	rs9353525	A	1240	-0.320	0.726	-2.053	**0.04**

Table 3. Table of results for logistic model with AUDIT and rs9353525 and rs507961. A1=Minor Allele. NMISS= Number of Non Missing individuals. OR= Odds Ratio. Stat= Coefficient t-statistic.

Tableau 4. – Table 4 (article 1).

Model	Training Accuracy	Testing Accuracy	Sign test(p)	CV Consistency
[rs806368]	0.53	0.49	8 (0.94)	15/20
[rs806368, rs10488693]	0.55	0.50	12 (0.17)	15/20
[rs484061, rs4963307, rs7766029]	0.58	0.541	16 (0.006)	19/20

Table 4. Table of results for the best combinations defined by GMDR for 69 SNPs for AUDIT. Model = SNPs included in the model. Sign test = Sign test result with p-value in parentheses. CV Consistency = cross validation consistency

Figures in Article

Figure 1. – Figure 1 (article 1) .

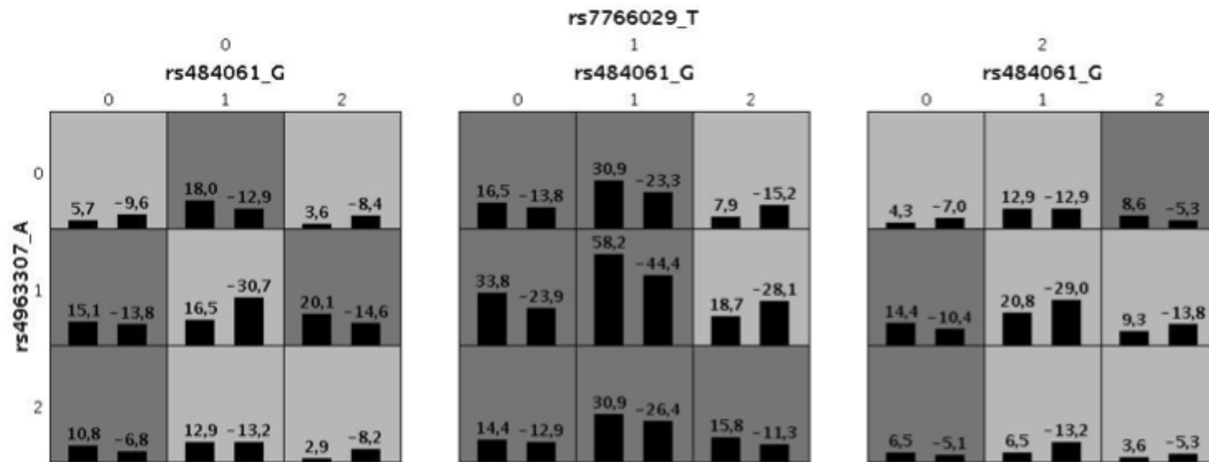


Figure 1 Illustration for the best combination defined by GMDR for 69 SNPs for AUDIT. The allele code is defined by minor allele numbers of rs484061(allele G), rs4963307(allele A) and rs7766029 (allele T). The numbers above the histogram bar, indicate the sum of 'positive' (above the averaged score= 0) and 'negative' (below the averaged score= 0) scores by the combination of genotypes. Also, the dark grey indicates a high-risk combination of the genotypes with alcoholism and light gray for low risk. It defined by sum of positive and negative score, when it's < 0 for low-risk and > 0 for high-risk.

Article 2: Independent contribution of polygenic risk for schizophrenia and cannabis use in predicting psychotic-like experiences in young adulthood: Testing gene x environment moderation and mediation.

Manuscript status: **Under Review, Psychological Medicine (March, 2021)**
Manuscript has been modified to fit stylistically with the present work.

Laurent Elkrief [1*], Bochao Lin[2], Mohammad H Afzali[1,3], Tobias Banaschewski[4]; Arun L.W. Bokde[5], Erin Burke Quinlan[6], Sylvane Desrivières[6], Herta Flor[7,8], Hugh Garavan[9], Penny Gowland [10], Andreas Heinz[11], Bernd Ittermann[12], Jean-Luc Martinot [13], Marie-Laure Paillère Martinot [14], Frauke Nees [4, 7], Dimitri Papadopoulos Orfanos[15], Tomáš Paus[16], Luise Poustka[17], Sarah Hohmann[4], Juliane H. Fröhner[18], Michael N. Smolka[18], Henrik Walter[11], Robert Whelan [19], Gunter Schumann[6, 20], Jurjen Luykx [2,21,22], Marco P. Boks[21], Patricia J. Conrod [1,3], and the IMAGEN consortium.
*Corresponding Author

Author Affiliations:
1. Sainte-Justine Hospital Research Center, Montréal, Québec, Canada.
2. Department of Translational Neuroscience, Brain Center University Medical Center, Utrecht University, Utrecht, the Netherlands
3. Département de psychiatrie et d'addictologie, Université de Montréal, Montréal, QC, Canada
4. Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany;
5. Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland;
6. Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, United Kingdom;
7. Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;
8. Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany;
9. Departments of Psychiatry and Psychology, University of Vermont, 05405 Burlington, Vermont, USA;
10. Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom;
11. Charité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charitéplatz 1, Berlin, Germany;
12. Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2 - 12, Berlin, Germany;
13. Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Saclay, University Paris Descartes - Sorbonne Paris Cité; and Maison de Solenn, Paris, France;
14. Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes; and AP-HP.Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France;
15. NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
16. Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada.
17. Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075, Göttingen, Germany;
18. Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany;
19. School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland;
20. PONS Research Group, Dept of Psychiatry and Psychotherapy, Campus Charite Mitte, Humboldt University, Berlin and Leibniz Institute for Neurobiology, Magdeburg, Germany, and Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China.
21. Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
22. Outpatient second opinion clinic, GGNet Mental Health, Warnsveld, The Netherlands.
23. Department Psychiatry, Brain Center University Medical Center Utrecht, Utrecht, the Netherlands

Corresponding author contact information:
Laurent Elkrief, CHU Ste Justine, Centre de Recherche, 3175 Ch Cote Ste-Catherine, Montreal, QC H3T1C5, laurent.elkrief@umontreal.ca

Declarations of competing interests: Dr. Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire. He received conference support or speaker's fee by Lilly, Medice, Novartis and Shire. He has been involved in clinical trials conducted by Shire & Viforpharma. He received royalities from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. The present work is unrelated to the above grants and relationships.

Author contributions

Laurent Elkrief conceptualized the analysis, ran the analysis, wrote and edited the manuscript. Bochao Lin helped conceptualize the analysis and ran analyses pertaining to the Urecth cannabis cohort. Patricia Conrod and Marco Boks supervised the project, helped conceptualize the project, secured access to the data and edited the manuscript.

Abstract

Background: It has not yet been determined if the commonly reported cannabis-psychosis association is limited to individuals with pre-existing genetic risk for psychotic disorders.

Methods: We examined whether the relationship between polygenic risk score for schizophrenia (PRS-Sz) and psychotic like experiences (PLE) is mediated or moderated by lifetime cannabis use at 16 years of age in 1740 of the individuals of the European IMAGEN cohort. Sensitivity analyses including covariates were conducted and results were replicated using data from 1223 individuals in the Dutch Utrecht cannabis cohort. Two PRS-Sz are studied in main analyses. The first, with a training set p-value threshold (P_t) of 0.05, which best predicts schizophrenia risk, and the second (P_t=0.5), which best predicted cannabis use in the IMAGEN cohort.

Results: PRS-Sz at both P_t (0.5, 0.05) significantly predicted cannabis use (optimal P_t=0.5; p=0.0026) and PLE (optimal P_t=0.5; p=0.011), in the IMAGEN cohort. In the full model, considering PRS-Sz and covariates, cannabis use also significantly associated with PLE in IMAGEN (optimal P_t=0.5; p= 0.035). Results remained consistent in the replication cohort and through sensitivity analyses. Nevertheless, there was no evidence of a mediation or moderation effects.

Conclusions: These results suggest that cannabis use remains a risk factor for psychotic-like experiences, over and above genetic vulnerability for schizophrenia. This research does not support the notion that the cannabis-psychosis link is limited to individuals who are predisposed to psychosis and suggests a need for research focusing on cannabis-related processes in psychosis that cannot be explained by genetic vulnerability.

Introduction

Cannabis use is a well-studied risk factor for psychosis, schizophrenia spectrum disorders, and psychopathology in general. Meta-analysis and systematic reviews have consistently shown that there is a higher incidence of psychotic outcomes among cannabis users (74)(75) and that this relationship is dose dependent (76). Using cannabis during adolescence further increases risk for psychosis (72)(223), earlier onset of psychotic symptoms (79) and worsened prognosis (80). While the epidemiological evidence, along with some experimental evidence(224), suggest a causal link between cannabis use and psychosis, the nature of this relationship remains the focus of fierce debate (225)(226). Generally, three different hypotheses are used to explain the mechanisms of the cannabis-schizophrenia association: (1) the relationship is fully causal, i.e cannabis use causes schizophrenia, (2) the relationship may be partially confounded by shared genetic and environmental confounders and/or reverse causation, (3) this link is entirely non-causal (146)(145).

Considering that part of the etiology of cannabis use and psychosis can be explained through heritable processes (227)(228), recent large scale genome wide association studies (GWAS) have demonstrated that multiple single nucleotide polymorphisms (SNP) are associated with risk for schizophrenia (139), and predict cannabis use behaviours(144)(229). Researchers can summarize the genetic risk for a disease through polygenic risk score (PRS) calculations, derived from the summary statistics generated in these large-scale GWAS studies. Although most PRS for psychiatric diseases can currently only account for a small portion of the variance of disease (approximately <10% (140)), PRS can inform of shared genetic etiology among complex traits, and can also be used to estimate the genetic risk to a trait at the individual level (135). In view of the purported cannabis-psychosis link, researchers have examined the link between schizophrenia PRS (PRS-Sz) and cannabis use. PRS-Sz has been consistently associated to varying levels of cannabis use across numerous cohorts (146) (147) (Carey et al., 2016)(149) (151) (154). Consequently, some have concluded that the relationship between PRS-Sz and cannabis use represents a pathway from genetic risk for schizophrenia to cannabis use (149), or that sensitivity to exposure to cannabis use is moderated by PRS-Sz (151). In contrast, one highly powered study reported that PRS-Sz was not associated to cannabis use disorder in healthy controls, or patients

with psychiatric disorders other than schizophrenia (230). Furthermore, they report that the association between prior cannabis use disorder and later development for schizophrenia was not altered after adjustment for PRS-Sz and PRS of other psychiatric disorders(230); suggesting that the association between cannabis use and development of schizophrenia is not explained by common genetic vulnerability (230). Nevertheless, the results of most studies utilizing PRS-Sz, along with experiments employing discordant relative designs(161) and studies using mendelian-randomization techniques (144)(162) support the second hypothesis – mainly that the relationship between schizophrenia and cannabis use is confounded by shared genetic vulnerability and reverse causation.

The relationship between cannabis use and psychosis development is particularly interesting in the adolescent "clinical high-risk for psychosis"(81) population. These individuals are at a high risk for psychosis in the presence of sub-clinical psychotic symptoms, functional decline and/or genetic risk(81). As such, research on the developmental origins of psychosis risk have focused on the emergence of psychotic-like experiences (PLE) during the adolescent period and how cannabis might influence such trajectories.

PLE are highly prevalent sub-clinical psychotic symptoms (85), reported in about 7% of individuals (Linscott and Van os 2013;Bourgin et al. 2020). Similar to the current models of symptomatology in patients along the psychotic spectrum these sub-clinical symptoms have been further subdivided into various dimensions, such as positive, negative and affective symptoms(100). While these sub-clinical experiences are transitory in about 80% of individuals, PLE are persistent in 20% of individuals (Van Os 2016). Moreover, the presence of PLEs in community samples is associated with increased odds for any mental disorder (OR=3.08 CI95= 2.26-4.21), and psychotic disorders (OR=3.96, CI 2.03-7.73) (Healy et al. 2019).

Considering the close relationship of psychotic-like experiences to psychotic disorders, many have tested the hypothesis that cannabis use also increases one's risk for PLE (see Ragazzi et al. 2016 for systematic-review). One study found that cannabis use is significantly associated with the positive PLE (beta=0.061, $p<1X10^{-4}$), even after controlling for numerous confounding factors(van Gastel et al. 2012). Another study found that the relationship between PLE and

cannabis use is increased in the heaviest of cannabis consumers (schubart); in those who spend >25 €/week on cannabis (i.e heaviest users), there was an increased odds for various domains of PLE such as, negative symptoms (OR 3.4, CI 2.9-4.1), positive symptoms (OR 3.0 CI 2.4-3.6), and depressive symptoms (OR 2.8, CI 2.3-3.3)(Schubart). Furthermore, cannabis use has also been shown to temporally precede PLE in adolescent cohorts (107), but psychotic-like experiences in childhood do not predict cannabis use(108). Overall, the study of PLE in cohorts of cannabis users, may be an interesting avenue to understand the nature and potential directionality of the cannabis-psychosis relationship.

PRS-Sz are also related to PLE. While initial studies reported no relationship between PRS-Sz and PLE (155)(156), more recent studies – with greater power – have found that PRS-Sz is associated to PLE (157)(158)(159)(160). But, there remains contradictory evidence in this field. For example some have reported that PRS-Sz is related to the negative and affective symptom domains (157)(158), but not positive symptoms (hallucinations, paranoia, thought disturbance), whereas others have reported an association between PRS-Sz and positive symptoms (159)(160).

Thus, while the relationship between polygenic risk for schizophrenia and cannabis use has been consistently described in the literature, and the link between cannabis use and psychotic-like symptoms is shown to be significant, the relationship between all three factors (polygenic risk, cannabis use, and PLE) is not yet fully understood. While other studies have attempted to find environmental factors that mediate the relationship between PRS-Sz and cannabis use (149), to our knowledge no study has examined if cannabis use mediates the relationship between PRS-Sz and PLE. Thus, considering that PRS-Sz may be directly or indirectly linked to cannabis use, the current study aims to investigate whether or not the pathway from genetic vulnerability to psychosis symptoms, is at least partially mediated by an indirect pathway through cannabis use.

In addition to the mediation hypothesis, we also test a moderation hypothesis, in which cannabis use might exacerbate genetic vulnerability to schizophrenia, and in turn increase the frequency of PLE. Clarifying the moderating role of genetic vulnerability on the relationship between cannabis and psychosis would also help to inform public health messaging guidelines for

recreational cannabis in which individuals with certain risk profile could be advised accordingly. These two hypotheses will be contrasted against a null hypothesis, which postulates that despite any potential common genetic vulnerability to cannabis use and psychosis risk, the relationship between cannabis use and psychosis risk holds, and is independent of (or in addition to) a common genetic vulnerability (i.e., cannot be explained by common genetic vulnerability). To test all hypotheses, we use a developmentally informed approach that focuses on temporal precedence to confirm mediation between variables. The current study uses data from two independent European cohorts: We use data from the IMAGEN (168) study, a longitudinal study of over 2000 European adolescents, as a discovery sample, and aim to replicate those results in an independent European sample, the Utrecht cannabis cohort (104). The use of the IMAGEN cohort is ideal considering that it allows for a longitudinal view of cannabis use and psychotic like experience development, during the critical years of adolescence. Furthermore, this cohort is relatively well powered to detect mediation effects, as similarly sized cohorts have attempted to discern such effects using similar phenotypes (149). This is compared with the Utrecht cannabis cohort, which is a cohort that has been enriched for PLE and heavy cannabis use; heavy cannabis use being a particularly strong risk factor for development of psychosis and PLE.

Methods

Participants

IMAGEN sample

The IMAGEN study is a longitudinal imaging genetics study of over 2000 healthy adolescents, mostly of European descent. Detailed descriptions of this study, genotyping procedures, and data collection have previously been published (168). The current study uses data for the 2087 who contributed their genetic data. The multicentric IMAGEN project had obtained ethical approval by the local ethics committees (at their respective sites) and written informed consent from all participants and their legal guardians. The parents and adolescents provided written informed consent and assent, respectively at 14 and 16, and then participants gave full consent at 18 and 21 years of age.

Utrecht cannabis cohort

Data from the Utrecht cannabis Cohort comes from a subset (N=1223) of a large (N=17,698) cohort of young Dutch participants, for which genetic, cannabis use and PLE data was available. Detailed descriptions of recruitment methods, genotyping procedures, and data collection was previously published (104)(Boks et al., 2020). Participants gave online informed consent, and the study received approval by the University Medical Centre Utrecht medical ethical commission. Of note is the enrichment for the extremes in PLE and cannabis use data in the Utrecht cannabis cohort. To increase power for gene x environment interactions in previous studies(Boks et al., 2007), data from individuals from the general population was combined with data of participants selected from the top or bottom quintile of total PLE scores, who are either non-users(<2 lifetime exposures to cannabis) or heavy users (i.e. current expenditure for personal cannabis use exceeded €10 weekly).

Phenotype measures

Cannabis use measures

IMAGEN participants were repeatedly assessed for cannabis use at 14, 16, 18 and 21 years of age using questions taken from the European School Survey of Alcohol and other Drugs (ESPAD) questionnaire. The ESPAD is a self-report questionnaire that measures use of various drugs of abuse, including cannabis(Hibell et al., 1997)(Hibell et al., 2004). With very few participants reporting cannabis use at 14 years of age, we focus our analyses on data that were collected at the 16-year-old assessment, using responses to the question "On how many occasions in your whole lifetime have you used marijuana (grass, pot) or hashish (hash, hash oil)?". Answers are scored on a scale ranging from 0-6: '0' = 0, '1-2 times'=1, '3-5 times'=2, '6-9 times'=3, '10-19 times'=4, '20-39 times'=5, '40 or more times'=6. In the Utrecht cannabis Cohort, lifetime cannabis use data was reported according to the following categories never = 0, '1 time' = 1, '2 times' = 2, '5-9 times' = 3, '>10 times' =4. Cannabis use data was dichotomized into "case/control", where >=10 lifetime uses are considered cases. This dichotomization allows for direct comparison of the IMAGEN cohort to the Utrecht cohort.

Psychotic-like experience measures

PLE data for both cohorts was drawn from the Community Assessment of Psychic Experiences - 42 (CAPE) questionnaire(99). CAPE-42 is a widely used self-report questionnaire that reliably measures lifetime psychotic-like experiences(189). The CAPE-42 has three subscales that measure positive, negative and depressive symptom dimensions. The CAPE-42 measures frequency of symptoms, along with distress caused by symptoms. In the primary analyses, we use the sum total of frequency scores, while we look at the various sub-dimensions in secondary analysis. Due to the skewed distribution of scores, the log-transformed sum score of each individual dimension and total score of the frequency of symptoms was used. We used CAPE-42 data from the 18-year-old follow up for IMAGEN cohort.

Genetic Data

IMAGEN

The genotyping was conducted using the Illumina Quad 610 chip and 660Wq at the "Centre National de Genotypage" (Paris, France). Only autosomal SNPs are kept for this study. Following quality control steps, genetic data remained for 1740 individuals. Baseline quality control steps and principal component analysis to control for ancestry are described in supplementary materials.

Utrecht

The genotyping in this cohort was conducted using either the Illumina® HumanOmniExpress (733,202 SNPs; 576 individuals) or the Illumina® Human610-Quad Beadchip (620,901 SNPs; 768 individuals). As with the IMAGEN sample, quality control steps and principal component analysis for ancestry are described in supplementary materials.

Analysis

Polygenic risk scores (PRS)

Polygenic Risk Scores for schizophrenia (PRS-Sz) were constructed for each of the IMAGEN and Utrecht cannabis individuals, who passed genetic quality control. PRS-Sz were built using data

from the most recent schizophrenia GWAS based on 40675 cases and 64643 controls (139) as a training set (for description of base set see supplementary materials). Polygenic risk scores were built in PRSice2 (194), following quality control protocols described in supplementary materials. First, for each individual, a PRS-Sz was built using the p-value threshold (P_t) of <0.05, as this was the Pt that optimally captures phenotypic variance in schizophrenia (142). Next, a PRS-Sz for each individual using 12 different p-value thresholds (P_t) [$5x10^{-8}$,$5X10^{-7}$,$5X10^{-6}$,$5X10^{-5}$,$5X10^{-4}$,$5X10^{-3}$,$5X10^{-2}$,0.1,0.2,0.3,0.4,0.5], was calculated to capture the PRS which explains the most variance (Nagelkerke's R^2) in the logistic regression between PRS-Sz and cannabis use (binary measure). We report the R^2 of the PRS, which is the R^2 of the full model subtracted by the variance explained by the covariates of interest. To control for type-1 error, we obtained an empirical p-value for the "best" PRS-Sz using 10,000 label swapped permutations in PRSice2 (194). We aligned our analyses closely to the replication study, using the same protocol to create PRS, and by using the same dichotomised phenotype measurements, in both IMAGEN and Utrecht cannabis cohorts. However, we added age (in years) as a covariate measure in the Utrecht cannabis cohort regressions, as this study is cross sectional and includes participants with a wide age range. To ease interpretability of results we scale the PRS, using the scale function in R(195). Using the same methods, we created a PRS for cannabis use (PRS-Can), using publicly available data from the GWAS studying lifetime cannabis use (144) (a detailed description of the base set can be found in supplementary materials). We use the PRS with a P_t of 0.05 for sensitivity analysis, as this is the PRS that best predicts cannabis use in our cohort (see supplementary materials for a more detailed description).

Regression analysis

Multiple multinomial linear and logistic regressions were used to assess the relationships between PRS-Sz, cannabis use and PLEs. For primary analyses, the PRS-Sz for the Pt with the highest Nagelkerke's R^2, as well as the PRS with Pt <0.05 are used as independent variables (IV). First, the relationship between PRS-Sz and total CAPE score (log-transformed) was examined. In this model sex as well as the first six principal components (which explain >90% of the variance in the IMAGEN cohort) were used as covariates, as previously published(175)

This model was then replicated using the Utrecht cannabis data, adding age as a covariate. Regression models were constructed using R (195). We correct for the use of two p-value thresholds and therefore consider a p-value of 0.025 (0.05/2) as significant. The association between PRS-Sz and the various sub-domains of the CAPE-42 questionnaire were analyzed in secondary analyses. Variance explained by the independent variables and covariates was calculated in the regression analyses as R^2 (or Nagelkerke's R^2 for dichotomous dependent variables).

For sensitivity analysis, we used PRS-Sz derived using more stringent P_t ($5X10^{-8}$, $5X10^{-5}$) as independent variables. Moreover, we used the PRS for cannabis use as a covariate in regression analyses. The PRS-Can was included as a covariate only for the IMAGEN dataset, as the Utrecht cannabis cohort was included in lifetime cannabis use GWAS cohort (144).

Mediation and moderation analysis

Using the *R* package '*mediation'* (197), mediation analysis was used to examine if cannabis use mediates the relationship between PRS-Sz and PLE. As alluded to above, our hypothesis, is that cannabis use (M) mediates the relationship between PRS-Sz (independent variable; IV) and PLE (dependent variable; DV). In all mediation analyses, sex and first 6 PC are used as covariates. A non-parametric bootstrap (n=10000) is used to estimate the sampling distribution of the indirect effect. If the confidence interval(CI) of the effect does not cross zero, the indirect effect or average causal mediation effect (ACME) is significant. Mediation analysis is only executed using the IMAGEN data as this dataset is the only sample that assessed cannabis use some years before the PLE assessment and therefore the only dataset that can provide a true estimate of a longitudinal relationship. Finally, to examine if cannabis use moderates the relationship between IV and DV, we use the *R* package '*MeMoBootR*' (232), which allows for "a complete two-way moderation analysis with one moderator, similar to model 1 in PROCESS by A. Hayes (2013)".

Results

Sample characteristics

Characteristics of participants who passed genetic QC and responded to cannabis use and PLE questionnaire data are detailed in supplementary Table S2. Data from a total of 1740 individuals were used to calculate PRS-Sz in IMAGEN, and 1223 individuals in Utrecht cannabis cohort. In both IMAGEN and Utrecht Cannabis cohort samples, males report higher cannabis use than females($p<0.001$). The total frequency of CAPE-42 symptoms reported is significantly greater in males($p<0.001$) in the IMAGEN cohort. There was no difference in the reported total CAPE-42 symptoms between males and females in the Utrecht cannabis cohort (Table S2). Female participants in both cohorts report significantly higher scores in the depression symptom sub-scale of the CAPE-42($p<0.001$). Finally, the mean age of the Utrecht cannabis cohort is 20.5 years.

Association of PRS-Sz and cannabis use

After accounting for covariates, the PRS-Sz using a P_t of 0.05 predicted cannabis use in both cohorts (β_{IMAGEN}=0.062, p=0.022, R^2=0.007; $\beta_{Utrecht\ cannabis}$=0.084, p=7.31x10^{-7}, R^2=0.024, Table 1). The P_t that best predicted cannabis use status in the IMAGEN cohort was P_t=0.5 (β_{IMAGEN}=0.062 p=0.0026, $p_{empirical}$=0.01, R^2= 0.013, Table 1). This result was replicated in the Utrecht cannabis cohort ($\beta_{Utrecht\ cannabis}$=0.076, p=9.62X10^{-9}, R^2=0.032, Table 1). The results for the regression model, including covariates, is shown in supplementary Table S3. The explained variance and p-value are shown in Figure 2 and detailed for each P_t in supplementary Table S4.

