Genetic Factors in Cannabinoid Use and Dependence

7

Megan K. Mulligan

Abstract

Cannabinoid use and dependence are heritable traits controlled in part by genetic factors. Despite a high incidence of use worldwide, genes that contribute to the risk of problematic use and dependence remain enigmatic. Here we review human candidate gene association studies, family-based linkage studies, and genome-wide association studies completed within the last two decades. These studies have expanded the list of candidate genes and intervals. However, there is little overlap between studies and generally low reproducibility in independent samples. Reasons for this lack of coherence vary but may depend on low sample size and statistical power, and the fact that most studies leverage populations ascertained for drug dependence other than cannabis. However, recent well-powered studies on lifetime cannabis use demonstrate that the genetic architecture of cannabis use resembles that of other substance use disorders and psychiatric disease in that many small effect genes contribute in an additive fashion. This finding suggests that increasing sample size and more focused recruitment of individuals based on cannabinoid use and dependence will identify more candidate genes. Follow-up of existing high priority candidates in preclinical model systems will facilitate better understanding of the genetic architecture and genetic risk factors for cannabis use and dependence.

Abbreviations

2-AG	2-arachidonoylglycerol
AA	African American
AEA	N-arachidonylethanolamide or anandamide
CADM2	cell adhesion molecule 2
CB1	cannabinoid receptor type 1
CB2	cannabinoid receptor type 2
CD	cannabis dependence
CGAS	candidate gene association studies
Chr	chromosome
CNR1	gene encoding CB1
CUD	cannabinoid use disorder
DAG	1,2-diacylglycerol
DSM	Diagnostic and Statistical Manual of Mental Disorders
EA	European Americans
FAAH	fatty acid amide hydrolase

M. K. Mulligan (✉)
Department of Genetics, Genomics and Informatics,
The University of Tennessee Health Science Center,
Memphis, TN, USA
e-mail: mkmulligan@uthsc.edu

FLS   family-based linkage studies
GPCR   G-protein coupled receptor
GTEx   Genotype-Tissue Expression
GWAS   genome-wide association studies
iPSYCH   Initiative   for   Integrative
           Psychiatric Research
LD   linkage disequilibrium
LOD   logarithm of the odds
MGLL   monoacylglycerol lipase
nAChR   neuronal acetylcholine receptor
NAG   Nicotine   Addiction   Genetics
           Program
NAPE-PLD   N-acylphosphatidylethanolamine-
           specific phospholipase D
NCAM1   neural cell adhesion molecule 1
NRG1   neuregulin 1
PPAR   peroxisome proliferator-act ivated
           receptor
SCOC   short coiled-coil protein
SCOC-AS1 short coiled-coil protein anti-
           sense RNA 1
SNP   single nucleotide polymorphism
THC   Δ9-tetrahydrocannabinol
TRP   transient receptor potential (ion
           channel)

7.1   Introduction

Similar to other abused substances, genetic factors contribute substantially to cannabinoid use and dependence. Heritability for abuse of cannabinoids, primarily cannabis and its derivatives, ranges from 30% to 80% [1, 2]. Heritability measures the contribution of segregating gene variants to the total variation in a trait or phenotype of interest. For substance use disorders heritability has often been estimated from twin studies in which the concordance rate for cannabis-related traits is compared in monozygotic (genetically identical) and dizygotic (fraternal) twins. Higher concordance of a trait in monozygotic versus dizygotic twin pairs indicates a substantial contribution of genetic factors relative to environmental factors. Thus, high heritability is an indication that genetic factors contribute significantly to cannabinoid use disorders (CUDs). However, few of these factors have been eluci-

dated. In this chapter we review the growing list of genes associated with CUDs based on recent candidate gene, linkage, and genome-wide association studies in humans. Finally, we discuss reasons for and possible solutions to address the paucity of known genetic factors contributing to CUDs.

7.1.1   What Is Cannabis Use Disorder?

The term "substance use disorder" is derived from the Diagnostic and Statistical Manual of Mental Disorders: DSM-5 [3]. This manual defines a set of criteria used to diagnose problematic and recurrent use of drugs or alcohol that can impact health and social well-being. Symptoms of CUD include disruption in normal function caused by cannabis use and development of tolerance, craving, and/or withdrawal symptoms associated with increased or continued use. A cluster of withdrawal symptoms, including sleep disruptions, anxiety, anger, depression have been associated with early abstinence from cannabis and may contribute to continued use. As of 2018, there are no FDA-approved treatments for CUD.