Association of PRS-Sz with PLE

PRS-Sz also predicted PLE in both cohorts (Table 2), when accounting for sex and PC as covariates. PRS-Sz(P_t =0.5) was significantly associated to CAPE-42 scores in both cohorts (β_{IMAGEN}=0.014 p=0.011, $\beta_{Utrecht\ cannabis}$=0.0048, p=<0.0001). The PRS-Sz (P_t=0.05) was nominally significant in the IMAGEN cohort, and remained significantly associated to CAPE-42 score, after correction for multiple tests in the Utrecht cannabis cohort ($\beta_{Utrecht\ cannabis}$=0.0053, p=<0.0001). The results for the full regression model, including covariates is shown in supplementary Table S5.

The relationship between both of the retained PRS-Sz and the different subdomains in the CAPE-42 questionnaire was studied. In the IMAGEN cohort, PRS-Sz at both P_t were significantly associated to the depression subscale (P_t =0.05; β_{IMAGEN}= 0.02 p=0.0046, P_t =0.5; β_{IMAGEN}= 0.022, p=0.0015). Only the PRS-Sz (P_t =0.5) predicted negative symptoms in the IMAGEN cohort (β= 0.016, p=0.024). On the other hand, in the Utrecht cannabis cohort, both PRS-Sz were significantly associated to all three sub-domains (supplementary Tables S6, S7, S8), but the strongest result was that for the negative symptom sub-scale(p<0.0001).

Sensitivity analysis

In the IMAGEN cohort, the relationship between PRS-Sz (P_t=5X10^{-5}) and CAPE-42, as well as the relationship between PRS-Sz (P_t=5X10^{-5}) and lifetime cannabis use, were significant, when considering the same covariates as the main analyses (p<0.05; see suppl. Table S9). The PRS-Sz using a P_t of 1X10^{-8}, was not predicative of either cannabis use or PLE (p>0.05, see suppl. Table S9). The lifetime cannabis use PRS (PRS-Can) did not predict dichotomous cannabis use measures at 16 years of age, or CAPE scores at 18 years of age (p>0.5). Nevertheless, the PRS-Can (Pt=0.05) did significantly predict cannabis use at 18 and 21 years of age in the IMAGEN cohort ($\beta_{18\ years}$=0.052, p=0.036, R^2= 0.005; $\beta_{21\ years}$=0.065, p=0.004, R^2=0.01), and was therefore included as a covariate in our sensitivity analysis regression. After including PRS-Can (Pt=0.05) as a covariable in a linear regression examining the relationship between PLE (dependent variable) and PRS-Sz (P_t=0.5) and cannabis use (independent variables) and all other covariables described above, both independent variables remained significant (β_{PRS}= 0.014 p= 0.013, $\beta_{Cannabis\ Use}$= 0.038, p= 0.037). Moreover, the relationship between PRS-Sz (P_t=0.5) (independent variable) and cannabis use (dependent variable) remained significant, after inclusion of PRS-Can into the regression (β= 0.025 p= 0.0022, supplementary Table S10).

Moderation and mediation analysis

In both mediation analyses (P_t =0.05 and 0.5), the average direct effects of PRS-Sz on PLE are significant (p<0.01). Although cannabis use at 16 years old was independently predictive of PLE at 18 years of age, and PRS-Sz predicted cannabis use, the average causal indirect pathway (i.e the mediation pathway) from PRS-Sz to PLE through cannabis was not significant (Figure 3;

supplementary Table S11), suggesting that cannabis use does not mediate the relationship of PRS-Sz to PLE in adolescents. In the full model, in both cohorts, cannabis use predicted total PLE, after accounting for PRS-Sz (at both P_t) and various covariates (Figure 4; supplementary Table S12). In our moderation model, the interaction between cannabis use and PRS-Sz was also not significant ($p>0.05$; Table 3), suggesting that both cannabis use and PRS-Sz independently predict PLE.

Discussion

In this study, we examine whether polygenic risk for schizophrenia predicts cannabis use, and higher levels of psychotic like experiences, in two independent European cohorts. Furthermore, we explore potential hypotheses, through mediation and moderation analyses. Our results, demonstrate that cannabis use can be reliably predicted by PRS-Sz, strengthening the existing literature (146)(147)(148)(149). We are not the first to study the relationship between PRS-Sz and cannabis use in the IMAGEN cohort. French et al. previously demonstrated that cannabis use at 14 years of age interacted with PRS-Sz in decreasing cortical thickness from 14.5 to 18.5 years old, using the IMAGEN dataset(233). Here we extend these findings by showing that the PRS-Sz predicts PLE. This too is in line with other work, using a variety in PLEs assessments in various sub-domains (157)(158)(159)(160). The current study confirms that PRS-SZ and cannabis use are linked to risk for PLE overall and in the depressive and negative domains in both samples and also predict positive symptoms in the older Utrecht cannabis cohort.

Considering the abundance of observational evidence showing temporal precedence of cannabis use in risk for psychosis, a reasonable alternative to a causal hypothesis is the proposal that cannabis use and PLE are explained through common genetic risk. However, our findings do not support this explanation, despite showing that PRS-SZ is correlated with both PLE and cannabis outcomes in both cohorts. This is in line with recent work that reported that various classes of cannabis use were associated to increased risk for psychotic experiences, even after adjusting for family history of schizophrenia (108) and other work adjusting for PRS-Sz(149). To our knowledge this is the first study to examine if cannabis use mediates the relationship between PRS-Sz and PLE. Through longitudinal and cross-sectional designs, our analyses did not find any evidence to support mediation or moderation hypotheses that explain the relationship between

lifetime cannabis use and PLE. Consequently, these null findings suggest that despite the common genetic vulnerability of psychotic experiences, cannabis use and schizophrenia (144)(150), both PRS-Sz and cannabis use independently increase one's risk for PLE, leaving room for alternative explanations of the cannabis-psychosis relationship.

Two recent studies employed Mendelian Randomization (MR) to investigate a causal links between cannabis use and schizophrenia (162)(234). In both of these works there was weak evidence to support the causal hypothesis in the direction schizophrenia to cannabis use, while the reverse relationship was strong (162)(234). While these studies are limited by the power of the respective GWAS studies used, recent work has called into question causal inferences made in MR studies of complex traits(235), and suggest the use of a latent causal variable (LCV) instead. In latent causal variable models, genetic correlation between "two traits is mediated by a latent variable which has a causal effect on each trait" (235). Accordingly, a recent study examined the causal link between schizophrenia and lifetime cannabis use employing LCV and found no evidence for a causal genetic link between the two (236). Taken together, these reports do not preclude the possibility of a causal mechanism linking cannabis use to psychosis. Instead, they – along with the results presented above – suggest that psychosis or psychosis risk, and cannabis use may be linked through another environmental mediator rather than being linked through a common genetic predisposition.

The findings of the current study suggest that the variance in cannabis use that is most linked to PLEs is that which is not accounted for by PRS-Sz. This is interesting and suggests that future studies could focus on environmental factors influencing cannabis behaviours, such as the type of cannabis used, or available in a given population, the effects of advertisements endorsed by the cannabis industry, differing legalization frameworks, and cannabis potency, when attempting to understand the link between cannabis and psychosis.

In secondary analysis of the current study, PRS-Sz was associated with depressive and negative sub-domains of the CAPE-42 in the adolescent cohort, while in the older Utrecht cannabis cohort PRS-Sz was associated to all three sub-domains. results are in line with previous reports that found no association between PLE-Sz and positive symptoms in adolescent

populations (155)(157)(158). In contrast, one study has reported an association between PRS-Sz and positive PLE symptoms in their adolescent cohort (159), however only when considering non-zero responders, i.e. those who have already manifested positive symptoms. Moreover, the association between PLE-Sz and the positive symptom domain in the Utrecht cannabis cohort supports recent evidence (150) which suggests that genetic overlap between positive and negative psychotic experiences and schizophrenia might be stronger in adulthood than in adolescence (150). As previously suggested by Jones (157), these results imply that genetic risk for schizophrenia is in fact associated to positive PLE, but that this risk may be expressed in young adulthood rather than adolescence. On the other hand, other environmental risk factors – such as cannabis use – may be what cause these same positive symptoms in adolescents.

Limitations

While the findings of this study are consistent with previous independent works, this study is not without limitations and results should be interpreted accordingly. First, polygenic risk scores can only explain very small portions of the variances of the phenotypes they study (140). In this study, PRS-Sz explained up to 3.2% of the variance of cannabis use and 2.1% of variance of PLE, when accounting for confounders. PRS also only incorporates data from common genetic variants; as such a significant portion of the genetic effects may not be captured through the PRS, such as the effects of rare variants and copy number variants, which also may play a role in the pathogenesis of schizophrenia (237). In addition, the IMAGEN data set had many missing data points, which could bias our results. Next, considering the self-report nature of our phenotypic measures, our results may be at risk for measurement error, due to underreporting of symptoms, leading to weakened power. Moreover, we use the PRS-Sz – which was built to predict outcomes of clinical schizophrenia in adults – to predict psychotic-like experiences in adolescent and young adult populations. While the PRS-Sz has been used to reliably predict PLEs (157)(158)(159)(160), the most discriminant SNPs for PLE may have not been captured by our PRS. However considering the genetic overlap between schizophrenia and PLE (150), our significant result remains informative. Finally, the focus on lifetime cannabis use as well as the use of a binary measure of cannabis use, did not allow for the proper investigation of dose-response relationships in the

IMAGEN study. This is an important consideration, as heavy cannabis use with high THC content poses a particular risk for psychosis(76).

Conclusion

In conclusion, although the current study could not confirm a mediated pathway between schizophrenia risk and PLE through cannabis use, the results contribute to the literature by showing the positive relationship between cannabis and future psychotic-like symptoms, while controlling for genetic vulnerability. This result is of public health importance. While cannabis producers would like to claim that cannabis use is only contra-indicated for individuals with a personal or family history of psychosis, the current findings suggest that cannabis use remains a risk factor for psychotic-like experiences, over and above known genetic vulnerability for schizophrenia. Moreover, there was no evidence that genetically vulnerable individuals were more susceptible to the psychosis-related outcomes of adolescent onset cannabis use. As suggested by other authors (157), identifying a causal mechanism in the pathway from cannabis use to psychosis is extremely important for the development of targeted preventative interventions aimed at reducing cannabis use and/or schizophrenia risk.

Acknowledgements

We would like to thank all of the participants of both IMAGEN and Utrecht Cannabis Cohort studies. We thank the investigators of the Utrecht Cannabis Cohort for making data available.

Financial support

Dr. Conrod is supported by a Tier 1 Canada Research Chair and this research was supported by a CIHR catalyst grant FRN155406 and the Canadian Cannabis and Psychosis Team Grant (FRN CA7170130)

This work received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant

‘STRATIFY’ (Brain network based stratification of reinforcement-related disorders) (695313), ERANID (Understanding the Interplay between Cultural, Biological and Subjective Factors in Drug Use Pathways) (PR-ST-0416-10004), BRIDGET (JPND: BRain Imaging, cognition Dementia and next generation GEnomics) (MR/N027558/1), Human Brain Project (HBP SGA 2, 785907), the FP7 project MATRICS (603016), the Medical Research Council Grant 'c-VEDA’ (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940/2), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1), the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: ANR (project AF12-NEUR0008-01 - WM2NA, and ANR-12-SAMA-0004), the Eranet Neuron (ANR-18-NEUR00002-01), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the fondation de l’Avenir (grant AP-RM-17-013); the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

Funding for the Utrecht Cannabis cohort was provided by NWO, the Dutch council for scientific research (ZonMW TOP grant no. 91207039).

Bibliography

See Bibliography section of thesis

Tables in publication

Tableau 5. – Table 1 (article 2).

IMAGEN (N=1407)	p-value threshold	β	std.error	R2	p value
	$p \leq 0.05$	0.062	0.027	0.0074	**0.022**
	$p \leq 0.5$	0.062	0.021	0. 0127	**0.0026**
Utrecht cannabis (N=1223)					
	$p \leq 0.05$	0.084	0.017	0.024	**7.31×10^{-7}**
	$p \leq 0.5$	0.39	0.069	0.032	**9.62×10^{-9}**

Table 1. Predictive value of PRS-Sz on cannabis use. This table shows results for the most predicative polygenic risk score on cannabis use (case/control status). Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. R^2 variance explained is the Nagelkerke's R^2. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the Utrecht cannabis cohort. β= main effect size, std.error = standard error

Tableau 6. – Table 2 (article 2) ..

	β	std.error	p value	R2	adjusted R2
IMAGEN (N=1156)					
$p \leq 0.05$	0.011	0.005	0.0357	0.016	0.009
$p \leq 0.5$	0.014	0.005	**0.011**	0.018	0.011
Utrecht cannabis (N=1223)					
$p \leq 0.05$	0.005	0.002	$\mathbf{2.64 \times 10^{-4}}$	0.017	0.009
$p \leq 0.5$	0.005	0.001	$\mathbf{1.63 \times 10^{-5}}$	0.021	0.014

Table 2. Predictive value of PRS-Sz on psychotic like experiences (PLE). This table shows results for the most predicative polygenic risk score on PLE. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the Utrecht cannabis cohort. β= main effect size, std.error = standard error

Tableau 7. – Table 3 (article 2).

IMAGEN (N=1061)			β	std error	p value	R2	adjusted R2
	Pt=0.05					0.024	0.015
		PRS-Sz	0.004	0.002	**0.025**		
		Cannabis Use	0.045	0.019	**0.016**		
		PRSSz:CannabisUse	-0.006	0.006	0.251		
	Pt=0.5					0.027	0.018
		PRS-Sz	0.004	0.001	**0.005**		
		Cannabis Use	0.046	0.019	**0.015**		
		PRSSz:CannabisUse	-0.006	0.004	0.146		
Utrecht cannabis (N=1223)							
	Pt=0.05					0.039	0.03
		PRS-Sz	0.005	0.002	**0.01**		
		Cannabis Use	0.067	0.013	**1.34×10^{-7}**		
		PRSSz:CanUse	-0.002	0.003	0.552		
	Pt=0.5					0.042	0.033
		PRS-Sz	0.004	0.001	**0.003**		
		Cannabis Use	0.065	0.013	**3.16×10^{-7}**		
		PRSSz:CanUse	-0.001	0.002	0.517		

Table 3. Moderation Analysis. This table shows results for the moderation analysis. Cannabis use (case/control) and PRS-Sz both predict total PLE in both cohorts at both Pt, but their interaction is insignificant. PRSSz:CanUse is the interaction term. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the Utrecht cannabis cohort. β= effect size, std.error = standard error

Figures in publication

Figure 2. – Figure 1 (article 2).

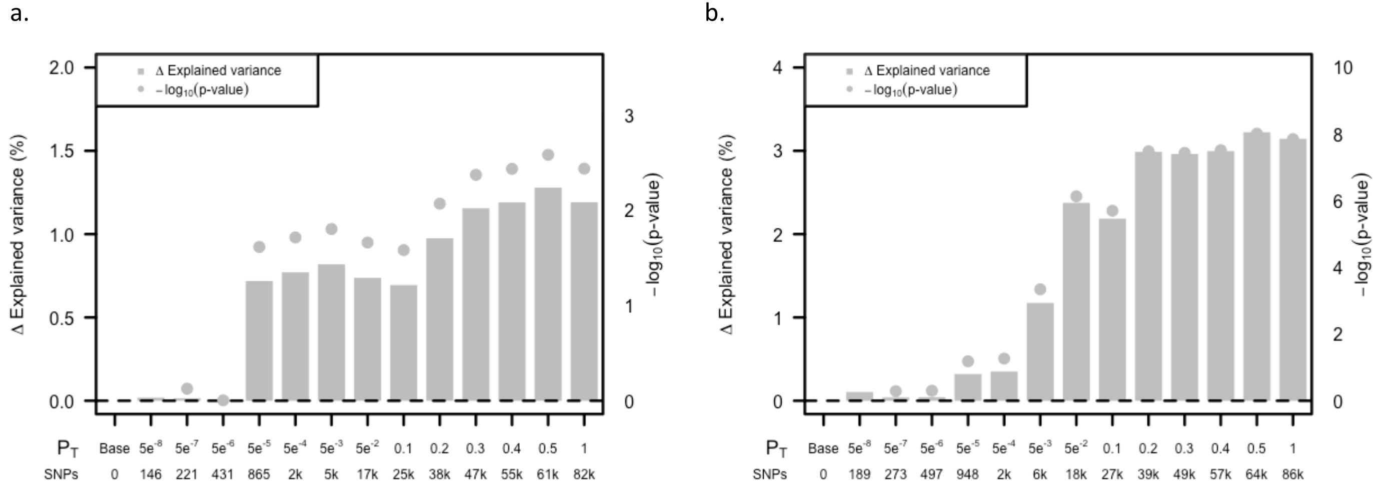


Figure 1. Predictive value of SCZ polygenic risk scores (PRS) on cannabis use. Left hand y-axis, Explained variance (R^2 as %). Right hand y-axis, –log(p-value) of regression between PRS-Sz and cannabis use. X-axis, various pt used for regression, and number of SNP included in PRS. (a) IMAGEN cohort. pt that best explains cannabis use pt=0.5 (β_{IMAGEN}=0.062 p=0.0026, $p_{empirical}$=0.01, R^2= 0.013). (b) Utrecht cannabis cohort. pt that best explains cannabis use pt=0.5 ($\beta_{Utrecht\ cannabis}$=0.076, p=9.62X$10^{-9}$, R^2=0.032).

Figure 3. – Figure 2 (article 2).

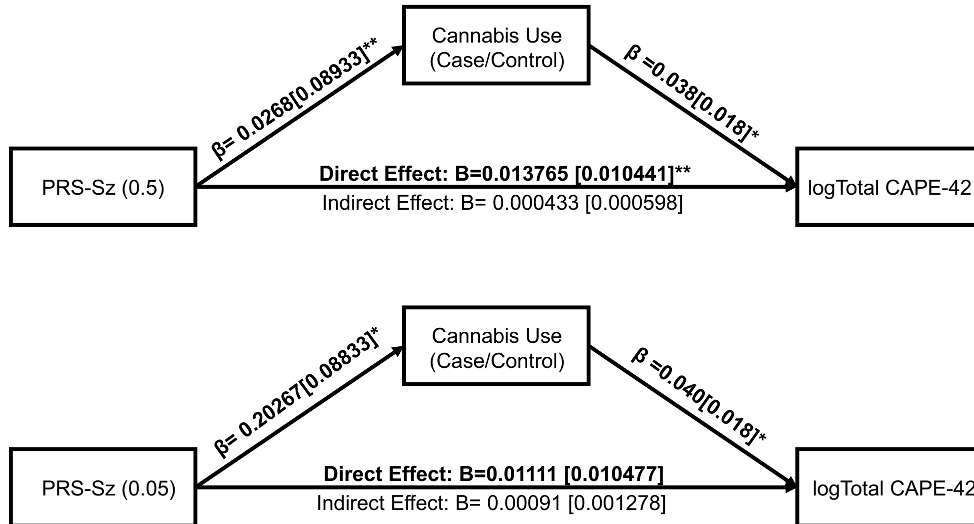


Figure 2. Results of mediation analysis. Independent variable= Polygenic risk score for schizophrenia (PRS-Sz), Dependent variable=logTotal of CAPE-42. Mediator = Cannabis use. As shown, the effect of PRS-Sz on psychotic like experiences as measured by the CAPE-42 questionaire was not mediated by cannabis use. The effect of PRS-Sz was significantly associated to cannabis us in both models as was cannabis use effect size onto CAPE-42 scores. The direct effect was significant. Unstandardized indirect effects were computed for each Of the 10000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5^{th} and 97.5^{th} percentiles. *=p<0.05 **=p<0.01

Conclusion

Interpretation and limitations of the presented studies

Through the use of various genetic and statistical approaches, I studied the relationship between genetic markers, psychopathology, cannabis and the endogenous cannabinoid system.

In study one, we report that genetic markers of the endocannabinoid system are related to alcohol use disorder patterns in a cohort of European adolescents. These results are not replicated reliably in an independent cohort. Additionally, we report evidence that the interaction of SNPs across the endocannabinoid system (specifically *CNR1-MGLL-DAGLA* genes) is significantly associated to scoring 7 or more on the AUDIT questionnaire. This finding is interesting considering that experimental research, in rodents, has shown that the CB1-MAGL-DAGL "system" plays a role in the physiological response to excessive alcohol consumption by modulating 2-AG levels.

Studies have shown that 2-AG levels are increased to modulate the HPA axis in situations of chronic stress (210). Moreover, these increased levels of 2-AG have been shown to play a role in the reduction of stress induced anxiety response in rodents (210). The anxiety-like alcohol withdrawal symptoms seen in rodent models of alcohol dependence are mediated by the HPA axis, specifically corticotropin releasing factor levels in the central nucleus of the amygdala (36). Hence, in models of alcohol dependence, it was demonstrated that 2-AG levels were decreased in the CeA, and that increasing levels of 2-AG – through inhibition of MAGL – reduced abstinence related anxiety behaviours and excessive alcohol consumption(238). Other work also demonstrated an explicit role for DAGL in these models of chronic stress and alcohol consumption. Rodents exposed to chronic stress have reduced *DAGLA* gene expression, and this reduction of *DAGLA* was associated to increased preference for alcohol(133). Overall, this experimental evidence would suggest that the CB1-MAGL-DAGL "system", through its regulation of 2-AG levels, could be involved in the maintenance stage of alcohol dependence (37)(38) by mediating anxiety levels during periods of withdrawal and drug seeking. As such, the results of study one may be seen as possible evidence in humans for the involvement of this system.

Nevertheless, this conclusion must be contrasted with the notion that the GMDR results was not replicated in the SYS cohort and may therefore be due to type-I error.

The results of this study must be interpreted in light of the limitations of the study. First, we must consider the study design. As a traditional candidate gene study, this study is at high risk for type-I error, due to the small effect size of common genetic variants and the power typically required to detect these effects (163). It could also be argued that our statistical correction used for multiple testing was not stringent enough. Moreover, we were unable to reproduce our findings in our replication cohort or the findings of older studies examining similar phenotypes. This is probably due to the fact that our set-based test eliminated many of the SNPs that have been previously reported. Also, other SNPs that have been previously studied were also not genotyped in our assay chips. The cohort used for both the original study and the replication sample, may have also been underpowered to detect reliably significant genetic effects. Considering that these are adolescent cohorts, the number of "cases" was quite small. Finally, to our knowledge there has been no reliable GWAS study to date that has identified an endocannabinoid related gene as significantly associated to alcohol or substance abuse behaviour. Taken together, while the animal and candidate gene literature suggest that the endocannabinoid system plays a role in the pathogenesis of alcohol use disorder, our results must be seen as inconclusive, or preliminary at best.

In study two we report that lifetime cannabis use does not mediate or moderate the relationship between PRS-Sz and PLE. Through multiple multinomial linear and logistic regressions, we also demonstrate that cannabis use and PRS-Sz independently predict PLE, suggesting that cannabis use is a risk factor for psychotic-like experiences, even when controlling for genetic risk for schizophrenia. This result is quite robust considering that the results of the sensitivity analyses and the replication of results in an independent cohort. Moreover, these results are in line with previously reported findings.

Our work is particularly interesting as we are the first study to report the study of cannabis use as an environmental mediator along the pathway of PRS-Sz and PLE, although one other group has studied other environmental mediators of the relationship between PRS-Sz and PLE (149).

These results are also of public health importance, because although cannabis producers would like to suggest that cannabis use is specifically contra-indicated for those with a genetic risk for schizophrenia, the current findings suggest that cannabis use remains a risk factor for psychotic experiences, over and above known genetic vulnerability for schizophrenia. This is noteworthy because it suggests an additive effect of cannabis use; cannabis use still predicts future psychotic-like symptoms, while not mediating the pathway between genetic risk for psychosis and psychosis. This conclusion is in line with a discordant relative study, examining the relationship between cannabis use and psychotic-like experiences(239). In this cross-sectional study of over 2000 pairs of twin and non-twin sibling pairs, Karcher and colleagues report that PLE were significantly associated to frequent cannabis use (β = 0.11; 95% CI, 0.08-0.14), cannabis use disorder(β = 0.13; 95% CI, 0.09-0.16), and current cannabis use (β = 0.07; 95% CI, 0.04-0.10) (239). While much of the relationship between cannabis use and PLE is explained by genetic influence (between 69.2% to 84.1% of the variance of this association), some of the association can only be explained by "person-specific" pathways (239). Therefore, cannabis use may be partially causally linked to the psychosis continuum in a pathway that is mediated by other environmental risk-factors. As such, future studies focusing on factors influencing cannabis behaviours such as cannabis availability, type of cannabis used, effects of advertisements endorsed by the cannabis industry, and cannabis potency, could potentially reveal causal pathways in the relationship between cannabis use and psychosis.

The results of the secondary analysis, examining the various domains of the CAPE-42 questionnaire are also of note. While PRS-Sz was associated to depressive and negative domains in the adolescent IMAGEN cohort, all three domains were significantly predicted by PRS-Sz in the older Utrecht cannabis cohort. This would imply that the genetic propensity to schizophrenia may first be expressed as negative or affective symptoms in younger adolescents, while it is associated to positive symptoms in young adulthood. Therefore, while positive symptoms are prevalent during adolescence, our results suggest that other environmental factors – such as cannabis consumption – may be the driving force behind these symptoms. This is a logical conclusion, considering that cannabis consumption is related to earlier expression of positive psychotic symptoms(79).

While study two utilized a more robust methodology, and is in line with previously reported works, it too is not without limitations. First, is the use the PRS-Sz as a predictor for cannabis use and PLE outcomes. This could understandably lead to a construction of a PRS which does not capture SNPs that are most informative for cannabis use and PLE outcomes. Nonetheless, considering the genetic overlap between cannabis use, schizophrenia and PLE(144)(150), and our use of the PRS for cannabis use as a covariate in sensitivity analyses, our results remain informative. We must also consider the fact that the cannabis use measure does not account for severity of use, or dose. This is a vital consideration, due to the important role of dose in the cannabis-schizophrenia relationship, as well as the meta-analysis results which suggest that the relationship between cannabis use in the clinical high risk for psychosis population is accounted for through heavy cannabis consumption rather than lifetime use measurements(76)(84). Although the decision to use the >10 cut-off allowed for direct comparison between the IMAGEN and Utrecht cannabis cohort, this study could have been more informative with a more precise measure of severity of cannabis use. Moreover, the use of a binary measure such as the one used in the above study could potentially lead to a loss of information about individual differences (for example, the potential effects of ten lifetime cannabis consumptions versus twenty uses).

The study of PRS in clinical psychiatry is exciting, as it represents a novel way of calculating an individual's lifetime genetic risk for a certain disease or disorder. Nevertheless, as mentioned above, to date, PRS represent at best 10% of the variance of complex traits(140). As such, PRS could be not used as clinical tools on their own, rather as an additional factor to be considered among other reliable risk factors(240). This highlights an important limitation of this study: models that only consider PRS are incomplete. In this study, the PRS-Sz explained up to 3.2% of the variance of cannabis use and 2.1% variance of PLE, when accounting for confounders. Therefore, while the results were statistically significant, they are probably not clinically significant. While we should not have expected a much larger explanation of the variance, these low R^2 could be explained by the fact the studied populations are more representative of the general population than clinical populations; the frequency of PLE symptoms and number of cannabis users is smaller than that of a clinical or "at-risk" population. Thus, the results of the

above study should not be used to make assumptions about clinical populations or the clinical significance of the relationship between PRS-Sz, cannabis use and psychotic-like experiences; these results only support the notion that the cannabis-psychosis spectrum relationship goes beyond one's genetic propensity to schizophrenia, and that further study in clinical samples is required. This logic should also be applied to the results of study one. The very low number of participants reporting "clinically significant" alcohol use may have reduced the statistical power of the study. Moreover, the fact that the populations studied are non-clinical reduces our ability to draw meaningful conclusions on the clinical significance of our results.

Although the advent of polygenic risk scoring research has led some to envision its use in clinical practice, an important number of ethical considerations could be addressed in any discussion pertaining to the use of PRS in clinical work. I believe that it is relevant to discuss two of these ethical issues, as they relate to both of the studies described in this master's thesis. First, we must consider the clear disparities in the predictive value of current polygenic-risk scores across different populations; mainly between European and non-European populations(241)(242). As many have previously shown, current polygenic risk scores invariably predict individual risk more accurately in European populations than non-European populations(142)(243)(244)(245). The reason for this is simple, to date most GWAS studies (on which all PRS are constructed) are based on population samples that are predominantly Eurocentric(241). And so, a widespread application of PRS would disproportionally advantage those of European genetic descent, while it would leave other populations – which already endure large disparities with regards to healthcare – behind (see (241) for a more detailed discussion on the subject). This is extremely relevant to the above-described works, considering that we only included individuals of European descent in these studies. While it would have been statistically unsound to have included non-Europeans in the current works (a total of 11 individuals of non-European descent are found among the 2000+ participants in the IMAGEN study), we are cognisant of the implications of this decision; specifically, that our results are not applicable to any populations of non-European ancestry. We also recognize that a continued effort should be made to remediate the underrepresentation of non-Europeans in genetic research, to be able to "realize the full and equitable potential of PRS" (241).

The second consideration is the particularity of PRS results in psychiatry, and the complexity of communicating these results in clinic. In their thought-provoking review on the subject, Palk and colleagues(246) highlight that while many of the ethical issues which arise in cases of genetic testing in general overlap with those in psychiatric practices, "psychiatric PRS testing arguably intensifies these concerns due to the fact that is likely that the disorders it would mostly be used to predict risk for, would be those with the highest heritability, such as schizophrenia and bipolar disorder, both of which are subject to high levels of stigmatisation(247)" (246). As such, individuals (and particularly adolescents) receiving results of a genetic test indicating high polygenic risk for severe psychiatric disorders, may be at a high risk for internalized stigma and the consequences associated with negative self-labelling(246). Considering the complexity of understanding a polygenic risk result, and the nature of psychopathology in general, the communication of a high polygenic risk score to a patient is a challenge in it of itself. While the goal of communicating such a result would be to "prevent onset or mitigate severity" of the disorder, a misunderstanding of a result could be detrimental (246). For example, misinterpretation of the biological result, may lead to deterministic assumptions in which the role of polygenic risk factors is overemphasized at the expense of psycho-social and environmental factors(246). This is an important consideration because deterministic interpretations have been associated to a "sense of fatalism, decreased agency or being 'at the mercy of one's genes' or biology"(248)(246). Moreover, Palk and colleagues explain that a nuanced, and clear explanation of a high polygenic risk score, may increase motivation to implement behavioural changes that may reduce one's risk to developing the disorder in question (246). These considerations, while clearly non-exhaustive, highlight some of the challenges that clinicians may face when implementing polygenic risk scoring into psychiatric practice.

Current state of psychiatric genetics research

A field made accessible by the Psychiatric Genomics Consortium

Understanding the heritable nature of psychiatric disorders has long been a primary goal of psychiatric research. While the field was limited to family based, or twin-set cohorts until relatively recently, the last 30 years has seen an explosion in the field of psychiatric genetics. In

the early years of psychiatric genetics research, most studies focused on candidate gene approaches. This was in part due to costs of genome-wide sequencing, and the difficulty of recruiting cohorts with enough power to detect genome-wide significant effects ($1X10^{-8}$). As the first meta-analysis of the early candidate gene studies were published, researchers quickly realized that many of the exciting results, were in fact false positives(163). This was highlighted in the seminal paper by Duncan and Keller, criticizing candidate gene studies in psychiatry(163). Interestingly, a recent meta-analysis of candidate gene studies published after the publication of the Duncan and Keller paper, found that the results of the new candidate gene studies were in fact motivated by GWAS findings, and did not "provide much insight into the pathology of schizophrenia in addition to GWAS" (249). This publication, along with others, created a movement within the psychiatric genetics field that rendered traditional candidate gene studies obsolete(250). Thus, to reduce the bias introduced in the field by small scale tradition candidate gene approaches, researchers were asked to focus their efforts studying SNP that were significant in reliable GWAS studies.

The eventual shift to GWAS studies in psychiatric genetics was is in large part due to the collective efforts of the Psychiatric Genomics Consortium (PGC); a group of 800 investigators from 150+ institutions that aim to "convert family history risk factor into biologically, clinically, and therapeutically meaningful insights"(251). Considering the minimal effect sizes that common genetic polymorphisms had on the development of these complex psychiatric disorders, large adequately powered cohorts were required. Thus, through the creation of PGC in 2009, an open-access framework was established; summary statistic data from PGC GWAS studies would be accessible to researchers across the world, and genotype data would be sent to qualified scientists(251). The "open-science" perspective of the PGC has effectively created a boom in the field of psychiatric genetics, with 24 main studies, and 51 secondary analyses being published by the group in its first 8 years, notwithstanding the 150+ articles published by external researchers using PGC data(251).

To demonstrate the impact that the PGC efforts have had on the understanding of the etiology of psychiatric disease, let us briefly summarize the findings of the schizophrenia working group. Among the most cited of the PGC papers is the 2014 schizophrenia GWAS (142). This study,

which included data from 36,989 cases and 113,075 controls, identified 128 SNPs at genome wide level significance ($p < 1X10^{-8}$) mapping onto 108 independent genetic loci that are associated with schizophrenia. Notably, among the SNPs identified are alleles located in the *DRD2* (a target of all antipsychotics), and multiple genes associated with glutamatergic neurotransmission(142). Moreover, most SNPs were not associated to nonsynonymous exomic polymorphisms, suggesting that most schizophrenia associated SNPs are more likely to alter gene expression rather than affecting protein structure(142). Interestingly, this study also identified that schizophrenia SNP sets are enriched in gene enhancer regions that are associated to immune function, providing genetic evidence for the immune dysregulation hypothesis of schizophrenia (142)(252). Building on this study, the most recent identified 179 GWAS significant SNPs across 143 independent loci, of which 93 were identified in the 2014 PGC study(139). By integrating novel fine-mapping techniques, Pardiñas and colleagues were able to identify candidate causal genes for schizophrenia(108). The identification of said "causal" genes, represent potential therapeutic targets for the treatment or prevention of schizophrenia.

Another interesting application of PGC data, is the elucidation of shared genetic influence across psychiatric disorders and traits. Although mainstream conceptions of the nosology of psychiatric diseases (including the Diagnostic and Statistical Manual of Mental Disorders 5[DSM-5](34) and ICD(253)) currently classify psychiatric disorders into several distinct categories, there has been a push – supported by research across disciplines – to apply a transdiagnostic approach(254). Part of the support for this re-conceptualization of psychiatric disease is based on genetic and familial studies, which have demonstrated that heritable components of psychiatric disease overlap between disorders (255)(256).

Thus, the most recent GWAS from the PGC cross-disorder group has recently reported – in a meta-analysis including 232,964 cases and 494,162 controls – that genetic correlation analysis can identify three distinct groups of interrelated psychiatric disorders(257). Group one includes bipolar disorder (BD), major depressive disorder (MDD) and schizophrenia (SZ). The second group includes anorexia nervosa (AN), obsessive compulsive disorder (OCD) and Tourette's Syndrome (TS). Finally, group three includes autism spectrum disorder (ASD), attention-deficit and hyperactivity disorder (ADHD) and also Tourette's Syndrome (TS)(257). Groups have used the

cross-disorder GWAS data studies to perform network and pathway analyses, which have identified plausible biological pathways that have given insights into the etiology of psychiatric disease(256). Moreover, using the summary statistics from these studies, authors have created PRS for these disorders, which can be used to study genomic overlap between traits and disorders(256). Of note, are the studies demonstrating the genomic overlap between cannabis use, schizophrenia and PLE(144)(159)(150).

An increasing trend in psychiatry, and epidemiological research in general, has been to examine potential causal mediators of relationships between related phenotypes or disorders. As such, several studies, incorporating genomic datasets, have applied a variety of mediation strategies to study the direct and indirect genetic effects on phenotypes of interest. These genomic strategies include the Mendelian Randomization (MR) strategies, which use genetic variants as instrumental variables to study the relationships between exposure and outcome variables(258). These studies can infer causality if three assumptions are respected(258):

(1) the genomic variable is associated to the phenotype of interest

(2) confounders of the exposure x outcome relationship are not associated to the genomic variables

(3) The genomic variable is not associated to the outcome variable considering the exposure variable and confounders.

With the improved computing power of most standard lab computers, and the open-source nature of psychiatric GWAS, made available by the PGC, independent researchers have been able to utilize various MR methodologies to infer causal relationships between psychiatric disease and phenotypes of interest. Among the multiple methodologies in the MR field, are two-sample MR studies. A two sample MR study models data from two independent GWAS studies as distinct instrumental variables (instrumental risk factor and instrumental outcome), allowing for the analysis of the causal relationship between the two (259). These studies rely on two additional assumptions: both samples need to be from the same genetic population, and participants cannot overlap between the groups(259).

Three studies have used two-sample MR to study the relationship between schizophrenia and cannabis use(162)(144)(234). The studies use different cannabis use measures (lifetime cannabis use (162) vs ever use of cannabis(144)(234)), and have reported differing results. While Gage et al. (162) and Pasman et al. (144) reported that there is stronger evidence to support a causal link from schizophrenia risk to cannabis use (i.e those who are at risk for developing schizophrenia will use cannabis), Vaucher et al. (234) only reports a causal link between cannabis use to schizophrenia (i.e that cannabis use may cause schizophrenia). It must be noted however that Vaucher's group did not analyze the reverse relationship. As alluded to in the introduction, these MR results support the second hypothesis, i.e that the cannabis-schizophrenia association is at least partly confounded by reverse-causality(145).

Recently some have criticized the causal inference made by MR studies, due to the confounding effects of pleiotropic mechanisms. Genetic pleiotropy refers to the notion that certain genetic markers or variants have effects of more than one phenotype. We can separate pleiotropic mechanisms into two distinct groups: horizontal pleiotropy and vertical pleiotropy. Horizontal pleiotropy (or biological pleiotropy) occurs when a variant or gene produces biological effects on multiple phenotypes(260). Vertical pleiotropy is when a gene or variant has causal influence on one trait, which in turn has a causal effect on a second trait(260). While MR depends on vertical pleiotropy, as genetic markers can influence downstream mechanisms, the presence of horizontal pleiotropy will cause a violation of the assumptions described above; the effects of genetic variants on the outcome would not exclusively be through the risk factor variable(259). Although multiple statistical techniques have been developed to account for horizontal pleiotropy, such as MR egger regression(261), the causal conclusions of MR studies may still be still be erroneous(235). As such, a novel methodology recently proposed – latent causal variable (LCV) models – claims to overcome the disadvantages of MR (235). In this method, "genetic correlation between two traits is mediated by a latent variable having a causal effect on each trait"(235). Accordingly, the causal link between schizophrenia and lifetime cannabis use was studied using LCV(262). This study found no evidence for a causal link between schizophrenia and lifetime cannabis use through an unobserved variable (262). Applying the authors findings to the question of the cannabis-schizophrenia association, we can conclude that the genetic correlation

commonly reported between cannabis use and schizophrenia are more likely to arise from horizontal pleiotropy than a causal pathway; common genetic variants influence the development of both disorders. Furthermore, as argued above, these reports, along with the findings reported in article 2, do not preclude the possibility of a causal mechanism linking cannabis use and psychosis. Rather, they suggest that the relationship between psychosis risk and cannabis use may be linked via another environmental mediator, instead of being linked through common genetic mechanisms. As such, these results call for the search of possible environmental and psychological mediators of the link between cannabis use and psychosis, such as personality traits, social isolation and the influence peer groups.

A roadmap for future studies

Hypothesis-based methods in the era of modern psychiatric genomics: Biologically informed polygenic risk scores

Bearing in mind the findings from study one, in the context of the rodent literature, it is reasonable to hypothesize that a gene by environment interaction of endocannabinoid related genetic markers on anxiety will have an effect on alcohol behaviour in humans. But the lack of reliable endocannabinoid related GWAS results in cohorts of alcohol users (or general substance abusers) makes the study of this hypothesis difficult, in the era of modern psychiatric genomics. One conceivable solution would be the application of “biologically informed” polygenic risk scores recently proposed and effectively verified by Dass and colleagues(263). In this method, one would create a “polygenic risk score” that is based on gene networks expressed in biologically relevant brain tissues, rather than a PRS constructed from statistically significant SNPs from GWAS studies (263). In their work, Daas and collegues show that a biologically informed polygenic risk scores (ePRS) can outperform traditional PRS in the prediction of certain endophenotypes. Particularly, they demonstrate that the ePRS for gene networks associated to insulin receptor expression in the mesolimbic areas (striatum and prefrontal cortex) or the hippocampus, predict impulsivity endophenotypes, addiction, and Alzheimer’s disease better than the respective traditional PRS. This is interesting considering that the insulin receptor, and insulin metabolism, is involved in the processes of reward(264), impulsivity(265), mood(266), cognition(267), and memory(268). The

SNPs used in the ePRS were not statistically significant in the GWAS for their respective diseases, and none of the SNPs could predict the outcomes studied alone(263). The authors therefore argue that their ePRS represents the effects of gene networks rather than the aggregated effects of individual, seemingly unrelated SNPs(263). Applying this logic, I propose a study – or collection of studies – that would look at the relationship between the ePRS representing the gene network associated to the expression of *CNR1 – MGLL – DAGLA* in the central nucleus of the amygdala and alcohol abuse outcomes in a cohort of adolescent individuals. Moreover, we could examine if the anxiety scores, biological markers of chronic stress, and other affective phenotypes could mediate this relationship. Thus, through the application of novel psychiatric genomic techniques, we could continue to answer biologically informed hypothesis-based questions in this field.

Translating psychiatric genomics research to clinical settings

Two stimulating articles, outlining the clinical utility of polygenic risk scoring, have recently been published by the group of Naomi Wray and Graham Murray in JAMA Psychiatry(140)(269). In these articles, the authors describe current and future clinical applications of polygenic risk scoring, and also outline the limitations of these methods. They describe the possibility of using PRS for the stratification of individuals as a part of populational screening programs, and the utility of using PRS in the guidance of clinical decision making(269). The authors argue that one potential application of PRS in clinical psychiatry could be in the diagnosis of young help seeking individuals(140). Considering that the early phases of severe psychiatric diseases are often characterized by a continuum of unspecific affective symptoms tools that could differentiate between disorders would be of great clinical value, but due to the high genetic correlation between psychiatric disorders(255)(256)(257), PRS cannot currently contribute to the differential diagnosis of disease(140). Rather, PRS could be useful in the early identification of severe illness trajectories, by identifying individuals with a high polygenic risk load instead of diagnosing a particular disease(140). This is a useful application, as it would help clinicians in their orientation of at-risk individuals, i.e. sending high PRS patients to a mental health prevention program versus usual care(140). Along these lines, one relevant future study would be to examine if the PRS from the main PGC GWAS studies could predict general psychopathology and mental health outcomes in clinical cohorts, such as future hospitalizations, global functioning, and suicidality. This could

be achieved by creating a risk calculator with multiple known risk factors for general mental illness (including substance abuse measures) and including PRS into the calculation. Similar studies have already been attempted in clinical-high risk for psychosis cohorts, using the PRS-Sz, with relative success(270). The results of the second study presented in this thesis provide evidence that the inclusion of PRS-Sz within a clinical risk calculator is logical. Our results suggest that the risk conferred by PRS-Sz and cannabis use onto PLE are the result of an additive effect rather than that of an interaction effect. Thus, any future study examining the utility of a clinical risk calculator for the prediction of PLE could include cannabis use as well as PRS-Sz, among other known independent risk factors for conversion to psychosis. Another interesting avenue is the use of PRS in the prediction of treatment response, and while cohorts currently exist for the detection of genomic predictors (including PRS) of treatment response to lithium(271), and anti-depressant medications(272)(273), Murray and colleagues note that the current limitation of these studies is lack of populational power, rather than fundamental issues with PRS(140). Overall, the literature shows that PRS show promise as a reliable tool for the prediction of various psychiatric phenotypes, when combined with other reliable predictive measures. Nevertheless, at the time of the composition of this work, polygenic risk score tools are not ready to be deployed in psychiatric clinical practice just yet. Moreover, they ought to be used judiciously and always in consideration of the ethical challenges facing their application and interpretation.

Bibliography

1. Long T, Wagner M, Demske D, Leipe C, Tarasov PE. Cannabis in Eurasia: origin of human use and Bronze Age trans-continental connections. Veget Hist Archaeobot. 2017 Mar 1;26(2):245–58.

2. Canada H. Canadian Tobacco Alcohol and Drugs (CTADS): 2015 summary [Internet]. aem. 2017 [cited 2019 Apr 6]. Available from: https://www.canada.ca/en/health-canada/services/canadian-tobacco-alcohol-drugs-survey/2015-summary.html

3. Québec I de la S du. Enquête québécoise sur le cannabis 2019 [Internet]. [cited 2020 Oct 14]. Available from: https://www.stat.gouv.qc.ca/statistiques/sante/cannabis-2019.html

4. Di Marzo V. A brief history of cannabinoid and endocannabinoid pharmacology as inspired by the work of British scientists. Trends Pharmacol Sci. 2006 Mar;27(3):134–40.

5. Freund TF, Katona I, Piomelli D. Role of endogenous cannabinoids in synaptic signaling. Physiol Rev. 2003 Jul;83(3):1017–66.

6. Herkenham M, Lynn AB, Little MD, Johnson MR, Melvin LS, de Costa BR, et al. Cannabinoid receptor localization in brain. Proc Natl Acad Sci USA. 1990 Mar;87(5):1932–6.

7. Gaoni Y, Mechoulam R. Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish. J Am Chem Soc. 1964 Apr 1;86(8):1646–7.

8. Devane WA, Dysarz FA, Johnson MR, Melvin LS, Howlett AC. Determination and characterization of a cannabinoid receptor in rat brain. Mol Pharmacol. 1988 Nov 1;34(5):605–13.

9. Munro S, Thomas KL, Abu-Shaar M. Molecular characterization of a peripheral receptor for cannabinoids. Nature. 1993 Sep 2;365(6441):61–5.

10. Howlett AC, Barth F, Bonner TI, Cabral G, Casellas P, Devane WA, et al. International Union of Pharmacology. XXVII. Classification of cannabinoid receptors. Pharmacol Rev. 2002 Jun;54(2):161–202.

11. Di Marzo V, De Petrocellis L. Why do cannabinoid receptors have more than one endogenous ligand? Philos Trans R Soc Lond, B, Biol Sci. 2012 Dec 5;367(1607):3216–28.

12. Onaivi ES. Commentary: Functional Neuronal CB2 Cannabinoid Receptors in the CNS. Curr Neuropharmacol. 2011 Mar;9(1):205–8.

13. Pacher P, Mechoulam R. Is lipid signaling through cannabinoid 2 receptors part of a protective system? Prog Lipid Res. 2011 Apr;50(2):193–211.

14. Zhao P, Abood ME. GPR55 and GPR35 and their relationship to cannabinoid and lysophospholipid receptors. Life Sciences. 2013 Mar 19;92(8):453–7.

15. Stelt M van der, Marzo VD. Endovanilloids. Euro Jour of Biochem. 2004;271(10):1827–34.

16. Di Marzo V. Endocannabinoid signaling in the brain: biosynthetic mechanisms in the limelight. Nat Neurosci. 2011 Jan;14(1):9–15.

17. Lu H-C, Mackie K. Review of the Endocannabinoid System. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2016 Apr;79(7):516-525.

18. Meyer HC, Lee FS, Gee DG. The Role of the Endocannabinoid System and Genetic Variation in Adolescent Brain Development. Neuropsychopharmacology. 2018 Jan;43(1):21–33.

19. Ohno-Shosaku T, Kano M. Endocannabinoid-mediated retrograde modulation of synaptic transmission. Current Opinion in Neurobiology. 2014 Dec 1;29:1–8.

20. Allegret S. The history of hemp. Hemp: industrial production and uses. 2013;4–26.

21. Hanuš LO, Meyer SM, Muñoz E, Taglialatela-Scafati O, Appendino G. Phytocannabinoids: a unified critical inventory. Nat Prod Rep. 2016 Nov 23;33(12):1357–92.

22. Pertwee RG. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ9-tetrahydrocannabinol, cannabidiol and Δ9-tetrahydrocannabivarin. British Journal of Pharmacology. 2008;153(2):199–215.

23. Lucas CJ, Galettis P, Schneider J. The pharmacokinetics and the pharmacodynamics of cannabinoids. British Journal of Clinical Pharmacology. 2018;84(11):2477–82.

24. Mato S, Robbe D, Puente N, Grandes P, Manzoni OJ. Presynaptic Homeostatic Plasticity Rescues Long-Term Depression after Chronic Δ9-Tetrahydrocannabinol Exposure. J Neurosci. 2005 Dec 14;25(50):11619–27.

25. Hoffman AF, Oz M, Yang R, Lichtman AH, Lupica CR. Opposing actions of chronic Δ9-tetrahydrocannabinol and cannabinoid antagonists on hippocampal long-term potentiation. Learn Mem. 2007;14(1–2):63–74.

26. Colizzi M, McGuire P, Pertwee RG, Bhattacharyya S. Effect of cannabis on glutamate signalling in the brain: A systematic review of human and animal evidence. Neuroscience & Biobehavioral Reviews. 2016 May 1;64:359–81.

27. López-Moreno JA, González-Cuevas G, Moreno G, Navarro M. The pharmacology of the endocannabinoid system: functional and structural interactions with other neurotransmitter systems and their repercussions in behavioral addiction. Addiction Biology. 2008;13(2):160–87.

28. ElSohly MA, Mehmedic Z, Foster S, Gon C, Chandra S, Church JC. Changes in Cannabis Potency over the Last Two Decades (1995-2014) - Analysis of Current Data in the United States. Biol Psychiatry. 2016 Apr 1;79(7):613–9.

29. Bukiya AN. Physiology of the Endocannabinoid System During Development. In: Bukiya AN, editor. Recent Advances in Cannabinoid Physiology and Pathology. Springer Internation Publishing. 2019:13-37.

30. Mechoulam R, Parker LA. The endocannabinoid system and the brain. Annu Rev Psychol. 2013;64:21–47.

31. Volkow ND, Baler RD, Compton WM, Weiss SRB. Adverse Health Effects of Marijuana Use. New England Journal of Medicine. 2014 Jun 5;370(23):2219–27.

32. Gobbi G, Atkin T, Zytynski T, Wang S, Askari S, Boruff J, et al. Association of Cannabis Use in Adolescence and Risk of Depression, Anxiety, and Suicidality in Young Adulthood: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2019 Apr 1;76(4):426–34.

33. Jacobus J, Tapert SF. Effects of Cannabis on the Adolescent Brain. Curr Pharm Des. 2014;20(13):2186–93.

34. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013. 1520 p.

35. Leshner AI. Addiction Is a Brain Disease, and It Matters. Science. 1997 Oct 3;278(5335):45–7.

36. Koob GF, Volkow ND. Neurocircuitry of Addiction. Neuropsychopharmacology. 2010 Jan;35(1):217–38.

37. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry. 2016 Aug;3(8):760–73.

38. Volkow ND, Koob GF, McLellan AT. Neurobiologic Advances from the Brain Disease Model of Addiction. New England Journal of Medicine. 2016 Jan 28;374(4):363–71.