Cannabis is one of the most widely consumed psychoactive substances worldwide [4]. In the United States, policy changes in individual state legislature beginning in 1996 for medical use and continuing in 2012 for recreational use are likely to increase cannabis use in this country. As the number of users increases, so does the risk of CUDs. According to recent epidemiological studies in the United States, CUD impacts ~6% of the population [5]. Key to the identification of individuals at risk for CUD and development of pharmacological interventions for CUD is a better understanding of the genetic factors contributing to the disease.

7.1.2   Introduction to Human Genetic Association Studies

Substance use disorders, like CUDs and other psychiatric diseases, are complex traits that are driven by the actions and interactions of multiple

genetic and environmental factors. In the case of genetic factors, variation in heritable traits is caused by inheritance of different gene alleles (e.g. polymorphisms or sequence variants) that confer differential gene regulation or function. Human genes often contain multiple polymorphisms which can impact a single base (single nucleotide polymorphism or SNP) or several (insertion or deletion). These polymorphisms are typically biallelic and exist in one of two forms that represent the ancestral sequence or the altered sequence. The frequency of each allele in the population being sampled is generally a major consideration. Alleles with a minor allele frequency less than 1% will be difficult to study because few individuals with the minor allele exist in the population. In this case, detection of significant associations between inheritance of alleles and trait variation will be difficult even with moderate sample sizes. For this reason, the majority of associations identified between human disease and allelic variation involve common variants with relatively abundant allele frequencies in most human populations. Although common, these variants typically have low penetrance (probability that inheritance of the variant will cause the phenotype being measured). Genetic studies of human disease over the past decade have revealed that most human diseases are complex polygenic traits that result from the inheritance of many small effect risk alleles acting in an additive fashion [6]. These associations typically involve common variants, but rare alleles can also impact disease. Rare alleles are present at low frequency or only in specific human populations, and can only be identified using specialized study designs and populations. It is important to note that for CUDs, and most other complex diseases, the absolute number of risk alleles, their frequency in the population, and their individual contribution to disease risk (genetic architecture) in humans is unknown. Thus, the goal of human genetic association studies is to evaluate the genetic architecture of disease and identify risk genes and alleles in order to identify vulnerable individuals and design effective intervention or treatment strategies.

Three main approaches have been used towards the goal of identifying the genes and gene variants associated with CUDs in humans. These are candidate gene association studies (CGAS), family-based linkage studies (FLS), and genome-wide association studies (GWAS). In the first approach (CGAS), known allelic variants within a candidate gene are tested for an association with the disease. Often the candidate gene is selected based on *a priori* evidence regarding involvement in a disease related pathway and/or the presence of known functional variants. In a typical CGAS, inheritance of candidate gene alleles are associated with disease risk using statistical models and a case-control or family-based study design. A benefit to CGAS is that only a few associations are tested at a time, resulting in less correction for multiple testing and more significant association scores. A caveat to CGAS is the biased and limited experimental design which may lead to inflation of the contribution of the candidate gene to the disease phenotype. In contrast, FLSs and GWASs represent unbiased methods to identify gene variants contributing to phenotypic variation or disease risk.

FLSs represent a genome-wide approach to identify loci that are associated with a trait or disease risk. FLSs compare pedigrees among families to assess the likelihood that affected individuals share the same allele at a polymorphic marker more often than would be expected by chance when compared to unaffected relatives. Markers found to be significantly linked to the disease or trait by FLS are postulated to be near the causal gene variant. However, the region of linkage in FLS is often quite large. Historically this has been due to smaller sample sizes and marker panels in the size range of hundreds to thousands. For these reasons, most FLS of CUDs typically result in the identification of large (~10 Mb) linked regions that contain hundreds of potential candidate genes. Resolution to a single candidate gene in the larger linked region is not possible.