39. Canada H. Addiction to cannabis [Internet]. aem. 2018 [cited 2021 Mar 27]. Available from: https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/health-effects/addiction.html

40. Ferland J-MN, Hurd YL. Deconstructing the neurobiology of cannabis use disorder. Nature Neuroscience. 2020 May;23(5):600–10.

41. Parsons LH, Hurd YL. Endocannabinoid signalling in reward and addiction. Nat Rev Neurosci. 2015 Oct;16(10):579–94.

42. Bloomfield MAP, Ashok AH, Volkow ND, Howes OD. The effects of Δ9-tetrahydrocannabinol on the dopamine system. Nature. 2016 Nov 17;539(7629):369–77.

43. Logrip ML, Koob GF, Zorrilla EP. Role of Corticotropin-Releasing Factor in Drug Addiction: Potential for Pharmacological Intervention. CNS Drugs. 2011 Apr;25(4):271–87.

44. Cousijn J, Wiers RW, Ridderinkhof KR, van den Brink W, Veltman DJ, Goudriaan AE. Grey matter alterations associated with cannabis use: results of a VBM study in heavy cannabis users and healthy controls. Neuroimage. 2012 Feb 15;59(4):3845–51.

45. Koenders L, Cousijn J, Vingerhoets WAM, van den Brink W, Wiers RW, Meijer CJ, et al. Grey Matter Changes Associated with Heavy Cannabis Use: A Longitudinal sMRI Study. PLoS One. 2016;11(5):e0152482.

46. D'Souza DC, Cortes-Briones JA, Ranganathan M, Thurnauer H, Creatura G, Surti T, et al. Rapid Changes in CB1 Receptor Availability in Cannabis Dependent Males after Abstinence from Cannabis. Biol Psychiatry Cogn Neurosci Neuroimaging. 2016 Jan 1;1(1):60–7.

47. Parsons LH, Hurd YL. Endocannabinoid signalling in reward and addiction. Nature Reviews Neuroscience. 2015 Oct;16(10):579–94.

48. Silvestri C, Di Marzo V. The endocannabinoid system in energy homeostasis and the etiopathology of metabolic disorders. Cell Metab. 2013 Apr 2;17(4):475–90.

49. Klein C, Hill MN, Chang SCH, Hillard CJ, Gorzalka BB. Circulating endocannabinoid concentrations and sexual arousal in women. J Sex Med. 2012 Jun;9(6):1588–601.

50. Trezza V, Baarendse PJJ, Vanderschuren LJMJ. The pleasures of play: pharmacological insights into social reward mechanisms. Trends Pharmacol Sci. 2010 Oct;31(10):463–9.

51. Solinas M, Justinova Z, Goldberg SR, Tanda G. Anandamide administration alone and after inhibition of fatty acid amide hydrolase (FAAH) increases dopamine levels in the nucleus accumbens shell in rats. J Neurochem. 2006 Jul;98(2):408–19.

52. Alvarez-Jaimes L, Polis I, Parsons LH. Regional Influence of Cannabinoid CB1 Receptors in the Regulation of Ethanol Self-Administration by Wistar Rats. Open Neuropsychopharmacol J. 2009;2:77–85.

53. Basavarajappa BS, Ninan I, Arancio O. Acute ethanol suppresses glutamatergic neurotransmission through endocannabinoids in hippocampal neurons. J Neurochem. 2008 Nov;107(4):1001–13.

54. Ferrer B, Bermúdez-Silva FJ, Bilbao A, Alvarez-Jaimes L, Sanchez-Vera I, Giuffrida A, et al. Regulation of brain anandamide by acute administration of ethanol. Biochemical Journal. 2007 Feb;404(1):97–104.

55. Rubio M, McHugh D, Fernández-Ruiz J, Bradshaw H, Walker JM. Short-term exposure to alcohol in rats affects brain levels of anandamide, other N-acylethanolamines and 2-arachidonoyl-glycerol. Neurosci Lett. 2007 Jun;421(3):270–4.

56. Hutchison KE, Haughey H, Niculescu M, Schacht J, Kaiser A, Stitzel J, et al. The Incentive Salience of Alcohol. Arch Gen Psychiatry. 2008 Jul;65(7):841–50.

57. Stinson FS, Ruan WJ, Pickering R, Grant BF. Cannabis use disorders in the USA : prevalence, correlates and co-morbidity. Psychol med (Print). 2006;36(10):1447–60.

58. Fischer AS, Tapert SF, Louie DL, Schatzberg AF, Singh MK. Cannabis and the Developing Adolescent Brain. Curr Treat Options Psych. 2020 Jun 1;7(2):144–61.

59. Silins E, Horwood LJ, Patton GC, Fergusson DM, Olsson CA, Hutchinson DM, et al. Young adult sequelae of adolescent cannabis use: an integrative analysis. Lancet Psychiatry. 2014 Sep;1(4):286–93.

60. Canada H. Canadian Tobacco Alcohol and Drugs (CTADS) Survey: 2017 summary [Internet]. aem. 2018 [cited 2020 Oct 14]. Available from: https://www.canada.ca/en/health-canada/services/canadian-tobacco-alcohol-drugs-survey/2017-summary.html

61. Rehm J, Shield KD. Global Burden of Alcohol Use Disorders and Alcohol Liver Disease. Biomedicines. 2019 Dec;7(4):99.

62. Franklin A, Chrystoja BR, Manthey J, Rehm J, Shield K. The Alcohol-Attributable Burden of Disease in Canada from 2000 to 2016. Canadian Journal of Addiction. 2020 Dec;11(4):6–12.

63. Schwarzinger M, Thiébaut SP, Baillot S, Mallet V, Rehm J. Alcohol use disorders and associated chronic disease – a national retrospective cohort study from France. BMC Public Health. 2017 Jul 21;18(1):43.

64. Subbaraman MS, Kerr WC. Simultaneous vs. concurrent use of alcohol and cannabis in the National Alcohol Survey. Alcohol Clin Exp Res. 2015 May;39(5):872–9.

65. Weinberger AH, Platt J, Goodwin RD. Is cannabis use associated with an increased risk of onset and persistence of alcohol use disorders? A three-year prospective study among adults in the United States. Drug Alcohol Depend. 2016 Apr 1;161:363–7.

66. Hasin DS, Kerridge BT, Saha TD, Huang B, Pickering R, Smith SM, et al. Prevalence and Correlates of DSM-5 Cannabis Use Disorder, 2012-2013: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions-III. Am J Psychiatry. 2016 01;173(6):588–99.

67. Blanco C, Hasin DS, Wall MM, Flórez-Salamanca L, Hoertel N, Wang S, et al. Cannabis Use and Risk of Psychiatric Disorders: Prospective Evidence From a US National Longitudinal Study. JAMA Psychiatry. 2016 Apr;73(4):388–95.

68. Arseneault L, Cannon M, Poulton R, Murray R, Caspi A, Moffitt TE. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. BMJ. 2002 Nov 23;325(7374):1212–3.

69. van Os J, Bak M, Hanssen M, Bijl RV, de Graaf R, Verdoux H. Cannabis use and psychosis: a longitudinal population-based study. Am J Epidemiol. 2002 Aug 15;156(4):319–27.

70. Zammit S, Allebeck P, Andreasson S, Lundberg I, Lewis G. Self reported cannabis use as a risk factor for schizophrenia in Swedish conscripts of 1969: historical cohort study. BMJ. 2002 Nov 23;325(7374):1199.

71. Fergusson DM, Horwood LJ, Swain-Campbell NR. Cannabis dependence and psychotic symptoms in young people. Psychol Med. 2003 Jan;33(1):15–21.

72. Mustonen A, Niemelä S, Nordström T, Murray GK, Mäki P, Jääskeläinen E, et al. Adolescent cannabis use, baseline prodromal symptoms and the risk of psychosis. The British Journal of Psychiatry. 2018 Apr;212(4):227–33.

73. Szoke A, Galliot A-M, Richard J-R, Ferchiou A, Baudin G, Leboyer M, et al. Association between cannabis use and schizotypal dimensions – A meta-analysis of cross-sectional studies. Psychiatry Research. 2014 Sep 30;219(1):58–66.

74. Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, et al. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. The Lancet. 2007 Jul 28;370(9584):319–28.

75. Semple DM, McIntosh AM, Lawrie SM. Cannabis as a risk factor for psychosis: systematic review: J Psychopharmacol. 2005 Mar;19(2):187-194.

76. Marconi A, Di Forti M, Lewis CM, Murray RM, Vassos E. Meta-analysis of the Association Between the Level of Cannabis Use and Risk of Psychosis. Schizophr Bull. 2016 Sep;42(5):1262–9.

77. Hindley G, Beck K, Borgan F, Ginestet CE, McCutcheon R, Kleinloog D, et al. Psychiatric symptoms caused by cannabis constituents: a systematic review and meta-analysis. The Lancet Psychiatry. 2020 Apr 1;7(4):344–53.

78. Forti MD, Quattrone D, Freeman TP, Tripoli G, Gayer-Anderson C, Quigley H, et al. The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study. The Lancet Psychiatry. 2019 May 1;6(5):427–36.

79. Galvez-Buccollini JA, Proal AC, Tomaselli V, Trachtenberg M, Coconcea C, Chun J, et al. Association between age at onset of psychosis and age at onset of cannabis use in non-affective psychosis. Schizophr Res. 2012 Aug;139(1–3):157–60.

80. Manrique-Garcia E, Zammit S, Dalman C, Hemmingsson T, Andreasson S, Allebeck P. Prognosis of schizophrenia in persons with and without a history of cannabis use. Psychological Medicine. 2014 Sep;44(12):2513–21.

81. Fusar-Poli P. The Clinical High-Risk State for Psychosis (CHR-P), Version II. Schizophr Bull. 2017 Jan 1;43(1):44–7.

82. Farris MS, Shakeel MK, Addington J. Cannabis use in individuals at clinical high-risk for psychosis: a comprehensive review. Soc Psychiatry Psychiatr Epidemiol. 2020 May 1;55(5):527–37.

83. Oliver D, Reilly TJ, Baccaredda Boy O, Petros N, Davies C, Borgwardt S, et al. What Causes the Onset of Psychosis in Individuals at Clinical High Risk? A Meta-analysis of Risk and Protective Factors. Schizophr Bull. 2020 Jan 4;46(1):110–20.

84. Kraan T, Velthorst E, Koenders L, Zwaart K, Ising HK, van den Berg D, et al. Cannabis use and transition to psychosis in individuals at ultra-high risk: review and meta-analysis. Psychol Med. 2016 Mar;46(4):673–81.

85. Ronald A, Sieradzka D, Cardno AG, Haworth CMA, McGuire P, Freeman D. Characterization of psychotic experiences in adolescence using the specific psychotic experiences questionnaire: findings from a study of 5000 16-year-old twins. Schizophr Bull. 2014 Jul;40(4):868–77.

86. Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. Psychol Med. 2013 Jun;43(6):1133–49.

87. Bourgin J, Tebeka S, Mallet J, Mazer N, Dubertret C, Le Strat Y. Prevalence and correlates of psychotic-like experiences in the general population. Schizophrenia Research. 2020 Jan 1;215:371–7.

88. Healy C, Brannigan R, Dooley N, Coughlan H, Clarke M, Kelleher I, et al. Childhood and adolescent psychotic experiences and risk of mental disorder: a systematic review and meta-analysis. Psychological Medicine. 2019 Jul;49(10):1589–99.

89. Werbeloff N, Drukker M, Dohrenwend BP, Levav I, Yoffe R, Os J van, et al. Self-reported Attenuated Psychotic Symptoms as Forerunners of Severe Mental Disorders Later in Life. Arch Gen Psychiatry. 2012 May 1;69(5):467–75.

90. Kelleher I, Cederlöf M, Lichtenstein P. Psychotic experiences as a predictor of the natural course of suicidal ideation: a Swedish cohort study. World Psychiatry. 2014;13(2):184–8.

91. van Os J, Reininghaus U. Psychosis as a transdiagnostic and extended phenotype in the general population. World Psychiatry. 2016 Jun;15(2):118–24.

92. Fusar-Poli P, Borgwardt S, Bechdolf A, Addington J, Riecher-Rössler A, Schultze-Lutter F, et al. The Psychosis High-Risk State: A Comprehensive State-of-the-Art Review. JAMA Psychiatry. 2013 Jan 1;70(1):107–20.

93. Fusar-Poli P, Cappucciati M, Borgwardt S, Woods SW, Addington J, Nelson B, et al. Heterogeneity of Psychosis Risk Within Individuals at Clinical High Risk: A Meta-analytical Stratification. JAMA Psychiatry. 2016 Feb;73(2):113–20.

94. DeRosse P, Karlsgodt KH. Examining the Psychosis Continuum. Curr Behav Neurosci Rep. 2015 May 1;2(2):80–9.

95. Cowan HR, Mittal VA. Three types of psychotic-like experiences in youth at clinical high risk for psychosis. Eur Arch Psychiatry Clin Neurosci. 2020 May.

96. Nelson MT, Seal ML, Pantelis C, Phillips LJ. Evidence of a dimensional relationship between schizotypy and schizophrenia: a systematic review. Neurosci Biobehav Rev. 2013 Mar;37(3):317–27.

97. Yung AR, Yung AR, Pan Yuen H, Mcgorry PD, Phillips LJ, Kelly D, et al. Mapping the Onset of Psychosis: The Comprehensive Assessment of At-Risk Mental States. Aust N Z J Psychiatry. 2005 Nov 1;39(11–12):964–71.

98. Miller TJ, McGlashan TH, Rosen JL, Somjee L, Markovich PJ, Stein K, et al. Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: preliminary evidence of interrater reliability and predictive validity. Am J Psychiatry. 2002 May;159(5):863–5.

99. Stefanis NC, Hanssen M, Smirnis NK, Avramopoulos DA, Evdokimidis IK, Stefanis CN, et al. Evidence that three dimensions of psychosis have a distribution in the general population. Psychol Med. 2002 Feb;32(2):347–58.

100. Os J van, Reininghaus U. Psychosis as a transdiagnostic and extended phenotype in the general population. World Psychiatry. 2016;15(2):118–24.

101. Mossaheb N, Becker J, Schaefer MR, Klier CM, Schloegelhofer M, Papageorgiou K, et al. The Community Assessment of Psychic Experience (CAPE) questionnaire as a screening-instrument in the detection of individuals at ultra-high risk for psychosis. Schizophrenia Research. 2012 Nov 1;141(2):210–4.

102. Addington J, Stowkowy J, Weiser M. Screening tools for clinical high risk for psychosis. Early Intervention in Psychiatry. 2015;9(5):345–56.

103. Bukenaite A, Stochl J, Mossaheb N, Schäfer MR, Klier CM, Becker J, et al. Usefulness of the CAPE-P15 for detecting people at ultra-high risk for psychosis: Psychometric properties and cut-off values. Schizophrenia Research. 2017 Nov 1;189:69–74.

104. Schubart CD, Gastel WA van, Breetvelt EJ, Beetz SL, Ophoff RA, Sommer IEC, et al. Cannabis use at a young age is associated with psychotic experiences. Psychological Medicine. 2011 Jun;41(6):1301–10.

105. van Gastel WA, Wigman JTW, Monshouwer K, Kahn RS, van Os J, Boks MPM, et al. Cannabis use and subclinical positive psychotic experiences in early adolescence: findings from a Dutch survey. Addiction. 2012 Feb;107(2):381–7.

106. Ragazzi TCC, Shuhama R, Menezes PR, Del-Ben CM. Cannabis use as a risk factor for psychotic-like experiences: A systematic review of non-clinical populations evaluated with the Community Assessment of Psychic Experiences. Early Interv Psychiatry. 2018;12(6):1013–23.

107. Bourque J, Afzali MH, Conrod PJ. Association of Cannabis Use With Adolescent Psychotic Symptoms. JAMA Psychiatry. 2018 Aug 1;75(8):864–6.

108. Jones HJ, Gage SH, Heron J, Hickman M, Lewis G, Munafò MR, et al. Association of Combined Patterns of Tobacco and Cannabis Use in Adolescence With Psychotic Experiences. JAMA Psychiatry. 2018 Mar 1;75(3):240–6.

109. Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. Molecular Psychiatry. 2019 Mar;24(3):409–20.

110. Luxenburger H. Vorläufiger Bericht über psychiatrische Serienuntersuchungen an Zwillingen. Z f d g Neur u Psych. 1928 Dec 1;116(1):297–326.

111. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. AJP. 2019 Mar 8;176(5):376–87.

112. Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC. No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. Biological Psychiatry. 2017 Nov 15;82(10):702–8.

113. Duncan LE, Ostacher M, Ballon J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology. 2019 Aug;44(9):1518–23.

114. Edenberg HJ, Gelernter J, Agrawal A. Genetics of Alcoholism. Curr Psychiatry Rep. 2019 Mar 9;21(4):26.

115. Storr M, Emmerdinger D, Diegelmann J, Pfennig S, Ochsenkühn T, Göke B, et al. The Cannabinoid 1 Receptor (CNR1) 1359 G/A Polymorphism Modulates Susceptibility to Ulcerative Colitis and the Phenotype in Crohn's Disease. PLOS ONE. 2010 Feb 26;5(2):e9453.

116. Mota N, Sumner JA, Lowe SR, Neumeister A, Uddin M, Aiello AE, et al. The rs1049353 polymorphism in the CNR1 gene interacts with childhood abuse to predict posttraumatic threat symptoms. J Clin Psychiatry. 2015 Dec;76(12):e1622–3.

117. López-Moreno JA, Echeverry-Alzate V, Bühler K-M. The genetic basis of the endocannabinoid system and drug addiction in humans. J Psychopharmacol. 2012 Jan 1;26(1):133–43.

118. Schmidt LG, Samochowiec J, Finckh U, Fiszer-Piosik E, Horodnicki J, Wendel B, et al. Association of a CB1 cannabinoid receptor gene (CNR1) polymorphism with severe alcohol dependence. Drug Alcohol Depend. 2002 Feb 1;65(3):221–4.

119. Ehlers CL, Slutske WS, Lind PA, Wilhelmsen KC. Association between single nucleotide polymorphisms in the cannabinoid receptor gene (CNR1) and impulsivity in southwest California Indians. Twin Res Hum Genet. 2007 Dec;10(6):805–11.

120. Proudnikov D, Kroslak T, Sipe J, Randesi M, Li D, Hamon S, et al. Association of polymorphisms of the cannabinoid receptor (CNR1) and fatty acid amide hydrolase (FAAH) genes with heroin addiction: impact of long repeats of CNR1. Pharmacogenomics J . 2010 Jun;10(3):232-242.

121. Herman AI, Kranzler HR, Cubells JF, Gelernter J, Covault J. Association study of the CNR1 gene exon 3 alternative promoter region polymorphisms and substance dependence. Am J Med Genet B Neuropsychiatr Genet. 2006 Jul 5;141B(5):499–503.

122. Hartman CA, Hopfer CJ, Haberstick B, Rhee SH, Crowley TJ, Corley RP, et al. The association between cannabinoid receptor 1 gene (CNR1) and cannabis dependence symptoms in adolescents and young adults. Drug Alcohol Depend. 2009 Sep 1;104(1–2):11–6.

123. Agrawal A, Wetherill L, Dick DM, Xuei X, Hinrichs A, Hesselbrock V, et al. Evidence for association between polymorphisms in the cannabinoid receptor 1 (CNR1) gene and cannabis dependence. Am J Med Genet B Neuropsychiatr Genet. 2009 Jul 5;150B(5):736–40.

124. Marcos M, Pastor I, de la Calle C, Barrio-Real L, Laso F-J, González-Sarmiento R. Cannabinoid receptor 1 gene is associated with alcohol dependence. Alcohol Clin Exp Res. 2012 Feb;36(2):267–71.

125. Benyamina A, Kebir O, Blecha L, Reynaud M, Krebs M-O. CNR1 gene polymorphisms in addictive disorders: a systematic review and a meta-analysis. Addict Biol. 2011 Jan;16(1):1–6.

126. Sipe JC, Chiang K, Gerber AL, Beutler E, Cravatt BF. A missense mutation in human fatty acid amide hydrolase associated with problem drug use. Proc Natl Acad Sci USA. 2002 Jun 11;99(12):8394–9.

127. Filbey FM, Schacht JP, Myers US, Chavez RS, Hutchison KE. Individual and Additive Effects of the CNR1 and FAAH Genes on Brain Response to Marijuana Cues. Neuropsychopharmacology. 2010 Mar;35(4):967–75.

128. Melroy-Greif WE, Wilhelmsen KC, Ehlers CL. Genetic variation in FAAH is associated with cannabis use disorders in a young adult sample of Mexican Americans. Drug Alcohol Depend. 2016 01;166:249–53.

129. Sloan ME, Gowin JL, Yan J, Schwandt ML, Spagnolo PA, Sun H, et al. Severity of alcohol dependence is associated with the fatty acid amide hydrolase Pro129Thr missense variant. Addict Biol. 2018 Jan;23(1):474–84.

130. Iwasaki S, Ishiguro H, Higuchi S, Onaivi ES, Arinami T. Association study between alcoholism and endocannabinoid metabolic enzyme genes encoding fatty acid amide hydrolase and monoglyceride lipase in a Japanese population. Psychiatr Genet. 2007 Aug;17(4):215–20.

131. Muldoon PP, Chen J, Harenza JL, Abdullah RA, Sim-Selley LJ, Cravatt BF, et al. Inhibition of monoacylglycerol lipase reduces nicotine withdrawal. Br J Pharmacol. 2015 Feb;172(3):869–82.

132. Carey CE, Agrawal A, Zhang B, Conley ED, Degenhardt L, Heath AC, et al. Monoacylglycerol lipase (MGLL) polymorphism rs604300 interacts with childhood adversity to predict cannabis dependence symptoms and amygdala habituation: Evidence from an endocannabinoid system-level analysis. J Abnorm Psychol. 2015 Nov;124(4):860–77.

133. Ishiguro H, Higuchi S, Arinami T, Onaivi ES. Association between alcoholism and the gene encoding the endocannabinoid synthesizing enzyme diacylglycerol lipase alpha in the Japanese population. Alcohol. 2018 May 1;68:59–62.

134. Sullivan PF. The Psychiatric GWAS Consortium: Big Science Comes to Psychiatry. Neuron. 2010 Oct 21;68(2):182–6.

135. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nature Protocols. 2020 Sep;15(9):2759–72.

136. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. PLoS Genet. 2013 Mar;9(3).

137. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016 May;40(4):268–72.

138. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics. 2010 Jul;42(7):565–9.

139. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018 Mar;50(3):381–9.

140. Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR. Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. JAMA Psychiatry. 2021 Feb; 78(2): 210-219.

141. Mullins N, Bigdeli TB, Børglum AD, Coleman JRI, Demontis D, Mehta D, et al. GWAS of Suicide Attempt in Psychiatric Disorders and Association With Major Depression Polygenic Risk Scores. AJP. 2019 Jun 5;176(8):651–60.

142. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014 Jul 24;511(7510):421–7.

143. Legge SE, Jones HJ, Kendall KM, Pardiñas AF, Menzies G, Bracher-Smith M, et al. Association of Genetic Liability to Psychotic Experiences With Neuropsychotic Disorders and Traits. JAMA Psychiatry. 2019 Dec 1;76(12):1256–65.

144. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci. 2018;21(9):1161–70.

145. Gillespie NA, Kendler KS. Use of Genetically Informed Methods to Clarify the Nature of the Association Between Cannabis Use and Risk for Schizophrenia. JAMA Psychiatry. 2020 Nov.

146. Hiemstra M, Nelemans SA, Branje S, van Eijk KR, Hottenga J-J, Vinkers CH, et al. Genetic vulnerability to schizophrenia is associated with cannabis use patterns during adolescence. Drug Alcohol Depend. 2018 01;190:143–50.

147. Hartz SM, Horton AC, Oehlert M, Carey CE, Agrawal A, Bogdan R, et al. Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. Biological Psychiatry. 2017 Nov 15;82(10):709–15.

148. Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, et al. Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. Front Genet. 2016 Aug.

149. Jones HJ, Hammerton G, McCloud T, Hines LA, Wright C, Gage SH, et al. Examining pathways between genetic liability for schizophrenia and patterns of tobacco and cannabis use in adolescence. Psychol Med. 2020 Jun 9;1–8.

150. Barkhuizen W, Pain O, Dudbridge F, Ronald A. Genetic overlap between psychotic experiences in the community across age and with psychiatric disorders. Translational Psychiatry. 2020 Mar 9;10(1):1–12.

151. Guloksuz S, Pries L-K, Delespaul P, Kenis G, Luykx JJ, Lin BD, et al. Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. World Psychiatry. 2019;18(2):173–82.

152. Hjorthøj C, Uddin MJ, Wimberley T, Dalsgaard S, Hougaard DM, Børglum A, et al. No evidence of associations between genetic liability for schizophrenia and development of cannabis use disorder. Psychological Medicine. undefined/ed;1–6.

153. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci. 2018 Sep;21(9):1161–70.

154. Verweij KJH, Abdellaoui A, Nivard MG, Sainz Cort A, Ligthart L, Draisma HHM, et al. Short communication: Genetic association between schizophrenia and cannabis use. Drug and Alcohol Dependence. 2017 Feb 1;171:117–21.

155. Zammit S, Hamshere M, Dwyer S, Georgiva L, Timpson N, Moskvina V, et al. A population-based study of genetic variation and psychotic experiences in adolescents. Schizophr Bull. 2014 Nov;40(6):1254–62.

156. Derks EM, Vorstman JAS, Ripke S, Kahn RS, Consortium TSPG, Ophoff RA. Investigation of the Genetic Association between Quantitative Measures of Psychosis and Schizophrenia: A Polygenic Risk Score Analysis. PLOS ONE. 2012 Jun 22;7(6):e37852.

157. Jones HJ, Stergiakouli E, Tansey KE, Hubbard L, Heron J, Cannon M, et al. Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. JAMA Psychiatry. 2016 Mar 1;73(3):221–8.

158. Jones HJ, Heron J, Hammerton G, Stochl J, Jones PB, Cannon M, et al. Investigating the genetic architecture of general and specific psychopathology in adolescence. Translational Psychiatry. 2018 Aug 8;8(1):1–11.

159. Pain O, Dudbridge F, Cardno AG, Freeman D, Lu Y, Lundstrom S, et al. Genome-wide analysis of adolescent psychotic-like experiences shows genetic overlap with psychiatric disorders. Am J Med Genet B Neuropsychiatr Genet. 2018;177(4):416–25.

160. Taylor MJ, Martin J, Lu Y, Brikell I, Lundström S, Larsson H, et al. Association of Genetic Risk Factors for Psychiatric Disorders and Traits of These Disorders in a Swedish Population Twin Sample. JAMA Psychiatry. 2019 Mar 1;76(3):280–9.

161. Giordano GN, Ohlsson H, Sundquist K, Sundquist J, Kendler KS. The association between cannabis abuse and subsequent schizophrenia: a Swedish national co-relative control study. Psychol Med. 2015 Jan;45(2):407–14.

162. Gage SH, Jones HJ, Burgess S, Bowden J, Smith GD, Zammit S, et al. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. Psychological Medicine. 2017 Apr;47(5):971–80.

163. Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. Am J Psychiatry. 2011 Oct;168(10):1041–9.

164. Hall FS, Drgonova J, Jain S, Uhl GR. Implications of genome wide association studies for addiction: Are our a priori assumptions all wrong? Pharmacol Ther. 2013 Dec;140(3):267–79.

165. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009 Feb;10(2):191–201.

166. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. Nat Rev Genet. 2009 Dec;10(12):872–8.

167. Reimers MA, Riley BP, Kalsi G, Kertes DA, Kendler KS. Pathway based analysis of genotypes in relation to alcohol dependence. The Pharmacogenomics Journal. 2012 Aug;12(4):342–8.

168. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry. 2010 Dec;15(12):1128–39.

169. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

170. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018 Feb; 27(2) e1608.

171. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010 Sep;467(7311):52–8.

172. Hamer D, Sirota L. Beware the chopsticks gene. Mol Psychiatry. 2000 Jan;5(1):11–3.

173. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort Profile: The Saguenay Youth Study (SYS). Int J Epidemiol. 2017 Apr;46(2):e19.

174. Boks MPM, Schipper M, Schubart CD, Sommer IE, Kahn RS, Ophoff RA. Investigating gene–environment interaction in complex diseases: increasing power by selective sampling for environmental exposure. Int J Epidemiol. 2007 Dec 1;36(6):1363–9.

175. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F, et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry. 2018 May 1;75(5):447–57.

176. Pausova Z, Paus T, Abrahamowicz M, Bernard M, Gaudet D, Leonard G, et al. Cohort Profile: The Saguenay Youth Study (SYS). Int J Epidemiol. 2017 Apr;46(2):e19.

177. Pausova Z, Paus T, Abrahamowicz M, Almerigi J, Arbour N, Bernard M, et al. Genes, maternal smoking, and the offspring brain and body during adolescence: design of the Saguenay Youth Study. Hum Brain Mapp. 2007 Jun;28(6):502–18.

178. Boks MP, He Y, Schubart CD, Gastel W van, Elkrief L, Huguet G, et al. Cannabinoids and psychotic symptoms: A potential role for a genetic variant in the P2X purinoceptor 7 (P2RX7) gene. Brain, Behavior, and Immunity. 2020 Aug 1;88:573–81.

179. Volk RJ, Steinbauer JR, Cantor SB, Holzer CE. The Alcohol Use Disorders Identification Test (AUDIT) as a screen for at-risk drinking in primary care patients of different racial/ethnic backgrounds. Addiction. 1997 Feb;92(2):197–206.

180. Chung T, Colby SM, Barnett NP, Rohsenow DJ, Spirito A, Monti PM. Screening adolescents for problem drinking: performance of brief screens against DSM-IV alcohol diagnoses. J Stud Alcohol. 2000 Jul;61(4):579–87.

181. Knight JR, Sherritt L, Harris SK, Gates EC, Chang G. Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcohol Clin Exp Res. 2003 Jan;27(1):67–73.

182. Fairlie AM, Sindelar HA, Eaton CA, Spirito A. Utility of the AUDIT for screening adolescents for problematic alcohol use in the emergency department. Int J Adolesc Med Health. 2006 Mar;18(1):115–22.

183. Aalto M, Alho H, Halme JT, Seppä K. AUDIT and its abbreviated versions in detecting heavy and binge drinking in a general population survey. Drug Alcohol Depend. 2009 Jul 1;103(1–2):25–9.

184. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. The Alcohol Use Disorders Identification Test; Guidelines for Use in Primary Care. 2001.

185. Rouquette A, Côté SM, Pryor LE, Carbonneau R, Vitaro F, Tremblay RE. Cohort profile: the Quebec Longitudinal Study of Kindergarten Children (QLSKC). Int J Epidemiol. 2014 Feb;43(1):23–33.

186. Hibell B, Andersson B, Bjarnason T, Kokkevi A, Morgan M, Narusk A, et al. The 1995 ESPAD report. Alcohol and other drug use among students in 26 European countiries. 1997 [cited 2020 Sep 25].

187. Keller MC. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol Psychiatry. 2014 Jan 1;75(1):18–24.

188. Hibell B, Andersson B, Bjarnason T, Ahlström S, Balakireva O, Kokkevi A, et al. The ESPAD Report 2003. Alcohol anb Other Drug Use Among Students in 35 European Countries [Internet]. 2004.

189. Mark W, Toulopoulou T. Psychometric Properties of "Community Assessment of Psychic Experiences": Review and Meta-analyses. Schizophr Bull. 2016 Jan;42(1):34–44.

190. Peters ER, Joseph SA, Garety PA. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). Schizophr Bull. 1999;25(3):553–76.

191. Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and Theoretical Foundations. The British Journal of Psychiatry. 1989 Nov;155(S7):49–52.

192. Kibel DA, Laffont I, Liddle PF. The composition of the negative syndrome of chronic schizophrenia. Br J Psychiatry. 1993 Jun;162:744–50.

193. Brenner K, Schmitz N, Pawliuk N, Fathalli F, Joober R, Ciampi A, et al. Validation of the English and French versions of the Community Assessment of Psychic Experiences (CAPE) with a Montreal community sample. Schizophrenia Research. 2007 Sep 1;95(1):86–95.

194. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience [Internet]. 2019 Jul;8(7).

195. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

196. Lou X-Y, Chen G-B, Yan L, Ma JZ, Zhu J, Elston RC, et al. A Generalized Combinatorial Approach for Detecting Gene-by-Gene and Gene-by-Environment Interactions with Application to Nicotine Dependence. Am J Hum Genet. 2007 Jun;80(6):1125–37.

197. Imai K, Keele L, Tingley D, Yamamoto T. Causal Mediation Analysis Using R. In: Vinod HD, editor. Advances in Social Science Research Using R. New York: Springer-Verlag; 2010.

198. United Nations Office On Drugs and Labour. World drug report 2018 (set of 5 booklets). Place of publication not identified: United Nations; 2018.

199. Pava MJ, Woodward JJ. A review of the interactions between alcohol and the endocannabinoid system: implications for alcohol dependence and future directions for research. Alcohol. 2012 May;46(3):185–204.

200. Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. Nature. 1990 Aug 9;346(6284):561–4.

201. Bühler K-M, Giné E, Echeverry-Alzate V, Calleja-Conde J, Fonseca FR de, López-Moreno JA. Common single nucleotide variants underlying drug addiction: more than a decade of research. Addiction Biology. 2015;20(5):845–71.

202. Nikolova YS, Ferrell RE, Manuck SB, Hariri AR. Multilocus Genetic Profile for Dopamine Signaling Predicts Ventral Striatum Reactivity. Neuropsychopharmacology. 2011 Aug;36(9):1940–7.

203. He Z, Zhang M, Lee S, Smith JA, Guo X, Palmas W, et al. Set-based Tests for Genetic Association in Longitudinal Studies. Biometrics. 2015 Sep;71(3):606–15.

204. Haughey HM, Marshall E, Schacht JP, Louis A, Hutchison KE. Marijuana withdrawal and craving: influence of the cannabinoid receptor 1 (CNR1) and fatty acid amide hydrolase (FAAH) genes. Addiction. 2008 Oct;103(10):1678–86.

205. Agarwal G, Tulsyan S, Lal P, Mittal B. Generalized Multifactor Dimensionality Reduction (GMDR) Analysis of Drug-Metabolizing Enzyme-Encoding Gene Polymorphisms may Predict Treatment Outcomes in Indian Breast Cancer Patients. World J Surg. 2016 Jul 1;40(7):1600–10.

206. Lin E, Kuo P-H, Liu Y-L, Yu YW-Y, Yang AC, Tsai S-J. A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers. Front Psychiatry. 2018;9:290

207. Chen G-B, Liu N, Klimentidis YC, Zhu X, Zhi D, Wang X, et al. A unified GMDR method for detecting gene–gene interactions in family and unrelated samples with application to nicotine dependence. Hum Genet. 2014 Feb 1;133(2):139–50.

208. Lee S, Teslovich TM, Boehnke M, Lin X. General Framework for Meta-analysis of Rare Variants in Sequencing Association Studies. Am J Hum Genet. 2013 Jul 11;93(1):42–53.

209. Zuo L, Kranzler HR, Luo X, Covault J, Gelernter J. CNR1 variation modulates risk for drug and alcohol dependence. Biol Psychiatry. 2007 Sep 15;62(6):616–26.

210. Morena M, Patel S, Bains JS, Hill MN. Neurobiological Interactions Between Stress and the Endocannabinoid System. Neuropsychopharmacology. 2016 Jan;41(1):80–102.

211. Lovallo WR. The hypothalamic–pituitary–adrenocortical axis in addiction. Int J Psychophysiol. 2006 Mar;59(3):193–4.

212. Kinsey SG, O'Neal ST, Long JZ, Cravatt BF, Lichtman AH. Inhibition of endocannabinoid catabolic enzymes elicits anxiolytic-like effects in the marble burying assay. Pharmacology Biochemistry and Behavior. 2011 Mar 1;98(1):21–7.

213. Sciolino NR, Zhou W, Hohmann AG. Enhancement of endocannabinoid signaling with JZL184, an inhibitor of the 2-arachidonoylglycerol hydrolyzing enzyme monoacylglycerol lipase, produces anxiolytic effects under conditions of high environmental aversiveness in rats. Pharmacological Research. 2011 Sep 1;64(3):226–34.

214. Shonesy BC, Bluett RJ, Ramikie TS, Báldi R, Hermanson DJ, Kingsley PJ, et al. Genetic Disruption of 2-Arachidonoylglycerol Synthesis Reveals a Key Role for Endocannabinoid Signaling in Anxiety Modulation. Cell Rep. 2014 Dec 11;9(5):1644–53.

215. Jenniches I, Ternes S, Albayram O, Otte DM, Bach K, Bindila L, et al. Anxiety, Stress, and Fear Response in Mice With Reduced Endocannabinoid Levels. Biological Psychiatry. 2016 May 15;79(10):858–68.

216. Serrano A, Pavon FJ, Buczynski MW, Schlosburg J, Natividad LA, Polis IY, et al. Deficient endocannabinoid signaling in the central amygdala contributes to alcohol dependence-related anxiety-like behavior and excessive alcohol intake. Neuropsychopharmacology. 2018 Aug;43(9):1840–50.

217. Colhoun HM, McKeigue PM, Smith GD. Problems of reporting genetic associations with complex outcomes. The Lancet. 2003 Mar 8;361(9360):865–72.

218. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, et al. Evaluating historical candidate genes for schizophrenia. Molecular Psychiatry. 2015 May;20(5):555–62.

219. Munafò MR, Gage SH. Improving the reliability and reporting of genetic association studies. Drug and Alcohol Dependence. 2013 Oct 1;132(3):411–3.

220. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, et al. Size matters: just how big is BIG?Quantifying realistic sample size requirements for human genome epidemiology. Int J Epidemiol. 2009 Feb 1;38(1):263–73.

221. Zuo L, Kranzler HR, Luo X, Yang B, Weiss R, Brady K, et al. Interaction between two independent CNR1 variants increases risk for cocaine dependence in European Americans: a replication study in family-based sample and population-based sample. Neuropsychopharmacology. 2009 May;34(6):1504–13.

222. Hopfer CJ, Young SE, Purcell S, Crowley TJ, Stallings MC, Corley RP, et al. Cannabis Receptor Haplotype Associated With Fewer Cannabis Dependence Symptoms In Adolescents. Am J Med Genet B Neuropsychiatr Genet. 2006 Dec 5;141B(8):895–901.

223. Kelley ME, Wan CR, Broussard B, Crisafio A, Cristofaro S, Johnson S, et al. Marijuana use in the immediate 5-year premorbid period is associated with increased risk of onset of schizophrenia and related psychotic disorders. Schizophrenia Research. 2016 Mar 1;171(1):62–7.

224. D'Souza DC, Perry E, MacDougall L, Ammerman Y, Cooper T, Wu Y, et al. The Psychotomimetic Effects of Intravenous Delta-9-Tetrahydrocannabinol in Healthy Individuals: Implications for Psychosis. Neuropsychopharmacology. 2004 Aug;29(8):1558–72.

225. Sommer IE, Brink W van den. High-potency cannabis and incident psychosis: correcting the causal assumption. The Lancet Psychiatry. 2019 Jun 1;6(6):464–5.

226. Forti MD, Morgan C, Selten J-P, Lynskey M, Murray RM. High-potency cannabis and incident psychosis: correcting the causal assumption – Authors' reply. The Lancet Psychiatry. 2019 Jun 1;6(6):466–7.

227. Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction. 2010 Mar;105(3):417–30.

228. Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, et al. Heritability Estimates for Psychotic Disorders: The Maudsley Twin Psychosis Series. Archives of General Psychiatry. 1999 Feb 1;56(2):162–8.

229. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. The Lancet Psychiatry. 2020 Oct:7(12):1032-1045.

230. Hjorthøj C, Uddin MJ, Wimberley T, Dalsgaard S, Hougaard DM, Børglum A, et al. No evidence of associations between genetic liability for schizophrenia and development of cannabis use disorder. Psychol Med. 2019 Dec 9;1–6.

231. Boks MP, He Y, Schubart CD, Gastel W van, Elkrief L, Huguet G, et al. Cannabinoids and psychotic symptoms: A potential role for a genetic variant in the P2X purinoceptor 7 (P2RX7) gene. Brain, Behavior, and Immunity. 2020 Aug 1;88:573–81.

232. Buchanan E. MeMoBootR: Mediation-Moderation with Bootstrapping in R. 2020.

233. French L, Gray C, Leonard G, Perron M, Pike GB, Richer L, et al. Early Cannabis Use, Polygenic Risk Score for Schizophrenia and Brain Maturation in Adolescence. JAMA Psychiatry. 2015 Oct 1;72(10):1002–11.

234. Vaucher J, Keating BJ, Lasserre AM, Gan W, Lyall DM, Ward J, et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. Molecular Psychiatry. 2018;23(5):1287–92.

235. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nature Genetics. 2018 Dec;50(12):1728–34.

236. Jang S-K, Saunders G, Liu M, 23andMe Research Team, Jiang Y, Liu DJ, et al. Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. Psychol Med. 2020 Aug 7;1–11.

237. Malhotra D, Sebat J. CNVs: Harbinger of a Rare Variant Revolution in Psychiatric Genetics. Cell. 2012 Mar 16;148(6):1223–41.

238. Serrano A, Pavon FJ, Buczynski MW, Schlosburg J, Natividad LA, Polis IY, et al. Deficient endocannabinoid signaling in the central amygdala contributes to alcohol dependence-related anxiety-like behavior and excessive alcohol intake. Neuropsychopharmacology. 2018 Aug;43(9):1840–50.

239. Karcher NR, Barch DM, Demers CH, Baranger DAA, Heath AC, Lynskey MT, et al. Genetic Predisposition vs Individual-Specific Processes in the Association Between Psychotic-like Experiences and Cannabis Use. JAMA Psychiatry. 2019 Jan 1;76(1):87–94.

240. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine. 2020 May 18;12(1):44.

241. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics. 2019 Apr;51(4):584–91.

242. Lewis ACF, Green RC. Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues. Genome Med. 2021 Jan 28;13(1):14.

243. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015 Oct 1;97(4):576–92.

244. Belsky DW, Moffitt TE, Houts R, Bennett GG, Biddle AK, Blumenthal JA, et al. Polygenic Risk, Rapid Childhood Growth, and the Development of Obesity: Evidence From a 4-Decade Longitudinal Study. Arch Pediatr Adolesc Med. 2012 Jun 1;166(6):515–21.

245. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry. 2017 Mar 15;81(6):470–7.

246. Palk AC, Dalvie S, de Vries J, Martin AR, Stein DJ. Potential use of clinical polygenic risk scores in psychiatry – ethical implications and communicating high polygenic risk. Philos Ethics Humanit Med. 2019 Feb; 14(1).

247. Rössler W. The stigma of mental disorders. EMBO Rep. 2016 Sep;17(9):1250–3.

248. Shen L, Condit CM, Wright L. The Psychometric Property and Validation of a Fatalism Scale. Psychol Health. 2009 Jun;24(5):597–613.

249. Liu C, Kanazawa T, Tian Y, Mohamed Saini S, Mancuso S, Mostaid MS, et al. The schizophrenia genetics knowledgebase: a comprehensive update of findings from candidate gene studies. Translational Psychiatry. 2019 Aug 27;9(1):1–7.

250. Duncan LE, Ostacher M, Ballon J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology. 2019 Aug;44(9):1518–23.

251. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. Psychiatric Genomics: An Update and an Agenda. Am J Psychiatry. 2018 Jan 1;175(1):15–27.

252. Benros ME, Mortensen PB, Eaton WW. Autoimmune diseases and infections as risk factors for schizophrenia. Ann N Y Acad Sci. 2012 Jul;1262:56–66.

253. Organization WH. ICD-10 : international statistical classification of diseases and related health problems : tenth revision [Internet]. World Health Organization; 2004 [cited 2020 Nov 15]. Available from: https://apps.who.int/iris/handle/10665/42980

254. Fusar-Poli P, Solmi M, Brondino N, Davies C, Chae C, Politi P, et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry. 2019 Jun;18(2):192–207.

255. Smoller JW. Disorders and borders: Psychiatric genetics and nosology. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2013;162(7):559–78.

256. Lee PH, Feng Y-CA, Smoller JW. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. Biological Psychiatry. 2020 Oct.

257. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell. 2019 Dec 12;179(7):1469-1482.e11.

258. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133–63.

259. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018 Jul; 362.

260. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics. 2013 Jul;14(7):483–95.

261. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015 Apr;44(2):512–25.

262. Jang S-K, Saunders G, Liu M, Team 23andMe Research, Jiang Y, Liu DJ, et al. Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. Psychological Medicine. undefined/ed;1–11.

263. Hari Dass SA, McCracken K, Pokhvisneva I, Chen LM, Garg E, Nguyen TTT, et al. A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions. EBioMedicine. 2019 Apr 1;42:188–202.

264. Figlewicz DP, Bennett JL, Aliakbari S, Zavosh A, Sipols AJ. Insulin acts at different CNS sites to decrease acute sucrose intake and sucrose self-administration in rats. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 2008 Aug 1;295(2):R388–94.

265. Schoffelmeer ANM, Drukarch B, Vries TJD, Hogenboom F, Schetters D, Pattij T. Insulin Modulates Cocaine-Sensitive Monoamine Transporter Function and Impulsive Behavior. J Neurosci. 2011 Jan 26;31(4):1284–91.

266. Kleinridders A, Cai W, Cappellucci L, Ghazarian A, Collins WR, Vienberg SG, et al. Insulin resistance in brain alters dopamine turnover and causes behavioral disorders. PNAS. 2015 Mar 17;112(11):3463–8.

267. Stoeckel LE, Arvanitakis Z, Gandy S, Small D, Kahn CR, Pascual-Leone A, et al. Complex mechanisms linking neurocognitive dysfunction to insulin resistance and other metabolic dysfunction. F1000Res. 2016 Jun;5:353.

268. Bomfim TR, Forny-Germano L, Sathler LB, Brito-Moreira J, Houzel J-C, Decker H, et al. An anti-diabetes agent protects the mouse brain from defective insulin signaling caused by Alzheimer’s disease–associated Aβ oligomers. J Clin Invest. 2012 Apr 2;122(4):1339–53.

269. Wray NR, Lin T, Austin J, McGrath JJ, Hickie IB, Murray GK, et al. From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. JAMA Psychiatry. 2020 Sep;78(1):101-109.

270. Perkins DO, Olde Loohuis L, Barbee J, Ford J, Jeffries CD, Addington J, et al. Polygenic Risk Score Contribution to Psychosis Prediction in a Target Population of Persons at Clinical High Risk. AJP. 2019 Nov 12;177(2):155–63.

271. Hou L, Heilbronner U, Degenhardt F, Adli M, Akiyama K, Akula N, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. Lancet. 2016 Mar 12;387(10023):1085–93.

272. Ward J, Graham N, Strawbridge RJ, Ferguson A, Jenkins G, Chen W, et al. Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts. PLoS One. 2018;13(9):e0203896.

273. Iniesta R, Hodgson K, Stahl D, Malki K, Maier W, Rietschel M, et al. Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. Sci Rep. 2018 03;8(1):5530.

Annex A – Supplementary data article 1

Endocannabinoid Gene x Gene Interaction Association to Alcohol Use Disorder in two Adolescent Cohorts.

Laurent Elkrief [1,2], Sean Spinney [2,3], Daniel E. Vosberg [4], Tobias Banaschewski M.D., Ph.D.[5]; Arun L.W. Bokde Ph.D.[6]; Erin Burke Quinlan PhD[7]; Sylvane Desrivières Ph.D.[7]; Herta Flor Ph.D.[8,9]; Hugh Garavan Ph.D.[10]; Penny Gowland Ph.D.[11]; Andreas Heinz M.D., Ph.D.[12]; Rüdiger Brühl Ph.D.[13]; Jean-Luc Martinot M.D., Ph.D.[14], Marie-Laure Paillère Martinot M.D., Ph.D.[15]; Frauke Nees Ph.D.[5,8]; Dimitri Papadopoulos Orfanos Ph.D.[16]; Luise Poustka M.D.[17]; Sarah Hohmann M.D.[5]; Sabina Millenet Dipl.-Psych.[5]; Juliane H. Fröhner MSc[18]; Michael N. Smolka M.D.[18]; Henrik Walter M.D., Ph.D.[12]; Robert Whelan Ph.D.[19]; Gunter Schumann M.D.[7,20]; Zdenka Pausova[21], Tomáš Paus [4,22], Guillaume Huguet [2,3]†, Patricia Conrod [2,3,23]*†, and the IMAGEN consortium.

Supplementary Methods

Phenotype Evaluated

While other studies focusing on adolescent alcohol abuse used a less stringent cutoff (1)(2)(3), the more stringent cut-off of 8 was chosen as this was more widely validated across different studies (see (4) for a review). Questionnaires were given to individuals, at 14, 16 and 18 years old. While there were missing data at various times for some participants, for the purpose of this analysis, if an individual's genetic data and AUDIT data were given at one time point, the individual was included in this study.

Set-Based Test

To determine the relevant SNPs for our analysis, a Set-Based test was performed using PLINK1.9. For a detailed explanation of how the analysis works, see the PLINK 1.9 manual (https://www.cog-genomics.org/plink/1.9/assoc#set). In this project, three set-based tests were carried using parameters of varying stringencies. The parameters that were adjusted between the tests were p-value for significant variants between tests, r^2 of variant pairs, and maximum set size. Data in all three set-based test underwent 10, 000 label-swapped permutation as well, using the --perm function in PLINK1.9. The first done, was the default test in PLINK1.9, with a p-value of 0.05, r^2 of 0.5, and a set-max of 5, the second test had a p-value of 0.05, r^2 of 0.3, and set-max of 3, while test 3 had a p of 0.01, r^2 of 0.1 and set-max of 2. Tests 2 and 3 were more stringent, and were run to challenge the data, to ensure robustness of our results. The three set-based tests were run,

with varying results (Suppl. Table1). Sixty nine SNPs appearing across five cannabinoid-related genes were analyzed for their relation to AUDIT scores. In the first set-based test, 9 SNPS returned with nominal p values of <0.05, of which 7 also passed linkage disequilibrium (LD) criterion. Through the first set-test criterion, only the CNR1 gene-set was significantly associated to an AUDIT score of eight or more (p=0.022). Within this set, only rs9353525 that was significantly and independently related to SUD. In the second set-based test, the same 9 SNPS returned with nominal p values of <0.05, of which, 5 SNPs passed the LD criterion. Again, only CNR1 was significantly associated to an AUDIT score of greater than seven (p=0.03). Finally, 4 SNPs returned with a nominal p value <0.01, in the third test, with 2 SNPs passing LD criterion. No genes remained significant after the third set-based test. As mentioned above, the 7 SNPs that had nominal p values of <0.05, in the first set-based test, and that passed LD criterion (r2<0.5) were extracted and only these were analyzed in the case-control, model analysis and logistic regression [For summary of set-based test, see Suppl. Table 6].