In contrast, a typical GWAS tests the association between a phenotype and allele frequency at hundreds of thousands or millions of individual polymorphisms (typically SNPs). Most mamma-

lian genomes have been sequenced and there are many high-throughput sequencing and genotyping platforms available to identify the allelic variation (genotype) at each locus on a global scale for individuals in a population. Currently, the most cost-effective and high throughput strategies include the use of genotyping microarrays that profile inheritance at millions of common variants. Due to linkage disequilibrium (LD), all variants do not need to be genotyped. LD occurs because regions of the genome are inherited as small blocks of DNA from either parent. Polymorphic genetic markers contained within each LD block will be highly correlated with one another because they are inherited as a unit. Polymorphisms in adjacent blocks will be less well correlated., Therefore, representative tag SNPs can be used as a proxy for all variants within a region of high LD (regions with high LD that are inherited together are referred to as haplotype blocks). Genotypes for adjacent polymorphisms can be imputed later for individuals based on population haplotypes. Although it is assumed that genetic polymorphisms modulate gene function or expression, the causal gene variant is generally not known following a GWAS. It is important to remember that GWAS can identify candidate gene loci, but cannot generally identify causal variants, the impact of variants on gene function, or the biological mechanism by which the gene contributes to disease.

Relative to CGAS, FLS and GWAS are unbiased approaches that can lead to the detection of multiple loci containing genes and alleles that contribute to risk. Each loci exerts a small effect on disease risk (i.e. CUDs) and the sum of all risk alleles, referred to as polygenic risk, is a better overall predictor of disease risk that captures more of the genetic variability or heritability of the disease. However, the large number of tests performed in FLS and GWAS requires correction for multiple testing and results in severe statistical penalties. To account for the many linkage or association tests performed for each marker and the phenotype of interest, empirical P-values adjusted for multiple test correction are computed. Usually the adjustment is made following the results of hundreds to thousands of permuta-

tions of genotypes for individuals in a genetic study. The adjusted P-value is represented as the number of times a permuted logarithm of the odds (LOD) score for association is greater or equal to the maximum observed LOD score divided by the number of permutations plus one. The empirical adjusted P-value is also referred to as the genome-wide corrected P-value. For most GWAS studies this is set very low ($P <$ E-09). For these reasons, the sample sizes and association scores required to reach statistical significance are much higher compared to CGAS. Increased sample sizes in recent GWAS has led to the identification of more candidate genes and better models of polygenic risk (for a review see [6]). In addition, genotyping a larger (or infinite) number of markers using microarray or next-generation DNA sequencing has the potential to resolve linkage region or loci down to a single gene.

7.2 Candidate Genes Identified Through Human Association Studies

Relative to alcohol and other drugs of abuse the number of association studies performed for CUDs and related traits remains relatively small. In this section candidate genes and the evidence supporting them will be reviewed.

7.2.1 Candidate Gene Association Studies

For CUD most CGAS have focused on gene variants within the endocannabinoid system. Endocannabinoid signaling is critical for modulation of numerous biological processes including, response to natural rewards, learning and memory, emotional processing, motor coordination, pain, energy metabolism, fertility, development, and immune response. Major endogenous lipid ligands of the endocannabinoid system include N-arachidonylethanolamide (AEA or anandamide) and 2-arachidonoylglycerol (2-AG). Both are synthesized from membrane precursors by N-acylphosphatidylethanolamine-

Table 7.1 Results from candidate gene association studies

Phenotype	Genes	Effect	Study reference
Problematic use	CNR1 (A/G; rs806380)	Minor allele G protective	[8]
Dependence	CNR1 (A/G; rs806380)	Minor allele G protective	[7]
Withdrawal severity and craving	FAAH (C/A; rs324420)	Minor allele A protective	[13]
Withdrawal severity and craving	FAAH (C/A; rs324420)	Minor allele A protective	[14]
Dependence	FAAH (C/A; rs324420)	Minor allele A protective	[12]

specific phospholipase D (NAPE-PLD) in the case of AEA, and by 1,2-diacylglycerol (DAG) lipases DAGLα and DAGLβ in the case of 2-AG. Both endocannabinoids are catabolized by one of two enzymes—fatty acid amide hydrolase (FAAH) for AEA and monoacylglycerol lipase (MGLL) for 2-AG. Both ligands (AEA and 2-AG) act as agonists primarily at two G-protein coupled receptors (GPCRs), cannabinoid receptor type 1 (CB1) and type 2 (CB2), although they can also activate other receptors including GPCRs 55 and 119, and peroxisome proliferator-activated receptor (PPARs). AEA also acts as an agonist at transient receptor potential ion channels (TRPs). Note that each member of the endocannabinoid system is encoded by genes located in distinct genomic regions as opposed to localization within several gene clusters.