Case-Control Analysis

Our first case/control association analysis was done using Fisher's exact test, through PLINK1.9. SNPs selected for this analysis were those that were both significant and independent in the first set-based tests. Cases were considered if an AUDIT score of eight or more was reported. Four case-control analyses were run. In the first, cases were considered if an AUDIT of eight was reached at any time point (ALL); if individuals scored greater than seven at multiple time points, the duplicate data were removed. The other analyses were done at each time point; one for age 14, one for 16 and one for 18. Analysis were adjusted for multiple tests with false discovery rate correction, using the --adjust function in the PLINK1.9 program, and false discovery rate (FDR) values are reported.

Genetic Analysis of Population Stratification

A principal components analysis based on the variance-standardized relationship matrix and displayed the 20 first genetic dimensions and associated-eigenvalues (*supplemental figure 1*) was performed. To avoid confounding factors related to ancestry the 6 first ancestry components were used as covariables in the logistic regression.

Covariables

A logistic regression was done on the SNPs that remained significant after correction for multiple tests. Our regressions had sex, the first six components of the MDS plot, parental alcohol abuse, and parental education as co-variables. Parental education was taken from self-report answers within the European School Survey Project on Alcohol and Other Drugs (ESPAD+) questionnaire administered at the first (fourteen years old) and second time point (sixteen years old) in IMAGEN. Alcohol abuse in parents was measured using the AUDIT information obtained at the first two time points in IMAGEN. If ESPAD+ and AUDIT information were missing at the 18 year old time point, the most complete and recent information was used at this time point. For our analyses that considered AUDIT information without a regard to time, the information used in the 14 and 16 year questionnaires were mixed, so that if a parent had signalled a DRINKING issue on AUDIT at any time, they were flagged as such. Moreover if parental information was missing, individuals were not included in the logistic regression.

Seattle Seq Annotations of SNPs

We used SeattleSeq (http://snp.gs.washington.edu/SeattleSeqAnnotation137/) to annotated SNPs in this study (Supplementary table 2).

The following annotations used in supplementary table 2 (information extract of SeattleSeq website) :

--- Name SNPs

--- chromosome (input from the user)

--- position (hg18) (input from user, location on the chromosome, hg18, 1-based)

--- position (hg19) (input from user, location on the chromosome, hg19, 1-based)

--- Major allele (Major allele frequency refers to the frequency at in an IMAGEN)

--- Minor allele (Minor allele frequency refers to the frequency at in an IMAGEN)

--- chimp. Allele (column chimpAllele: UCSC alignments)

--- Gene region (Name gene observed for SNPs)

--- FunctionGVS (GVS class of variation function, using only hg19 and your submitted alleles; see description)

--- Function in protein (Description of genetic variation for amino acids in protein)

--- PolyPhen Prediction (column polyPhen: amino acid substitution impacts)

--- Grantham Score (column granthamScore: the Grantham score of any amino acid changes, as per Grantham (1974) Science, Table 2)

--- Conservation Score phastCons (column scorePhastCons: UCSC, 46 placental mammalian species, range of 0 to 1, with 1 being the most conserved)

--- Conservation Score GERP (column consScoreGERP: rejected-substitution score from the program GERP, Stanford University, range of -12.3 to 6.17, with 6.17 being the most conserved)

--- CADD C Score (column scoreCADD: phred-like Combined Annotation Dependent Depletion scores from Kircher et al., University of Washington, range 0 though 99)

--- HapMap Frequencies (3 columns AfricanHapMapFreq, EuropeanHapMapFreq, AsianHapMapFreq: African, European, and Asian, in percent)

--- Has Genotypes (column hasGenotypes: whether dbSNP has genotypes available for the variation)

--- dbSNP Validation (column dbSNPValidation: dbSNP validation status codes, dealing with e.g. whether the variation has been seen at least twice)

--- Repeats (2 columns repeatMasker and tandemRepeat)

--- Clinical Association (column clinicalAssociation: links to NCBI pages and PubMed)

--- Distance to Nearest Splice Site (column distanceToSplice: how close the variation is to a splice site)

--- CpG Islands (column cpgIslands: whether in a region where CpGs are present at a high level, from the UCSC genome annotation database)

--- NHLBI ESP Allele Counts (column genomesESP: the allele counts observed in the Exome Sequencing Project, optionally split by two ancestries)

--- ExAC Allele Counts (column genomesExAC: the allele counts of the Exome Aggregation Consortium, optionally split by 7 populations)

Supplementary Results

We performed case-control analysis of the IMAGEN cohort, where cases are considered if an individual scored eight or more on AUDIT at 14, 16, or 18 years of age. At age 14, none of the 7 SNPs were significantly associated with AUD ($p>0.05$) [*Table S7*]. At age 16, one SNP was significantly associated to case-control status; rs9343535 ($p=0.012$, $OR=0.656$) however it did not remain significant after FDR correction [*Table S7*]. Finally, at age 18, 4 SNPs were significantly

associated to case-control status group membership: rs782446 in *MGLL* (p=0.04, OR=0.813), rs484061 in *MGLL* (p=0.014, OR=0.814), rs507961 in *MGLL* (p=0.004, OR=0.742), rs9353525 in *CNR1* (p=0.025, OR=0.748). After correction for multiple tests, none remained significant [*Table S7*].

Logistic models were done for both SNPs, at each time point, as well as for any positive screen for AUD (ALL) analysis. After controlling for the effects of the first six principal components, sex, parental AUDIT scores (at any time) and parental education, both rs9353525 and rs507961 were still significantly associated with positive AUDIT screen in the ALL analysis [Table 3] (p<0.01), with both SNPs minor allele acting as protective factors (OR<1). For complete results of logistic regression see *Table S9.*

Supplementary Tables

Tableau 8. – Suppl. Table S1 (article 1).

Age	Sex	Control (AUDIT <8)	Case (AUDIT >=8)	Total
14	Male	973	32	1005
	Female	997	42	1039
	Total	1970	74	2044
16	Male	638	145	783
	Female	744	91	835
	Total	1382	236	1618
18	Male	442	231	673
	Female	585	172	757
	Total	1027	403	1430
All	Male	683	325	1008
	Female	793	250	1043
	Total	1476	575	2051

Supplementary Table S1: Description of subjects in IMAGEN. Subjects are classified by status (AUDIT score > or = to 8 or control), sex and age.

Tableau 9. – Suppl. Table S2 (article 1).

Fichier trop grand – à intégré lors du dépôt finale

Tableau 10. – Suppl. Table S3 (article 1).

Information		SYS Cohort
N (% female)		772 (52.07%)
Pedigree by numbres of childrens	All	401
	1	86
	2	266
	>= 3	49
Age months	Mean	180.1930052
	SD	22.25195432
	Min	144
	Max	228
Phenotype (% samples)	0	381 (49.35%)
	1	343 (44.43%)
	2	38 (4.92%)
	3	6 (0.78%)
	4	4 (0.52%)

Supplementary Table S3: Description of subjects in SYS cohort.

Tableau 11. – Suppl. Table S4 (article 1).

AUDIT	GRIP		
1. How often do you have a drink containing alcohol?	1	Have you tried alcohol in your life ?	Yes or No
	2	How often do you consume alcohol?	1=Alittlefromtimetotime 2=Alittledaytoday 3=Muchfromtimetotime 4=MuchDayToDay (h16)
2. How many drinks containing alcohol do you have on a typical day when you are drinking?			
3. How often do you have six or more drinks on one occasion?	3	In the last year, have you had 6 or more drinks?	Yes or No
4. How often during the last year have you found that you were not able to stop drinking once you had started?	4	Have you already tried decreasing your alcohol consumption?	Yes or No
	5	Have you already had throughs of decreasing alcohol conusmption?	Yes or No
	6	Have you already asked for help because of your alcohol use?	Yes or No
5.How often during the last year have you failed to do what was normally expected from you because of drinking?	7	Have you missed school to go drink or because you were hungover?	Yes or No
6.How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?			
7.How often during the last year have you had a feeling of guilt or remorse after drinking?	8	Have you felt guilty because of your alcohol consumption ?	Yes or No
8.How often during the last year have you been unable to remember what happened the night before because you had been drinking?			
9. Have you or someone else been injured as a result of your drinking?	9	Have you ever driven under the influence of alcohol?	1=Often2=Some Times 3=Once Or Twice4=Never or I Do Not Drive
	10	Have you ever been in trouble with the law or police because you were drunk or drank alcohol?	Yes or No
10.Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down?	11	Have you ever gotten into an argument or fight with family and/or friends because of your drinking?	Yes or No

Supplementary Table S4: Comparing the AUDIT and GRIP questionnaires. Questions from the GRIP questionnaire were selected for comparativeness to the AUDIT. Questions 1,3-8, 1

Tableau 12. – Suppl. Table S5 (article 1).

SNP	A1 SYS	A2 SYS	MAF
rs782446	C	A	0.2613
rs484061	A	G	0.4955
rs604300	A	G	0.1049
rs507961	T	C	0.1989
rs9353525	A	G	0.1223
rs7766029	T	C	0.496
rs4729873	G	A	0.3654
rs4963307	G	A	0.4914
rs10488693	T	C	0.05227

Supplementary Table S5: Description of SNPs in SYS cohort. A1=Minor allele. A2=Major Allele. MAF=Minor Allele Frequency.

Tableau 13. – Suppl. Table S6 (article 1).

Tests	Genes	N_{SNP}	N_{SIG}	I_{SIG}	pvalue	SNPs
Set1	CNR1	16	2	1	0.022	rs9353525
	NAPEPLD	3	1	1	0.07557	rs4729873
	FAAH	5	0	0	1	-
	MGLL	31	5	4	0.1487	rs507961 | rs484061 | rs782446 | rs604300
	DAGLA	14	1	1	0.2023	rs10488693
Set2	CNR1	16	2	1	0.02995	rs9353525
	NAPEPLD	3	1	1	0.07587	rs4729873
	FAAH	5	0	0	1	-
	MGLL	31	5	2	0.0689	rs507961 | rs484061
	DAGLA	14	1	1	0.2033	rs10488693
Set3	CNR1	16	2	1	0.05276	rs9353525
	NAPEPLD	3	0	0	1	-
	FAAH	5	0	0	1	-
	MGLL	31	2	1	0.09677	rs507961
	DAGLA	14	0	0	1	-

Supplementary Table S6: Table of results for set genes analysis with AUDIT in IMAGEN. Set1 test was done using the default parameters in PLINK (p-value of 0.05, r of 0.5, and a set-max of 5) Parameters for set2 were p-value of 0.05, r of 0.3, and set-max of 3, while set3 had a p of 0.01, r of 0.1 and set-max of 2. "NSNP" = number of SNPs analyzed. "NSIG" = number of significant SNPs (significance as defined by set test rules). SNPs. "ISIG" = Number of significant and independent (LD criterion of test). "pvalue" = p value of the gene set. "SNPs" = name of independent and significant SNPs.

Tableau 14. – Suppl. Table S7 (article 1).

Age	Name SNP	Allele 1	Allele 2	Freq Allele 1	Freq Allele 2	Odd-Ratio	Pvalue	FDR Pvalue
14	rs782446	C	A	0.2095	0.25	0.7949	0.2874	1
	rs484061	G	A	0.4324	0.4985	0.7666	0.1315	1
	rs604300	A	G	0.08108	0.1085	0.7247	0.3445	1
	rs507961	T	C	0.1824	0.2282	0.7548	0.2296	1
	rs9353525	A	G	0.1081	0.1274	0.8301	0.614	1
	rs4729873	G	A	0.3378	0.3572	0.9181	0.6628	1
	rs10488693	T	C	0.0473	0.07263	0.6339	0.3278	1
16	rs782446	C	A	0.2288	0.2475	0.9023	0.4174	1
	rs484061	G	A	0.4597	0.4902	0.8849	0.2319	0.8417
	rs604300	A	G	0.08085	0.1104	0.7086	0.062	0.417
	rs507961	T	C	0.1992	0.2297	0.8338	0.1525	0.6921
	rs9353525	A	G	0.08936	0.1301	0.6563	0.01244	0.2258
	rs4729873	G	A	0.3199	0.3645	0.8202	0.06893	0.417
	rs10488693	T	C	0.06992	0.07344	0.9483	0.8484	0.173
18	rs782446	C	A	0.2196	0.2571	0.8133	0.03812	0.2668
	rs484061	G	A	0.4541	0.5054	0.8142	0.01419	0.1288
	rs604300	A	G	0.09701	0.1127	0.846	0.2549	0.6608
	rs507961	T	C	0.1873	0.2371	0.7418	0.003796	0.0689
	rs9353525	A	G	0.107	0.138	0.7479	0.02566	0.1552
	rs4729873	G	A	0.33	0.3683	0.8449	0.0564	0.2047
	rs10488693	T	C	0.05707	0.07157	0.7852	0.1849	0.5593

Supplementary table S7: Description of fisher test for differents age in IMAGEN.
Freq = frequency.

Tableau 15. – Suppl. Table S8 (article 1).

Interaction	Variables	Quasi poisson					Binary				
		BETA	Std.Error	DF	t-value	p-value	BETA	Std.Error	DF	t-value	p-value
Combo	Intercept	-5.593	0.331	400	-16.914	8.47E-49	-29.455	1.661	400	-17.734	2.45E-52
	Combo(1)	0.078	0.076	368	1.02	3.08E-01	0.501	0.27	368	1.857	6.42E-02
	Sex(2)	0.03	0.074	368	0.399	6.90E-01	-0.269	0.304	368	-0.883	3.78E-01
	Age(months)	0.027	0.002	368	16.011	3.69E-44	0.124	0.008	368	16.25	3.83E-45
rs782446 (C)	Intercept	-5.531	0.325	400	-17.004	3.46E-49	-27.715	1.605	400	-17.27	2.48E-50
	rs782446	-0.009	0.061	368	-0.142	8.87E-01	-0.69	0.29	368	-2.382	1.77E-02
	Sex(2)	0.025	0.074	368	0.34	7.34E-01	-0.365	0.314	368	-1.162	2.46E-01
	Age(months)	0.027	0.002	368	16.035	2.95E-44	0.119	0.008	368	15.803	2.63E-43
rs484061 (G)	(Intercept)	-5.484	0.324	400	-16.906	9.10E-49	-29.374	1.669	400	-17.598	9.53E-52
	rs484061	-0.082	0.054	368	-1.529	1.27E-01	-1.097	0.241	368	-4.545	7.47E-06
	Sex(2)	0.032	0.074	368	0.429	6.68E-01	-0.192	0.313	368	-0.613	5.40E-01
	Age(months)	0.027	0.002	368	16.117	1.35E-44	0.129	0.008	368	16.084	1.84E-44
rs604300 (A)	(Intercept)	-5.547	0.325	400	-17.063	1.93E-49	-28.723	1.57	400	-18.297	8.96E-55
	rs604300	0.044	0.1	368	0.436	6.63E-01	-0.448	0.373	368	-1.2	2.31E-01
	Sex(2)	0.025	0.074	368	0.338	7.36E-01	-0.244	0.305	368	-0.798	4.25E-01
	Age(months)	0.027	0.002	368	16.022	3.31E-44	0.122	0.007	368	16.53	2.66E-46
rs507961 (T)	(Intercept)	-5.532	0.325	400	-17.013	3.18E-49	-29.77	1.596	400	-18.654	2.52E-56
	rs507961	-0.007	0.07	368	-0.095	9.24E-01	0.614	0.296	368	2.072	3.89E-02
	Sex(2)	0.025	0.074	368	0.34	7.34E-01	-0.245	0.3	368	-0.818	4.14E-01
	Age(months)	0.027	0.002	368	16.026	3.19E-44	0.125	0.007	368	16.945	5.10E-48
rs9353525 (A)	(Intercept)	-5.532	0.324	400	-17.083	1.58E-49	-28.561	1.552	400	-18.398	3.27E-55
	rs9353525	-0.08	0.078	368	-1.027	3.05E-01	0.699	0.295	368	2.37	1.83E-02
	Sex(2)	0.021	0.074	368	0.278	7.82E-01	-0.283	0.301	368	-0.937	3.49E-01
	Age(months)	0.027	0.002	368	16.031	3.04E-44	0.12	0.007	368	16.426	7.19E-46
rs4729873 (G)	(Intercept)	-5.543	0.326	400	-17.028	2.73E-49	-29.14	1.573	400	-18.525	9.16E-56
	rs4729873	0.01	0.056	368	0.172	8.63E-01	-0.294	0.217	368	-1.354	1.77E-01
	Sex(2)	0.024	0.074	368	0.328	7.43E-01	-0.165	0.308	368	-0.535	5.93E-01
	Age(months)	0.027	0.002	368	16.042	2.76E-44	0.124	0.007	368	16.695	5.55E-47
rs7766029 (T)	(Intercept)	-5.615	0.329	400	-17.08	1.63E-49	-28.905	1.576	400	-18.339	5.85E-55
	rs7766029	0.088	0.052	368	1.684	9.31E-02	0.822	0.268	368	3.066	2.33E-03
	Sex(2)	0.03	0.074	368	0.401	6.89E-01	-0.286	0.299	368	-0.956	3.40E-01
	Age(months)	0.027	0.002	368	15.96	5.97E-44	0.118	0.007	368	15.921	8.67E-44
rs4963307 (G)	(Intercept)	-5.537	0.328	400	-16.898	9.83E-49	-28.675	1.577	400	-18.178	2.92E-54
	rs4963307	0.001	0.055	368	0.019	9.85E-01	0.111	0.294	368	0.377	7.06E-01
	Sex(2)	0.025	0.074	368	0.34	7.34E-01	-0.277	0.305	368	-0.91	3.63E-01
	Age(months)	0.027	0.002	368	16.038	2.86E-44	0.121	0.007	368	16.495	3.72E-46

Supplementary table S8: Summary of replication test statistics in SYS cohort. Combo = Result of GMDR 3 SNP model x phenotype. DF= Degrees of Freedom. The combo defined two groups for all combinations observed in GMDR, one group for low risk coding with 0 and the second group high risk with 1

Tableau 16. – Suppl. Table S9 (article 1).

Age	SNP	Covariables	BETA	Odd-Ratio	STAT	Pvalue
ALL	rs507961 (N=2030)	SNP	-0.2696	0.76368491	-3.064	0.002188
		PC1	5.087	161.903422	1.901	0.05732
		PC2	3.148	23.2894391	1.201	0.2296
		PC3	1.726	5.61813636	0.658	0.5107
		PC4	9.703	16366.6335	3.820	0.0001333
		PC5	11.37	86681.8675	4.690	2.73E-06
		PC6	-6.93	0.000978	-1.005	0.3147
		Alcohol abuse history in family	0.364	1.43907421	2.414	0.0158
		Education of mother	-0.02348	0.97679351	-0.672	0.5018
		Education of father	-0.01814	0.98202354	-0.567	0.5706
		Gender	0.3802	1.46257708	3.747	0.0001787
	rs9353525 (N=2026)	SNP	-0.3015	0.73970783	-2.605	0.009178
		PC1	4.581	97.6119574	1.707	0.08774
		PC2	2.497	12.1460012	0.935	0.35
		PC3	1.556	4.73982398	0.593	0.553
		PC4	9.934	20619.6545	3.906	9.37E-05
		PC5	11.18	71682.3621	4.594	4.35E-06
		PC6	-7.443	0.00058553	-1.078	0.281
		Alcohol abuse history in family	0.3605	1.43404626	2.393	0.01673
		Education of mother	-0.02567	0.97465667	-0.733	0.4635
		Education of father	-0.01165	0.9884176	-0.363	0.7168
		Gender	0.3638	1.43878643	3.586	0.0003359
14	rs507961 (N=2024)	SNP	-0.2436	0.7838	-1.102	0.2705
		PC1	7.909	2723	0.8821	0.3777
		PC2	-4.481	0.01132	-0.8342	0.4042
		PC3	7.155	1281	1.135	0.2562
		PC4	1.662	5.272	0.272	0.7856
		PC5	1.339	3.816	0.2165	0.8286
		PC6	-9.929	4.88E-05	-0.6029	0.5465
		Alcohol abuse history in family	0.3631	1.438	1.473	0.1407
		Education of mother	0.9061	2.475	2.874	0.004047
		Education of father	0.1455	1.157	1.747	0.08064
		Gender	0.1332	1.142	1.739	0.0821
	rs9353525 (N=2020)	SNP	-0.2148	0.8067	-0.7725	0.4398
		PC1	7.482	1775	0.8323	0.4052
		PC2	-5.02	0.006602	-0.9217	0.3567
		PC3	7.032	1132	1.126	0.2602
		PC4	1.762	5.822	0.2887	0.7728

		PC5	1.24	3.455	0.1999	0.8416
		PC6	-10.19	3.77E-05	-0.6173	0.5370
		Alcohol abuse history in family	0.3574	1.43	1.45	0.1471
		Education of mother	0.8877	2.43	2.816	0.0049
		Education of father	0.1482	1.16	1.78	0.0750
		Gender	0.1345	1.144	1.752	0.0797
16	rs507961 (N=1535)	SNP	-0.1902	0.82679376	-1.486	0.1374
		PC1	4.505	90.468344	0.9124	0.3616
		PC2	13.32	609259.765	1.641	0.1007
		PC3	-6.465	0.00155699	-1.125	0.2607
		PC4	3.895	49.1560534	0.8842	0.3766
		PC5	6.622	751.446488	1.501	0.1334
		PC6	3.451	31.5319085	0.3453	0.7298
		Alcohol abuse history in family	-0.5608	0.57075228	-3.775	0.0002
		Education of mother	0.3117	1.36574491	1.234	0.2170
		Education of father	0.04655	1.04765046	0.9311	0.3518
		Gender	-0.03174	0.96875843	-0.6799	0.4966
	rs9353525 (N=1532)	SNP	-0.4612	0.63052656	-2.489	0.0128
		PC1	4.506	90.5588576	0.9033	0.3664
		PC2	12.82	369534.727	1.517	0.1292
		PC3	-6.861	0.00104787	-1.167	0.2433
		PC4	3.321	27.6880247	0.742	0.4581
		PC5	6.452	633.968964	1.434	0.1515
		PC6	2.5	12.182494	0.2484	0.8038
		Alcohol abuse history in family	-0.5414	0.58193298	-3.64	0.0003
		Education of mother	0.3031	1.35404986	1.198	0.2307
		Education of father	0.04648	1.04757713	0.9236	0.3557
		Gender	-0.02175	0.97848483	-0.4627	0.6436
18	rs507961 (N=1243)	SNP	-0.3041	0.73778708	-2.588	0.009645
		PC1	0.1963	1.21689192	0.049	0.9607
		PC2	4.757	116.396213	1.256	0.2089
		PC3	4.724	112.617824	1.197	0.2313
		PC4	12.72	334368.849	3.693	0.0002215
		PC5	20.18	580848196	6.179	6.44E-10
		PC6	6.209	497.203799	0.674	0.5001
		Alcohol abuse history in family	0.1232	1.13111062	0.532	0.595
		Education of mother	-0.06384	0.93815509	-1.380	0.1675
		Education of father	0.0163	1.01643357	0.387	0.6986
		Gender	0.5777	1.78193526	4.331	1.49E-05
	rs935 3525	SNP	-0.3213	0.72520566	-2.059	0.03949
		PC1	0.01436	1.0144636	0.003592	0.9971

		PC2	4.564	95.9665794	1.204	0.2284
		PC3	3.921	50.4508704	1.003	0.3160
		PC4	12.61	299539.028	3.652	0.0003
		PC5	20.44	753319501	6.249	0.0000
		PC6	5.929	375.778547	0.6428	0.5204
		Alcohol abuse history in family	-0.5436	0.58065413	-4.094	0.0000
		Education of mother	0.09048	1.09469961	0.3899	0.6966
		Education of father	-0.06477	0.93728301	-1.398	0.1621
		Gender	0.02071	1.02092594	0.4912	0.6233

Supplementary table S9: Logistic regression of rs507961 and rs9353525 by phenotype in IMAGEN. (STAT = coefficient t-statistic). PC1-PC6, First six ancestry components.

Tableau 17. – Suppl. Table S10 (article 1).

Condition	SNP	Covariables	BETA	Odd-Ratio	STAT	Pvalue
ALL	rs507961 (N=2030)	ADD	2030	0.7508	-3.217	0.001294
		Sex	2030	0.6798	-3.8	0.0001444
		PC1	2030	114.7	1.763	0.07798
		PC2	2030	18.97	1.082	0.2791
		PC3	2030	0.119	-0.6629	0.5074
		PC4	2030	18410	3.838	0.0001241
		PC5	2030	90900	4.66	3.16E-06
		PC6	2030	0.0006935	-1.055	0.2913
		FlagAllTimes	2030	1.438	2.411	0.01593
		EduMom	2030	0.978	-0.636	0.5248
		EduDad	2030	0.9803	-0.6208	0.5348
		ADDxPC3	2030	44410	2.043	0.04108
	rs9353525 (N=2026), Interaction 1	ADD	2026	0.6716	-3.024	0.002496
		Sex	2026	0.6937	-3.6	0.0003176
		PC1	2026	8.169	0.7632	0.4454
		PC2	2026	0.4376	-0.2674	0.7892
		PC3	2026	4.403	0.5529	0.5803
		PC4	2026	29440	3.987	6.69E-05
		PC5	2026	64400	4.454	8.42E-06
		PC6	2026	0.0005432	-1.087	0.2771
		FlagAllTimes	2026	1.441	2.423	0.0154
		EduMom	2026	0.9731	-0.7749	0.4384
		EduDad	2026	0.9897	-0.3204	0.7487
		ADDxPC1	2026	2.21E+14	2.111	0.03474
	rs9353525 (N=2026), Interaction 2	ADD	2026	0.7553	-2.394	0.01665
		Sex	2026	0.6957	-3.569	0.0003577
		PC1	2026	107.4	1.74	0.08192
		PC2	2026	15.71	1.03	0.303
		PC3	2026	5.207	0.6293	0.5292
		PC4	2026	24420	3.966	7.30E-05
		PC5	2026	76930	4.608	4.06E-06
		PC6	2026	2.39E-08	-2.261	0.02378
		FlagAllTimes	2026	1.428	2.358	0.01835
		EduMom	2026	0.9754	-0.7102	0.4776
		EduDad	2026	0.988	-0.3755	0.7073
		ADDxPC6	2026	7.70E+18	2.854	0.004314

Supplementary table S10: Logistic regression of rs507961 and rs9353525 with significant interaction of covarite of no interest by phenotype in IMAGEN. (STAT = coefficient t-statistic). PC1-PC6, First six ancestry components.

Tableau 18. – Suppl. Table S11 (article 1).

Covariants	Combo 110				Combo 220			
	BETA	SE	t value	p	BETA	SE	t value	p
Intercept	-1.211	0.648	-1.867	6.18E-02	-1.251	0.645	-1.939	5.25E-02
Combo	-0.691	0.241	-2.871	4.09E-03	-1.455	0.612	-2.378	1.74E-02
Sex	-0.375	0.102	-3.689	2.25E-04	-0.37	0.102	-3.644	2.68E-04
PC1	4.379	2.712	1.615	1.06E-01	4.553	2.707	1.682	9.25E-02
PC2	3.615	2.697	1.34	1.80E-01	3.533	2.692	1.313	1.89E-01
PC3	2.255	2.665	0.846	3.98E-01	2.22	2.655	0.836	4.03E-01
PC4	11.269	2.699	4.176	2.97E-05	11.408	2.695	4.232	2.31E-05
PC5	11.999	2.481	4.836	1.32E-06	11.831	2.482	4.767	1.87E-06
PC6	-7.407	6.865	-1.079	2.81E-01	-6.976	6.859	-1.017	3.09E-01
EduMom1	13.631	324.745	0.042	9.67E-01	13.709	324.745	0.042	9.66E-01
EduMom2	13.663	324.745	0.042	9.66E-01	13.748	324.745	0.042	9.66E-01
EduMom3	13.559	324.745	0.042	9.67E-01	13.624	324.745	0.042	9.67E-01
EduMom4	13.584	324.745	0.042	9.67E-01	13.659	324.745	0.042	9.66E-01
EduMom5	13.653	324.745	0.042	9.66E-01	13.746	324.745	0.042	9.66E-01
EduMom6	13.509	324.745	0.042	9.67E-01	13.62	324.745	0.042	9.67E-01
EduMom7	12.487	324.745	0.038	9.69E-01	12.535	324.745	0.039	9.69E-01
EduMom8	13.655	324.744	0.042	9.66E-01	13.701	324.744	0.042	9.66E-01
EduDad1	-13.148	324.744	-0.04	9.68E-01	-13.237	324.744	-0.041	9.67E-01
EduDad2	-13.04	324.744	-0.04	9.68E-01	-13.092	324.744	-0.04	9.68E-01
EduDad3	-13.246	324.744	-0.041	9.67E-01	-13.295	324.744	-0.041	9.67E-01
EduDad4	-13.035	324.744	-0.04	9.68E-01	-13.116	324.744	-0.04	9.68E-01
EduDad5	-13.47	324.744	-0.041	9.67E-01	-13.526	324.744	-0.042	9.67E-01
EduDad6	-12.954	324.744	-0.04	9.68E-01	-13.023	324.744	-0.04	9.68E-01
EduDad7	-13.907	324.745	-0.043	9.66E-01	-14.056	324.745	-0.043	9.65E-01
EduDad8	-13.583	324.744	-0.042	9.67E-01	-13.668	324.744	-0.042	9.66E-01

Supplementary table S11: Description of 2 significant model with risk SNPs combo in IMAGEN. The combo code is defined by minor allele numbers of rs484061(allele G), rs4963307(allele A) and rs7766029 (allele T) in this order.

Tableau 19. – Suppl. Table S12 (article 1).

Code Combo	Combo SNPs	SYS cohorts N(%)	IMAGEN cohorts N(%)
000	Low combo	9(0.01)	42(0.02)
001	High combo	20(0.03)	70(0.03)
002	Low combo	13(0.02)	29(0.01)
010	High combo	28(0.04)	68(0.03)
011	High combo	52(0.07)	130(0.06)
012	High combo	26(0.03)	57(0.03)
020	High combo	13(0.02)	38(0.02)
021	High combo	28(0.04)	66(0.03)
022	High combo	12(0.02)	27(0.01)
100	High combo	22(0.03)	71(0.03)
101	High combo	48(0.06)	122(0.06)
102	High combo	15(0.02)	64(0.03)
110	**Low combo**	**56(0.07)**	**132(0.06)**
111	High combo	91(0.12)	236(0.12)
112	Low combo	49(0.06)	131(0.06)
120	Low combo	30(0.04)	65(0.03)
121	High combo	54(0.07)	136(0.07)
122	Low combo	24(0.03)	54(0.03)
200	Low combo	7(0.01)	35(0.02)
201	Low combo	20(0.03)	65(0.03)
202	High combo	15(0.02)	31(0.02)
210	High combo	24(0.03)	79(0.04)
211	Low combo	43(0.06)	126(0.06)
212	Low combo	25(0.03)	59(0.03)
220	**Low combo**	**10(0.01)**	**33(0.02)**
221	High combo	17(0.02)	61(0.03)
222	Low combo	21(0.03)	23(0.01)
All observation	-	772	2050

Supplementary Table S12: Table description of combo SNPs for low risk (low combo) and high risk (high combo) for alcoholism. The combo code is defined by minor allele numbers of rs484061(allele G), rs4963307(allele A) and rs7766029 (allele T) in this order.

Tableau 20. – Suppl. Table S13 (article 1).

Name SNP	Description	Chromatin states (Core 15-state model)	Chromatin states (25-state model using 12 imputed marks)	H3K4me1	H3K4me3	H3K27ac	H3K9ac	DNase
rs484061	Brain Hippocampus Middle	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh		
rs484061	Brain Substantia Nigra	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs484061	Brain Anterior Caudate		17_EnhW2	H3K4me1_Enh		H3K27ac_Enh		
rs484061	Brain Cingulate Gyrus	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh		
rs484061	Brain Inferior Temporal Lobe	7_Enh	15_EnhAF	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
rs484061	Brain Angular Gyrus	7_Enh	15_EnhAF	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
rs484061	Brain_Dorsolateral_Prefrontal_Cortex	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac Pro	
rs484061	Brain Germinal Matrix		17_EnhW2					
rs484061	Fetal Brain Female			H3K4me1_Enh				
rs484061	Fetal Brain Male	7_Enh		H3K4me1_Enh				
rs4963307	Brain Hippocampus Middle	7_Enh	12_TxEnhW	H3K4me1_Enh		H3K27ac_Enh		
rs4963307	Brain Substantia Nigra		17_EnhW2			H3K27ac_Enh		
rs4963307	Brain Anterior Caudate	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs4963307	Brain Cingulate Gyrus	7_Enh	12_TxEnhW	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs4963307	Brain Inferior Temporal Lobe		17_EnhW2	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs4963307	Brain Angular Gyrus		17_EnhW2	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs4963307	Brain_Dorsolateral_Prefrontal_Cortex		17_EnhW2	H3K4me1_Enh		H3K27ac_Enh		
rs4963307	Brain Germinal Matrix							
rs4963307	Fetal Brain Female			H3K4me1_Enh				
rs4963307	Fetal Brain Male			H3K4me1_Enh				
rs7766029	Brain Hippocampus Middle							
rs7766029	Brain Substantia Nigra							
rs7766029	Brain Anterior Caudate							
rs7766029	Brain Cingulate Gyrus							
rs7766029	Brain Inferior Temporal Lobe							
rs7766029	Brain Angular Gyrus							
rs7766029	Brain_Dorsolateral_Prefrontal_Cortex							
rs7766029	Brain Germinal Matrix							
rs7766029	Fetal Brain Female							
rs7766029	Fetal Brain Male							
rs507961	Brain Hippocampus Middle	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh		
rs507961	Brain Substantia Nigra	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs507961	Brain Anterior Caudate	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh		
rs507961	Brain Cingulate Gyrus	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs507961	Brain Inferior Temporal Lobe		15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs507961	Brain Angular Gyrus	7_Enh	15_EnhAF	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs507961	Brain_Dorsolateral_Prefrontal_Cortex	7_Enh	10_TxEnh5	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
rs507961	Brain Germinal Matrix	7_Enh	15_EnhAF	H3K4me1_Enh				

rs507961	Fetal Brain Female		15_EnhAF	H3K4me1_Enh				
rs507961	Fetal Brain Male		15_EnhAF	H3K4me1_Enh				
rs9353525	Brain Hippocampus Middle							
rs9353525	Brain Substantia Nigra							
rs9353525	Brain Anterior Caudate			H3K4me1_Enh		H3K27ac_Enh		
rs9353525	Brain Cingulate Gyrus							
rs9353525	Brain Inferior Temporal Lobe							
rs9353525	Brain Angular Gyrus							
rs9353525	Brain_Dorsolateral_Prefrontal_Cortex							
rs9353525	Brain Germinal Matrix							
rs9353525	Fetal Brain Female							DNase
rs9353525	Fetal Brain Male			H3K4me1_Enh				DNase

Supplementary table S13 : Regulatory chromatin states from DNAse and histone ChIP-Seq in brain for SNPs in this study (Roadmap Epigenomics Consortium, 2015 in https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php)(Black = missing data)

Tableau 21. – Suppl. Table S14 (article 1).

Name SNP	Chromosome	BP	A1	A2	Freq AC	Freq AU	Odd-Ratio	Pvalue
rs806368	6	88906819	C	T	0.1794	0.2174	0.7871	0.006816

Supplementary Table S14: Table of results for Fisher test of rs806368 (ALL) in IMAGEN. A1= minor allele. A2= major allele. Freq AC= Frequency of minor allele in cases. Freq AU = frequency of minor allele in controls. OR= Odds ratio

Tableau 22. – Suppl. Table S15 (article 1).

SNP	Covariables	Odd-Ratio	STAT	Pvalue
rs806368 (N=2027)	SNP	0.7953	-2.499	0.01246
	PC1	138.9	1.838	0.066
	PC2	22.09	1.173	0.2409
	PC3	4.893	0.607	0.5436
	PC4	1.65E+04	3.825	0.000131
	PC5	8.24E+04	4.664	3.10E-06
	PC6	0.001044	-0.996	0.3193
	Alcohol abuse history in family	1.419	2.323	0.0202
	Education of mother	0.9773	-0.657	0.5113
	Education of father	0.9846	-0.484	0.6286
	Gender	0.69	-3.665	0.0002471

Supplementary Table S15: Table of results for logistic model with AUDIT and rs806368 N= Number of Non Missing individuals in IMAGEN. OR= Odds Ratio. Stat= Coefficient t-statistic.

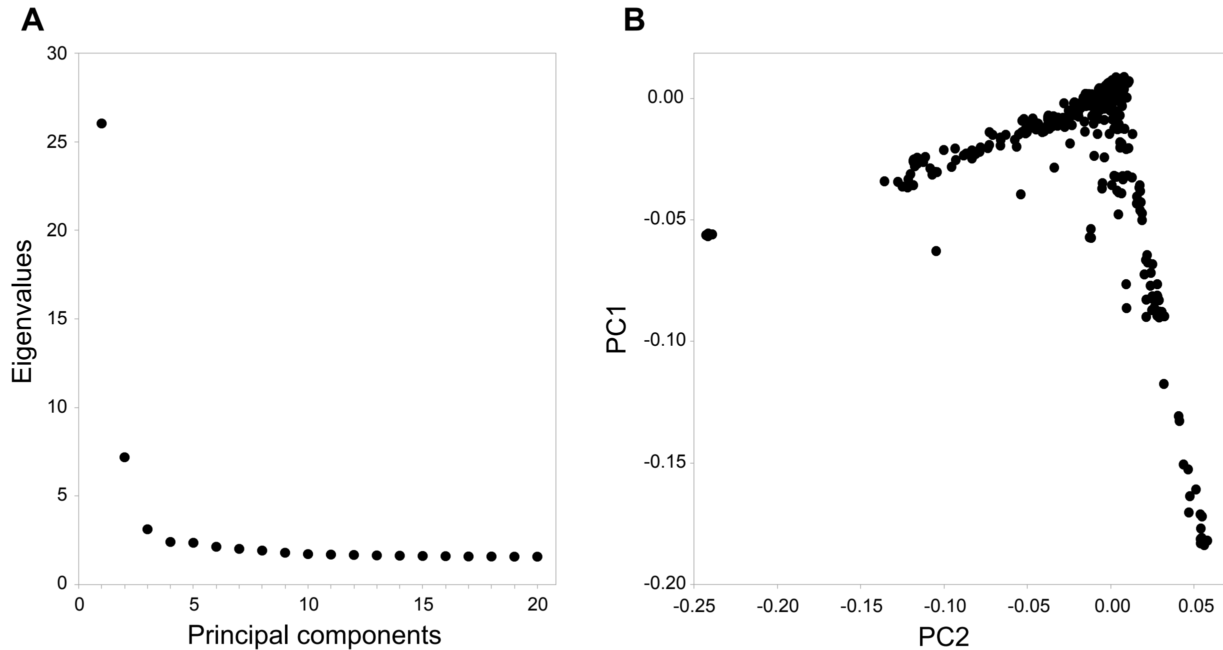
Tableau 23. – Suppl. Table S16 (article 1).

Name SNP	Position Weight Matrix ID(Library from Kheradpour and Kellis, 2013)
rs7766029	TCF4_known2
rs484061	AP-1_disc8
rs484061	Dlx3
rs484061	Hoxb6
rs484061	STAT_known13

Supplementary Table S16: Description of regulatory motifs altered for SNPs associated to GMDR model

Supplementary Figures

Figure 4. – Suppl. Figure S1 (article 1).



Supplementary Figure S1. Principal Component Analysis of Ancestry IMAGEN. Illustrations to ancestries information in Imagen. A) Distribution of Eigenvalues by principal components calculated on genetic distances in Imagen. B) Multidimensional scaling plots of the two first principal components.

Bibliography – Annex A

1. Chung T, Colby SM, Barnett NP, Rohsenow DJ, Spirito A, Monti PM. Screening adolescents for problem drinking: performance of brief screens against DSM-IV alcohol diagnoses. J Stud Alcohol. 2000 Jul;61(4):579–87.

2. Knight JR, Sherritt L, Harris SK, Gates EC, Chang G. Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcohol Clin Exp Res. 2003 Jan;27(1):67–73.

3. Fairlie AM, Sindelar HA, Eaton CA, Spirito A. Utility of the AUDIT for screening adolescents for problematic alcohol use in the emergency department. Int J Adolesc Med Health. 2006 Mar;18(1):115–22.

4. Reinert DF, Allen JP. The alcohol use disorders identification test: an update of research findings. Alcohol Clin Exp Res. 2007 Feb;31(2):185–99.

Annex B - Supplementary data Article 2

Independent contribution of polygenic risk for schizophrenia and cannabis use in predicting psychotic-like experiences in young adulthood: Testing gene x environment moderation and mediation.

Laurent Elkrief [1*], Bochao Lin[3], Mohammad H Afzali[1,3], Tobias Banaschewski[4]; Arun L.W. Bokde[5]; Erin Burke Quinlan[6]; Sylvane Desrivières[6]; Herta Flor[7,8]; Hugh Garavan[9]; Penny Gowland [10]; Andreas Heinz M.D., Ph.D.[11]; Bernd Ittermann [12]; Jean-Luc Martinot [13], Marie-Laure Paillère Martinot [14]; Frauke Nees [4, 7]; Dimitri Papadopoulos Orfanos Ph.D.[9]; Tomáš Paus [15]; Luise Poustka[16]; Sarah Hohmann[4]; Sabina Millenet[4]; Juliane H. Fröhner[17]; Michael N. Smolka[17]; Henrik Walter[11]; Robert Whelan [18]; Gunter Schumann[6, 19]; Jurjen Luykx [2,20,21]; Marco P. Boks[21]; Patricia J. Conrod [1,3], and the IMAGEN consortium.

Supplemental material

Imagen Dataset

The IMAGEN study is a longitudinal imaging genetics study of 2087 healthy adolescents, mostly of European descent. Detailed descriptions of this study, genotyping procedures, and data collection have previously been published[1]. The current study uses data for all 2087 individuals who completed the IMAGEN assessment battery at 16, and 18 and who contributed their genetic data at 14 years of age. The multicentric IMAGEN project had obtained ethical approval by the local ethics committees (at their respective sites) and written informed consent from all participants and their legal guardians. The data to be used in this study are genetic, biological sex, behavioural measures (cannabis use), and psychotic-like experiences measures (CAPE questionnaire). The parents and adolescents provided written informed consent and assent, respectively at 14 and 16, and then participants gave full consent at 18 and 21 years of age. IMAGEN participants at 14, 16, 18 and 21 years of age were repeatedly assessed on cannabis use outcomes using the ESPAD questionnaire, and on psychotic like experiences using the CAPE questionnaire at 18 and 21 years of age. These questionnaires are completed by participants at home via the Psytools portal. Data is then de-identified through a two-step procedure were participants genetic and neuroimaging data are linked to the cognitive psytools data using two “pseudocodes” to ensure personal and test center anonymity [1].

Utrecht Cannabis Cohort

This study recruited over 20,000 Dutch-speaking adolescents and young adults. Recruitment strategies are described in more detail in [2]. Participants answered questions on cannabis use history, along with the Community Assessment of Psychic Experiences questionnaire and provided their age, educational level and contact details. Data was collected from June 2006 to February 2009, which resulted in 21 838 participants. The assessment included two verification questions to protect against random answers, if participants failed to answer these questions they were excluded. After exclusion 17 698 participants remained (81% of 21 838). The study was approved by the University Medical Centre Utrecht medical ethical commission and all participants gave online informed consent. Of these 17698 participants, 1259 provided genetic information, via two waves.

Quality control

IMAGEN

SNPs with a minor allele frequency (MAF) of less than 2%, a genotyping rate of 2% or SNPs that did not respect Hardy Weinberg Equilibrium (HWE) (<1x10^-6$^{\wedge-6}$) were removed from analysis. Individuals with disproportionate levels of individual missingness (<2%), ambiguous sex, evidence of cryptic relatedness (>0.125), excessive heterozygosity were removed. The SNP coordinates were updated from hg18 to hg19 using Illumina information and the liftover tool from the genome browser. (http://genome.ucsc.edu/cgi-bin/hgLiftOver). After the first steps of quality control, 1950 individuals remained. The data of the IMAGEN cohort were then combined with data from HapMap III, and principal component analysis was performed in PLINK1.9[3] to determine ancestry information. We removed individuals who did not fall within 3SD of the mean of the first 2 principal components of the CEU + TSI populations from HapMapIII[4]. In all, 1740 individuals remained for polygenic risk score and regression analysis.

Utrecht Cannabis Cohort

DNA was extracted from whole blood of two 10 ml EDTA tubes obtained using venipuncture. For logistic reasons, genotype data for individuals of Dutch ancestry were generated on two different

array platforms; 576 individuals on Illumina® HumanOmniExpress (733,202 SNPs) and 768 individuals on the Illumina® Human610-Quad Beadchip (620,901 SNPs).

Base Data Set

CLOSUK + PGS (Pardiñas et al. 2018)

To build polygenic risk score for schizophrenia (PRS-Sz) we use a most recent and largest schizophrenia GWAS to date [5] as a training set. The summary statistics used for PRS construction came from a meta-analysis of the CLOZUK sample (treatment resistant schizophrenia) and independent psychiatric genomics consortium (PGC) datasets (schizophrenia) (total 40,675 cases and 64,643 control). The CLOZUK sample was demonstrably similar to previous PGC schizophrenia sample[5]. The entirety of the sample (meta-analysis results) was included for construction of PRS, as it was determined that the CLOZUK + PGC sample was independent and non-overlapping to the IMAGEN and CannabisQuest samples.

Cannabis Use (Pasman et al., 2018)

We use the publicly available GWAS meta-analysis results of from the International Cannabis Consortium (ICC) for the creation of the cannabis use polygenic risk score. The ICC GWAS separated cases and controls as a binary lifetime measure of cannabis use, "yes" or "no". The CannabisQuest cohort data contributed to this GWAS, and as such, a cannabis use PRS was not constructed for the CannabisQuest cohort: a leave-one out dataset was not available. Although data from a 23andMe cohort is available online for use with the ICC cannabis cohort (PGC), we did not have data use agreements in place, and thus the did not include this data into our PGC calculations. Overall, data for PRS construction was from 162 082 individuals.

Polygenic Risk Score

Polygenic Risk Scores for schizophrenia (PRS-Sz) were constructed for each IMAGEN and Utrecht individuals based on data from the most recent schizophrenia GWAS based on 40 675 cases and 64 643 controls [5] as a training set. Prior to PRS building, SNPs from the base set were removed if they had an MAF <0.01, and SNP with imputation quality < 0.8. Overlapping SNPs between the GWAS summary statistics (base dataset), 1000 reference Genome (reference dataset), and our

dataset (target dataset; IMAGEN or CannabisQuest) were selected. Then 1) insertion or deletion, ambiguous SNPs, 2) SNPs with minor allele frequency (MAF)<0.01 and SNP with imputation quality (R2) < 0.8 in both training dataset and target dataset 3) SNPs located in complex-LD regions (supplementary table 1) were excluded[6]. Odds ratios for autosomal SNPs reported in the summary statistics were log-converted to beta values. Linkage disequilibrium-based clumping, using the PRSice default setting (i.e. clumb index of 250kb, r2 0.1, clump-p 0.1) was used to identify a set of independent. PRS scores for 12 different p-value thresholds are calculated (Pt levels) for each individual as a sum of each effect size for included SNPs. PRS were calculated using PRSice with for 12 GWAS p-value thresholds: 5 x 10-8, 5 x 10-7, 5 x 10-6, 5 x 10-5, 5 x 10-4, 5 x 10-3, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5. The covariates gender, age, and first ten principal components were included in analyses.

The PRS for cannabis use (PRS-CanUse) was constructed similarly to previously published methods [7]. While the Johnson group used PLINK1.9[3] *clump* and *score* procedures, we created PRS using comparable methods in PRSice2[8]. Clumping was however with respect to 1000 Genome Phase 3 european samples, as it was done in the Johnson work. Moreover a 500kb physicial distance and LD threshold of r^2>=0.25 were used. We calculated PRS for 12 GWAS p-value thresholds: 5 x 10-8, 5 x 10-7, 5 x 10-6, 5 x 10-5, 5 x 10-4, 5 x 10-3, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5. Considering that PRS-Can use was not predictive of cannabis use at 16 years of age, we calculated PRS for cannabis users at 18 years of age in the IMAGEN cohort. The PRS most predictive of case/control (>10 lifetime cannabis uses), was Pt = 0.05, and this PRS was therefore used as a covariable in sensitivity analyses.

Supplementary Figures and Tables

Figure 5. – Suppl. Figure S1 (article 2).

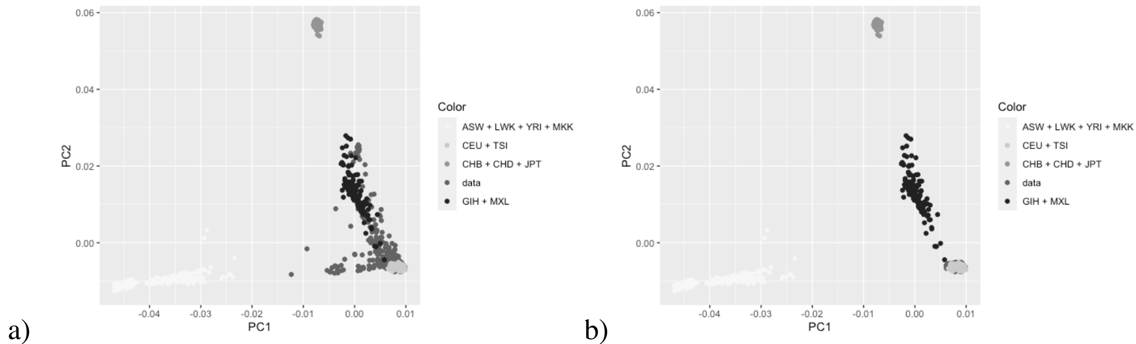


Figure S1. Principle component analysis. A comparison of the IMAGEN dataset including non-europeans individuals (a., N=1950) to the final IMAGEN dataset after removal of individuals who are not within 3SD of the mean of the first two principal components of the CEU+TSI populations of HapMap III (b., N=1740).

Abbreviations: ASW (n=49): African ancestry in Southwest USA; LWK (n=90): Luhya in Webuye, Kenya; MKK (n=143): Maasai in Kinyawa, Kenya; YRI (n=116): Yoruba in Ibadan, Nigeria; CHB (n=84): Han Chinese in Beijing, China; CHD (n=85): Chinese in Metropolitan Denver, Colorado; JPT (n=86): Japanese in Tokyo, Japan; GIH (n=88): Gujarati Indians in Houston, Texas; MEX (n=52): Mexican ancestry in Los Angeles, CA, United States; CEU (n=112): Utah residents with Northern and Western European ancestry from the CEPH collection; TSI (n=88): Toscani in Italia. Data = IMAGEN dataset

Tableau 24. – Suppl. Table S1 (article 2).