Of the known endocannabinoid signaling genes, variants in two—cannabinoid receptor type 1 (*CNR1*) and *FAAH*—have repeatedly been tested for associations with CUDs and related traits (Table 7.1). It is unknown whether variants in *CNR1* influence CB1 expression or function, however several variants have been associated with cannabis use or dependence [7, 8]. In particular the minor G allele of the rs806380 SNP is thought to exert a protective effect. A common missense variant (rs324420; C/A) that results in substitution of the amino acid threonine for pro-line in the FAAH enzyme has also been associated with CUDs and related traits.

In contrast to variants in *CNR1*, the missense mutation in *FAAH* has an impact on enzyme stability and function. Inheritance of both copies of the minor A allele (A/A homozygous genotype) results in lower expression and activity of the enzyme due to decreased stability and increased proteolysis [9, 10]. Of interest, the major allele associated with normal enzymatic activity is more frequently associated with risk or problematic cannabis use [11]. For example, inheritance of the major allele (C/C genotype) has been associated with CUD [12] and high cannabis withdrawal symptoms and craving [13, 14]. Although associations between traits related to CUD and variants in *CNR1* and *FAAH* have been reported, there are several studies for which these associations were not replicated [15] and overall there is no consensus regarding the involvement of these mutations in cannabis intake, withdrawal, and dependence.

Gene variants that impact the function of key enzymes involved in drug metabolism can also influence drug use and risk of developing use disorders. An example of this are functional variants in genes involved in alcohol metabolism (alcohol dehydrogenase and aldehyde dehydrogenase) which are among the strongest protective factors against development of alcohol dependence (reviewed in [16]). Functional variants in cannabinoid metabolizing genes exist but have not yet been associated with CUD. In the liver, the cytochrome P450 family of enzymes plays a role in processing cannabinoids. There are several functional variants that modulate expression or enzymatic activity of family members, in particular, polymorphisms in the P450 family member *CYP2C9* were found to influence metabolism of the synthetic cannabinoid JWH-018 that is a high affinity agonist at cannabinoid receptors [17]. Two mutations *CYP2C9*2* (cysteine substitution for arginine at amino acid residue 144) and *CYP2C9*3* (leucine substitution for isoleucine at amino acid residue 359) were found to increase or decrease metabolism of the synthetic cannabinoid, respectively. However, the impact of these variants on synthetic cannabinoid or cannabis use and dependence is not known.

7.2.2 Family-Based Linkage Studies

The first FLS was performed for adolescent cannabis use and dependence by Hopfer and colleagues [18]. The population used in this study included adolescents (ethnic distribution roughly 8% African American or AA, 37% Hispanic, 52% European American or EA, and 4% other) in a substance abuse treatment program in Denver and their genetically related siblings. Participants were part of a larger Colorado Center on Antisocial Drug Dependence study [19]. In total, 324 adolescent sibling pairs from 192 families were included. Cannabis use was also measured in an age-matched control sample drawn from the same population (community sample) consisting of 4843 individuals. The community sample was used to standardize cannabis dependence (CD) scores in the treatment samples. Repeated cannabis use was defined as using cannabis at least six times and CD was measured as the number of lifetime symptoms based on DSM-IV criteria. Cannabis use and dependence were much higher (99% and 59%, respectively) in the adolescent treatment probands relative to their siblings (55% and 14%, respectively) and compared to the community sample (5% prevalence of CD in age-matched controls). Parents and sibling pairs were genotyped for 374 markers covering the 22 autosomal chromosomes (the X and Y sex chromosomes were excluded). Two linkage regions were identified that met the criterion for suggestive linkage (P = 0.0004, LOD > 2.5) between inheritance of parental alleles at a marker and CD. Suggestive linkage regions were located on Chrs 3 (3q21 near marker D3S1267) and 9 (9q34 near marker D9S1826). No significant loci were found. The interval for linkage on Chr 3 was located roughly between markers D3S1271 and D3S1292 (101 to 132 Mb using the GRCh38/hg38 human genome assembly) and the interval for Chr 9 was located between marker D9S290 to the end of the chromosome (128.6 to 138 Mb). Because of the small number of markers used in the analysis, many genes (376 for Chr 3 and 305 for Chr 9) were located in each linkage interval. Although the Chr 3 suggestive linkage region includes *MGLL*, the gene encoding the major