Chromosome	Base pair start	Base pair end
6	25392021	33392022
8	111930824	114930824
11	46043424	57243424
1	48287980	52287979
2	86088342	101041482
2	134666268	138166268
2	183174494	190174494
3	47524996	50024996
3	83417310	86917310
3	88917310	96017310
5	44464243	50464243
5	97972100	100472101
5	128972101	131972101
5	135472101	138472101
6	56892041	63942041
6	139958307	142458307
7	55225791	66555850
8	7962590	11962591
8	42880843	49837447
10	36959994	43679994
11	87860352	90860352
12	33108733	41713733
12	111037280	113537280
20	32536339	35066586

Suppl. Table S1 - Complex-LD regions removed prior to PRS construction

Tableau 25. – Suppl. Table S2 (article 2).

		Female (N=880)	Male (N=860)	Total (N=1740)	p value
<u>IMAGEN</u>					
Cannabis Use					
Case/Control					
	0	669 (92.7%)	588 (85.8%)	1257 (89.3%)	< 0.001
	1	53 (7.3%)	97 (14.2%)	150 (10.7%)	
	Missing	158	175	333	
CAPE42					
Positive Dimension					
	Mean (SD)	24.79 (4.22)	24.82 (4.52)	24.8 (4.36)	0.8
	Min - Max	19 - 54	19 - 55	19 - 55	
	Missing	269	315	584	
Depressive Dimension					
	Mean (SD)	14.3 (3.81)	12.77 (3.1)	13.58 (3.57)	< 0.001
	Min - Max	8 – 30	8 – 28	8 – 30	
	Missing	269	315	584	
Negative Dimension					
	Mean (SD)	22.46 (5.61)	22.05 (5.61)	22.265 (5.61)	0.2
	Min - Max	14 - 45	14 - 44	14 – 45	
	Missing	269	315	584	
Total CAPE-42					
	Mean (SD)	61.54 (11.65)	59.64 (11.32)	60.65 (11.53)	0.005
	Min - Max	41 - 115	41 - 121	41 - 121	
	Missing	269	315	584	
<u>Utrecht cannabis</u>					
Age					
	Mean (SD)	20.34 (2.39)	20.58 (2.46)	20.45 (2.42)	0.08
	Min - Max	16 - 40	16 - 39	16 – 40	
	Missing	0	0	0	
Cannabis Use					
Case/Control					
	0	516 (78.4%)	227 (40.2%)	743 (60.8%)	< 0.001
	1	142 (21.6%)	338 (59.8%)	480 (39.2%)	
	Missing	0	0	0	
CAPE42					
Positive Dimension					
	Mean (SD)	28.24 (8.32)	28.01 (5.41)	28.13 (7.12)	0.6
	Min - Max	20 – 175	20 – 49	20 - 175	
	Missing	0	0	0	
Depressive Dimension					

	Mean (SD)	14.45 (3.67)	13.09 (3.32)	13.82 (3.58)	< 0.001
	Min - Max	8 - 40	8 – 28	8 - 40	
	Missing	0	0	0	
Negative Dimension					
	Mean (SD)	25.42 (6.61)	25.93 (6.93)	25.65 (6.76)	0.2
	Min - Max	15 - 58	15 - 47	15 - 58	
	Missing	0	0	0	
Total CAPE-42					
	Mean (SD)	67.77 (14.31)	67.03 (13.59)	67.43 (13.98)	0.4
	Min - Max	23 - 146	43 - 110	23 – 146	
	Missing	0	0	0	

Supplementary Table S2. Summary statistics. T-test are performed to determine difference of means between male and female subjects. Missing = number of individuals with missing phenotype data.

Tableau 26. – Suppl. Table S3(article 2).

IMAGEN	**p-value threshold**		**β**	**std.error**	**p**
	p ≤ 5e−2				
		PRS-Sz	0.203	0.088	0.022
		SEX	-0.780	0.183	2.10×10^{-05}
		PC1	0.072	0.118	0.545
		PC2	-0.062	0.091	0.495
		PC3	0.045	0.120	0.708
		PC4	0.217	0.120	0.070
		PC5	-0.075	0.090	0.408
		PC6	-0.285	0.160	0.075
	p ≤ 0.5				
		PRS-Sz	0.269	0.089	0.003
		SEX	-0.768	0.184	2.90×10^{-05}
		PC1	0.084	0.119	0.479
		PC2	-0.061	0.091	0.499
		PC3	0.056	0.121	0.640
		PC4	0.219	0.120	0.069
		PC5	-0.072	0.091	0.426
		PC6	-0.282	0.160	0.078
Utrecht cannabis					
	p ≤ 5e−2				
		PRS-Sz	0.084	0.017	7.31×10^{-7}

	SEX	-1.725	0.131	$2.00x10^{-16}$
	Age	0.046	0.027	0.082
	PC1	-7.989	8.150	0.327
	PC2	-4.728	8.334	0.571
	PC3	-5.698	8.319	0.493
	PC4	-6.840	8.393	0.415
	PC5	-12.125	8.380	0.148
	PC6	-5.500	8.254	0.505
p ≤ 0.5	PRS-Sz	0.393	0.069	$9.62x10^{-9}$
	SEX	-1.725	0.131	$2.00x10^{-16}$
	Age	0.046	0.027	0.082
	PC1	-7.989	8.150	0.327
	PC2	-4.728	8.334	0.571
	PC3	-5.698	8.319	0.493
	PC4	-6.840	8.393	0.415
	PC5	-12.125	8.380	0.148
	PC6	-5.500	8.254	0.505

Supplementary Table S3. Predictive value of PRS-Sz on cannabis use. This table shows results for the most predicative polygenic risk score on cannabis use. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and Sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. β= Beta, std.error = standard error, p = p value. PRS-Sz = polygenic risk score for schizophrenia. pt=p-value threshold.

Tableau 27. – Suppl. Table S4(article 2).

	p threshold	Nagelkerke R2	p value	CNT
IMAGEN				
	$5x10^{-8}$	0.019	0.715	146
	$5x10^{-7}$	0.015	0.747	221
	$5x10^{-6}$	0.000	0.989	431
	$5x10^{-5}$	0.719	0.024	865
	$5x10^{-4}$	0.771	0.019	2k
	0.005	0.819	0.016	5k
	0.05	0.738	0.022	17k
	0.1	0.694	0.026	25k
	0.2	0.975	0.009	38k
	0.3	1.155	0.004	47k
	0.4	1.191	0.004	55k
	0.5	1.279	0.003	61k
	1	1.191	0.004	82k
Utrecht cannabis				
	$5x10^{-8}$	0.105	0.291	189
	$5x10^{-7}$	0.040	0.515	273
	$5x10^{-6}$	0.043	0.500	497
	$5x10^{-5}$	0.321	0.065	948
	$5x10^{-4}$	0.351	0.054	2k
	0.005	1.174	$4.53 x10^{-4}$	6k
	0.05	2.376	$7.31 x10^{-7}$	18k
	0.1	2.185	$1.98 x10^{-6}$	27k
	0.2	2.987	$3.25 x10^{-8}$	39k
	0.3	2.962	$3.63 x10^{-8}$	49k
	0.4	2.997	$3.02 x10^{-8}$	57k
	0.5	3.223	$9.62 x10^{-9}$	64k
	1	3.144	$1.40 x10^{-8}$	86k

Supplementary Table S4. Predictive value of various PRS-Sz on cannabis use. Above are results from the logistic regression measuring predictive value of PRS-Sz (independent variable) and case-control cannabis use (dependent variable). We include the first 6 PC and Sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. Nagelkerke R2 is used to calculate variance. CNT is number of SNPs included in the PRS. p threshold = p-value threshold.

Tableau 28. – Suppl. Table S5 (article 2).

			β	**std.error**	**p.value**
IMAGEN (N=1156)	p-value threshold				
	p ≤ 5e−2				
		PRS-Sz	0.011	0.005	*3.57×10^{-2}*
		SEX	0.030	0.010	*3.85×10^{-3}*
		PC1	0.005	0.007	4.96×10^{-1}
		PC2	-0.004	0.005	4.47×10^{-1}
		PC3	-0.010	0.007	1.46×10^{-1}
		PC4	-0.006	0.007	3.54×10^{-1}
		PC5	-0.003	0.005	6.19×10^{-1}
		PC6	0.006	0.010	5.54×10^{-1}
	p ≤ 0.5				
		PRS-Sz	0.013	0.005	*1.11×10^{-2}*
		SEX	0.031	0.010	*3.25×10^{-3}*
		PC1	0.006	0.007	4.37×10^{-1}
		PC2	-0.004	0.005	4.50×10^{-1}
		PC3	-0.010	0.007	1.60×10^{-1}
		PC4	-0.006	0.007	3.72×10^{-1}
		PC5	-0.003	0.005	6.32×10^{-1}
		PC6	0.006	0.010	5.49×10^{-1}
Utrecht cannabis (N=1213)	p-value threshold				
	p ≤ 5e−2				
		PRS-Sz	0.005	0.001	*2.64×10^{-4}*
		SEX	0.010	0.011	3.79×10^{-1}
		age	-0.001	0.002	6.43×10^{-1}
		PC1	1.037	0.715	1.47×10^{-1}
		PC2	1.361	0.728	6.19×10^{-2}
		PC3	0.035	0.733	9.62×10^{-1}
		PC4	-0.693	0.732	3.44×10^{-1}
		PC5	-0.512	0.732	4.85×10^{-1}
		PC6	-0.296	0.732	6.86×10^{-1}
	p ≤ 0.5				
		PRS-Sz	0.005	0.001	*1.63×10^{-5}*
		SEX	0.011	0.011	3.23×10^{-1}
		age	-0.001	0.002	5.94×10^{-1}
		PC1	1.179	0.716	1.00×10^{-1}
		PC2	1.335	0.726	6.63×10^{-2}
		PC3	0.004	0.732	9.95×10^{-1}
		PC4	-0.663	0.731	3.64×10^{-1}
		PC5	-0.502	0.731	4.92×10^{-1}
		PC6	-0.373	0.730	6.09×10^{-1}

Supplementary Table S5. Predictive value of PRS-Sz on psychotic-like experiences. This table shows results for the most predicative polygenic risk score on CAPE-42 total responses. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. β= Beta, std.error = standard error, p = p value. PRS-Sz = polygenic risk score for schizophrenia. pt=p-value threshold.

Tableau 29. – Suppl. Table S6 (article 2).

IMAGEN				
p-value threshold		β	std.error	p.value
p ≤ 5e−2				
	PRS-Sz	0.020	0.007	***4.58×10^{-3}***
	SEX	0.106	0.014	***5.67×10^{-14}***
	PC1	0.005	0.010	5.99×10^{-1}
	PC2	-0.004	0.007	5.98×10^{-1}
	PC3	-0.005	0.009	5.71×10^{-1}
	PC4	-0.008	0.009	4.04×10^{-1}
	PC5	-0.008	0.007	2.53×10^{-1}
	PC6	-0.010	0.013	4.42×10^{-1}
p ≤ 0.5				
	PRS-Sz	0.022	0.007	***1.53×10^{-3}***
	SEX	0.107	0.014	***3.29×10^{-14}***
	PC1	0.006	0.010	5.29×10^{-1}
	PC2	-0.004	0.007	6.11×10^{-1}
	PC3	-0.005	0.009	6.06×10^{-1}
	PC4	-0.007	0.009	4.25×10^{-1}
	PC5	-0.008	0.007	2.63×10^{-1}
	PC6	-0.010	0.013	4.49×10^{-1}
Utrecht cannabis				
p ≤ 5e−2				
	PRS-Sz	0.006	0.002	***9.15×10^{-4}***
	SEX	0.101	0.013	***1.06×10^{-13}***
	AGE	0.004	0.003	1.11×10^{-1}
	PC1	0.128	0.839	8.79×10^{-1}

	PC2	1.583	0.854	6.39×10^{-2}
	PC3	-0.742	0.860	3.88×10^{-1}
	PC4	-1.077	0.859	2.10×10^{-1}
	PC5	-0.955	0.859	2.67×10^{-1}
	PC6	-0.129	0.858	8.81×10^{-1}
$p \leq 0.5$				
	PRS-Sz	0.005	0.001	***4.50×10^{-4}***
	SEX	0.102	0.013	***5.67×10^{-14}***
	AGE	0.004	0.003	1.26×10^{-1}
	PC1	0.230	0.841	7.85×10^{-1}
	PC2	1.549	0.853	6.97×10^{-2}
	PC3	-0.754	0.860	3.80×10^{-1}
	PC4	-1.048	0.858	2.22×10^{-1}
	PC5	-0.951	0.858	2.68×10^{-1}
	PC6	-0.215	0.857	8.02 × 10-1

Supplementary Table S6. Predictive value of PRS-Sz on depressive sub-scale of CAPE-42. This table shows results for the most predicative polygenic risk score on CAPE-42 the depression sub-scale responses. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. β= Beta, std.error = standard error, p = p value. PRS-Sz = polygenic risk score for schizophrenia. pt=p-value threshold.

Tableau 30. – Suppl. Table S7 (article 2).

IMAGEN				
p-value threshold		β	std.error	p.value
p ≤ 5e−2				
	PRS-Sz	0.003	0.005	4.76×10^{-1}
	SEX	-0.002	0.009	8.64×10^{-1}
	PC1	-0.001	0.007	9.13×10^{-1}
	PC2	-0.003	0.005	5.59×10^{-1}
	PC3	-0.019	0.006	***3.17×10^{-3}***
	PC4	0.002	0.006	7.58×10^{-1}
	PC5	0.003	0.005	4.90×10^{-1}
	PC6	0.027	0.008	***1.41×10^{-3}***
p ≤ 0.5	PRS-Sz	0.005	0.005	3.02×10^{-1}
	SEX	-0.001	0.009	8.78×10^{-1}
	PC1	0.000	0.007	9.53×10^{-1}
	PC2	-0.003	0.005	5.54×10^{-1}
	PC3	-0.018	0.006	***3.45×10^{-3}***
	PC4	0.002	0.006	7.45×10^{-1}
	PC5	0.003	0.005	4.85×10^{-1}
	PC6	0.027	0.008	***1.40×10^{-3}***
Utrecht cannabis				
p ≤ 5e−2				
	PRS-Sz	0.004	0.001	***5.66×10^{-3}***
	SEX	-0.001	0.011	9.34×10^{-1}
	AGE	-0.007	0.002	***2.57×10^{-3}***
	PC1	0.523	0.704	4.58×10^{-1}
	PC2	1.002	0.717	1.63×10^{-1}
	PC3	-0.207	0.722	7.74×10^{-1}
	PC4	-0.877	0.721	2.24×10^{-1}
	PC5	-0.313	0.721	6.64×10^{-1}
	PC6	-0.147	0.720	8.38×10^{-1}
p ≤ 0.5				
	PRS-Sz	0.003	0.001	***2.73×10^{-3}***
	SEX	0.000	0.011	9.92×10^{-1}
	AGE	-0.007	0.002	***2.14×10^{-3}***
	PC1	0.601	0.706	3.95×10^{-1}
	PC2	0.978	0.716	1.72×10^{-1}
	PC3	-0.218	0.722	7.62×10^{-1}
	PC4	-0.856	0.720	2.35×10^{-1}
	PC5	-0.310	0.721	6.67×10^{-1}
	PC6	-0.207	0.719	7.73×10^{-1}

Supplementary Table S7. Predictive value of PRS-Sz on positive sub-scale of CAPE-42. This table shows results for the most predicative polygenic risk score on CAPE-42 the positive sub-scale

responses. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. β= Beta, std.error = standard error, p = p value. PRS-Sz = polygenic risk score for schizophrenia. pt=p-value threshold.

Tableau 31. – Suppl. Table S8 (article 2).

IMAGEN				
p value threshold		β	std.error	p.value
p ≤ 5e−2				
	PRS-Sz	0.014	0.007	5.90×10^{-2}
	SEX	0.020	0.014	1.53×10^{-1}
	PC1	0.012	0.010	2.35×10^{-1}
	PC2	-0.004	0.007	5.47×10^{-1}
	PC3	-0.003	0.010	7.40×10^{-1}
	PC4	-0.015	0.009	1.14×10^{-1}
	PC5	-0.007	0.007	3.70×10^{-1}
	PC6	-0.009	0.013	4.79×10^{-1}
p ≤ 0.5				
	PRS-Sz	0.016	0.007	***2.36×10^{-2}***
	SEX	0.021	0.014	1.40×10^{-1}
	PC1	0.013	0.010	2.04×10^{-1}
	PC2	-0.004	0.007	5.51×10^{-1}
	PC3	-0.003	0.010	7.72×10^{-1}
	PC4	-0.015	0.009	1.21×10^{-1}
	PC5	-0.006	0.007	3.79×10^{-1}
	PC6	-0.009	0.013	4.83×10^{-1}
Utrecht Cannabis				
p ≤ 5e−2				
	PRS-Sz	0.007	0.002	***8.85×10^{-5}***
	SEX	-0.015	0.015	2.90×10^{-1}
	AGE	0.004	0.003	1.86×10^{-1}
	PC1	1.556	0.911	8.79×10^{-2}
	PC2	1.244	0.927	1.80×10^{-1}
	PC3	0.260	0.934	7.81×10^{-1}
	PC4	-0.619	0.932	5.07×10^{-1}
	PC5	-0.658	0.933	4.81×10^{-1}
	PC6	0.011	0.932	9.91×10^{-1}
p ≤ 0.5				
	PRS-Sz	0.006	0.001	***1.12×10^{-5}***
	SEX	-0.014	0.015	3.41×10^{-1}
	AGE	0.004	0.003	2.11×10^{-1}
	PC1	1.718	0.913	6.00×10^{-2}
	PC2	1.204	0.926	1.93×10^{-1}
	PC3	0.231	0.933	8.04×10^{-1}
	PC4	-0.579	0.931	5.34×10^{-1}
	PC5	-0.649	0.931	4.86×10^{-1}
	PC6	-0.097	0.930	9.17×10^{-1}

Supplementary Table S8. Predictive value of PRS-Sz on negative sub-scale of CAPE-42. This table shows results for the most predicative polygenic risk score on CAPE-42 the negative sub-scale responses. Also presented are results for PRS-Sz with a Pt of <0.05, as this Pt best explains schizophrenia risk. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. β= Beta, std.error = standard error, p = p value. PRS-Sz = polygenic risk score for schizophrenia. pt=p-value threshold.

Tableau 32. – Suppl. Table S9 (article 2).

Dependent Variable	p-value threshold	**β**	**std.error**	**p.value**

Lifetime cannabis use (case/control)				
	p ≤ 5e−8	0.032	0.088	0.715
	p ≤ 5e-5	0.201	0.089	0.024
Total Cape-42 - frequency				
	p ≤ 5e−8	0.001	0.005	0.801
	p ≤ 5e-5	0.012	0.005	0.021

Supplementary Table S9. Predictive value of various stringent PRS-Sz on cannabis use and CAPE-42 total scores. This table shows results for the logistic and linear regressions of various schizophrenia polygenic risk score on cannabis use (case/control status) and CAPE-42 responses. We include the first 6 PC and sex as covariates for all analyses. β= Beta, std.error = standard error, p = p value.

Tableau 33. – Suppl. Table S10 (article 2).

Dependent Variable	Independent Variable	β	std.error	p.value
Total Cape-42 - frequency				
	PRS-Sz (pt=0.5)	0.014	0.006	**0.013**
	PRS-CanUse (pt=0.05)	0.003	0.005	0.622

	Lifetime cannabis use (case/control)	0.038	0.018	**0.037**
Lifetime cannabis use (case/control)				
	PRS-Sz (pt=0.5)	0.268	0.089	0.002
	PRS-CanUse (pt=0.05)	0.090	0.089	0.313

Supplementary Table S10. Predictive value of PRS-Sz (Pt=0.5) on cannabis use and CAPE-42 total scores, when considering the polygenic-risk score for cannabis use as a covariate. This table shows results for the logistic and linear regressions of the schizophrenia polygenic risk score (Pt=0.5) on cannabis use (case/control status) and CAPE-42 responses. We include the first 6 PC and sex as covariates and the polygenic risk score for cannabis use (Pt=0.05) for all analyses. PRS-Sz = polygenic risk score for schizophrenia. PRS-CanUse = polygenic risk score for cannabis use. pt=p-value threshold. β= Beta, std.error = standard error, p = p value.

Tableau 34. – Suppl. Table S11 (article 2).

Models (Pt<0.5)	**Model Parameters**	**Estimate**	**95% CI Lower**	**95% CI Upper**	**p-value**
X: PRS-SZ	ACME	0.0004	-0.0002	0.0010	0.0952
M: Cannabis Use (Case/Control)	ADE	0.0138	0.0033	0.0242	0.0096
Y:PLE	Total Effect	0.0142	0.0046	0.0238	0.0060
	Prop. Mediated	0.0305	-0.0151	0.0762	0.1006
Models (Pt<0.05)					
X: PRS-SZ	ACME	0.0009	-0.0004	0.0022	0.1790
M: Cannabis Use (Case/Control)	ADE	0.0111	0.0006	0.0216	0.0370
Y:PLE	Total Effect	0.0120	0.0015	0.0225	0.0250
	Prop. Mediated	0.0757	-0.0579	0.2093	0.1970

ACME: Average Causal Mediation Effects, ADE: Average Direct Effects.

N=1061; Simulation=10000 (non-parametric confidence intervals)

Supplementary TableS11. Results of mediation analysis. Independent variable= Polygenic risk score for schizophrenia (PRS-Sz), Dependent variable=logTotal of Cape-42. Mediator = Cannabis use. As shown, the effect of PRS-Sz on psychotic like experiences as measured by the CAPE-42 questionaire was not mediated by cannabis use. The effect of PRS-Sz was significantly associated to cannabis us in both models as was cannabis use effect size onto CAPE-42 scores. The direct effect was significant. Unstandardized indirect effects were computed for each Of the 10000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles.

Tableau 35. – Suppl. Table S12 (article 2).

			β	**std.error**	**p.value**
IMAGEN	Pt<0.05				
		PRS-Sz	0.011	0.006	*4.82×10^{-2}*

		Cannabis Use	0.040	0.018	*2.68×10^{-2}*
		SEX	0.035	0.011	*1.36×10^{-3}*
		PC1	0.005	0.008	4.81×10^{-1}
		PC2	-0.003	0.006	5.51×10^{-1}
		PC3	-0.009	0.007	2.01×10^{-1}
		PC4	-0.006	0.007	3.84×10^{-1}
		PC5	-0.002	0.006	6.83×10^{-1}
		PC6	0.010	0.010	3.25×10^{-1}
	Pt<0.5				
		PRS-Sz	0.014	0.006	*1.30×10^{-2}*
		Cannabis Use	0.038	0.018	*3.52×10^{-2}*
		SEX	0.036	0.011	*1.18×10^{-3}*
		PC1	0.006	0.008	4.22×10^{-1}
		PC2	-0.003	0.006	5.40×10^{-1}
		PC3	-0.009	0.007	2.13×10^{-1}
		PC4	-0.006	0.007	4.04×10^{-1}
		PC5	-0.002	0.006	6.99×10^{-1}
		PC6	0.010	0.010	3.15×10^{-1}
Utrecht cannabis					
	Pt<0.05				
		PRS-Sz	0.004	0.001	*3.82×10^{-3}*
		Cannabis Use	0.067	0.013	*1.43×10^{-7}*
		age	-0.002	0.002	4.63×10^{-1}
		SEX	0.035	0.012	*3.92×10^{-3}*
		PC1	1.148	0.708	1.05×10^{-1}
		PC2	1.425	0.720	*4.82×10^{-2}*
		PC3	0.114	0.725	8.75×10^{-1}
		PC4	-0.597	0.724	4.10×10^{-1}
		PC5	-0.351	0.725	6.29×10^{-1}
		PC6	-0.225	0.724	7.56×10^{-1}
	Pt<0.5				
		PRS-Sz	0.004	0.001	*5.70×10^{-4}*
		Cannabis Use	0.065	0.013	*3.57×10^{-7}*
		age	-0.002	0.002	4.34×10^{-1}
		SEX	0.036	0.012	*3.64×10^{-3}*
		PC1	1.259	0.709	7.61×10^{-2}
		PC2	1.402	0.719	5.14×10^{-2}
		PC3	0.087	0.724	9.04×10^{-1}
		PC4	-0.575	0.723	4.26×10^{-1}
		PC5	-0.348	0.724	6.31×10^{-1}
		PC6	-0.289	0.722	6.90×10^{-1}

Supplementary Table S12. Predictive value of PRS-Sz and cannabis use on CAPE-42. This table shows results for the various PRS-Sz on CAPE-42 while considering lifetime cannabis use (case/control status) as an independent covariable. We include the first 6 PC and sex as covariates for all analyses and age is included for all analyses of the CannabisQuest cohort. PRS-Sz =

polygenic risk score for schizophrenia. pt=p-value threshold. β= Beta, std.error = standard error, p = p value.

Bibliography – Annex B

1. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry. 2010;15:1128–1139.

2. Boks MP, He Y, Schubart CD, Gastel W van, Elkrief L, Huguet G, et al. Cannabinoids and psychotic symptoms: A potential role for a genetic variant in the P2X purinoceptor 7 (P2RX7) gene. Brain, Behavior, and Immunity. 2020;88:573–581.

3. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4.

4. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52–58.

5. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50:381–389.

6. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. Long-Range LD Can Confound Genome Scans in Admixed Populations. Am J Hum Genet. 2008;83:132–135.

7. Johnson EC, Tillman R, Aliev F, Meyers JL, Salvatore JE, Anokhin AP, et al. Exploring the relationship between polygenic risk for cannabis use, peer cannabis use and the longitudinal course of cannabis involvement. Addiction. 2019;114:687–697.

8. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience. 2019;8.