enzyme responsible for catabolism of the endogenous cannabinoid 2-AG, the precise genes contributing to trait variation in this first study of CD remain elusive.

Following closely behind the first linkage analysis for CD were several larger FLS. Agrawal and colleagues [20] leveraged data from the Collaborative Study on the Genetics of Alcoholism (COGA; [21]) to perform a linkage analysis based on DSM-IV criteria for CD. The COGA population was unique because it consisted of many generations of families (~90% EA and ~10 AA) at high risk for alcoholism. Genotyping was performed for 1364 individuals with genetic high-risk for alcoholism using a microarray platform consisting of 1717 SNPs. A community sample of 984 individuals was not genotyped but used to address and correct for possible confounds associated with linkage analysis (e.g. gender, race, age). A suggestive locus (adjusted $p = 0.71$, LOD = 1.9) on Chr 14 spanning ~14 Mb from markers rs759364 to rs872945 (89.3 to 103 Mb and containing 311 gene models) was associated with CD in the mostly EA COGA cohort carrying risk alleles for alcohol dependence.

Agrawal and colleagues [22] performed linkage analysis for CD based on DSM-IV requirements on 3431 individuals from 289 Australian families comprising the Nicotine Addiction Genetics Program (NAG) [23]. These families (>90% Anglo-Celtic or Northern European ethnic origin) included siblings and parents with a lifetime history of heavy smoking (40 cigarettes in a 24 h period or 20 cigarettes per day during periods of heavy smoking). A community-based control sample of 5776 individuals was used to standardize phenotypes and correct for possible confounds associated with linkage analysis in the NAG cohort. The NAG cohort was genotyped for a panel of 381 autosomal markers. Factor analysis was performed on the abuse and dependence criteria to create a single cannabis problems factor score which accounted for the majority of the variance (>60%) among measures. Suggestive linkage regions for the cannabis problems factor score were identified on Chr 1 (~10 cM interval centered on marker D1S2841 located at 78.9 Mb)

and Chr 4 (~25 cM interval centered on marker D4S419 located at 18.7 Mb). Note that the cM is a unit of genetic distance measured in map units. This unit of measure has historically been used in association studies where 1 cM corresponds to a recombination frequency of 1%. In humans, 1 cM is roughly equivalent to 1200 kb, but this varies between sexes and physical location on the chromosome. Again, linkage intervals identified in this study were too large to nominate single candidate genes.

In a separate and larger family study Agrawal and colleagues [24] performed linkage analysis for lifetime cannabis use, early-onset cannabis use, and frequency of cannabis use. The Australian cohort consisted of 5600 adult Australian twins, parents, and siblings from 2352 families genotyped at 1461 markers per individual. No markers passed the threshold for genome-wide significance. A suggestive linkage region ($P \approx 0.65$) on distal Chr 18 near marker D18S1360/GATA129F05 was identified for cannabis initiation (LOD = 1.97) and frequency of use (LOD = 2.14). A suggestive region was also located on proximal Chr 19 for early-onset cannabis use (LOD = 1.92). Marker position was not provided for all traits in the study so approximate linkage regions for this study are included in Table 7.2. Similar to previous FLS, relatively small sample sizes and marker panels provided low statistical power to detect linkage regions as well as poor resolution within suggestive linkage regions (hundreds of candidate genes located within large regions of linkage).

Ehlers and colleagues [25] analyzed a separate cohort of 1647 adults (92% Caucasian) from families with a history of alcoholism in order to identify loci associated with CD, craving, and withdrawal (feeling nervous, tense, restless, or irritable during abstinence from cannabis use). The probands were genotyped for 811 markers and a control sample of 147 individuals was used to access baseline phenotype rates. For CD, two suggestive loci were identified on Chrs 1 (LOD = 2.1, 17 cM interval near marker D1S498) and 2 (LOD = 2.6, 22 cM interval near marker D2S2361). Five loci were identified for craving on Chrs 7 (LOD = 5.7, 13 cM interval near

D7S502), 3 (LOD = 4.4, 12 cM interval near D3S1279), Chr 1 (LOD = 3.6, 12 cM interval near D1S199), and 6 (LOD = 3.2, 7 cM interval near D6S281). An additional two suggestive loci for craving were identified on Chrs 9 (LOD = 2.6, 19 cM interval near D9S157) and 15 (LOD = 2.3, 9 cM interval near D15S127). For withdrawal, the strongest linkage region was identified on Chr 9 (LOF = 3.6, 10 cM interval near D9S1838). Additional suggestive loci for withdrawal were identified on Chrs 3 (LOD = 2.5, 13 cM interval near D3S1566) and 7 (LOD = 2.2, 25 cM interval near D7S506). The withdrawal loci on Chrs 9 and 3 also demonstrated evidence of linkage for a phenotype related to sleep disruptions ("sleeplessness"). Because the population under study was recruited based on a family history of alcoholism, Ehlers and colleagues examined whether linkage regions for alcohol overlapped with CD and associated traits. None of the linkage regions identified for CD or craving and withdrawal phenotypes had previously been associated with alcohol related traits measured in the same cohort.

Finally, Han and colleagues [26] used a multi-stage design to identify gene variants associated with CD. Linkage analysis was first performed in two different ethnic study cohorts—AA (1022 individuals from 384 families) and EA (874 individuals from 355 families). Both cohorts were ascertained for cocaine and opioid dependence and selected families included at least two affected siblings for opioid or cocaine dependence based on DSM-IV criteria. Linkage was performed in each ethnic sample separately. The strongest linkage peak was identified on Chr 8 (8p2.11, LOD = 2.9) for the AA samples and another suggestive peak for these samples was also detected on Chr 14 (LOD = 2.26). In the EA samples, a suggestive linkage peak was detected on Chr 7 (LOD = 1.85). In the next stage of the analysis the authors used an independent data set from the Study of Addiction: Genetics and Environment (SAGE) that included 4036 unrelated individuals (275 AA cases and 401 controls and 422 EA cases and 1049 controls). GWAS was performed dependent on ethnic background for the strongest suggestive linkage peak identi-

Table 7.2 Results from family-based linkage analysis

Phenotype	Individuals	Families	Study information	Criterion	Approximate linkage region	Annotated genes	Ethnicity	Replication	Study reference
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Strong evidence	Chr1: 12 to 27 Mb	346	Caucasian	No	[25]
Dependence	3431	289	Nicotine Addiction Genetics Program	Suggestive	Chr1:73 to 85 Mb	85	Caucasian	No	[22]
Dependence	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr1: 141 to 161 Mb	579	Caucasian	No	[25]
Dependence	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr2: 202 to 229	304	Caucasian	No	[25]
Withdrawal	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr3: 62 to 78 Mb	122	Caucasian	No	[25]
Dependence	648	192	Colorado Center on Antisocial Drug Dependence	Suggestive	Chr3:101 to 132 Mb	376	Mixed	No	[18]
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Strong evidence	Chr3: 144 to 158 Mb	142	Caucasian	No	[25]
Dependence	3431	289	Nicotine Addiction Genetics Program	Suggestive	Chr4:4 to 34 Mb	374	Caucasian	No	[22]
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Strong evidence	Chr6: 165 to 174 Mb	118	Caucasian	No	[25]
Withdrawal	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr7: 38 to 68 Mb	312	Caucasian	No	[25]
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Strong evidence	Chr7: 60 to 75 Mb	161	Caucasian	No	[25]
Dependence	Multiple cohorts	Multiple cohorts	Multistage Analysis	Suggestive and strong evidence	Chr8: 32.6 Mb	NRGN1 (rs17664708)	AA	Yes	[26]
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr9: 6 to 30 Mb	151	Caucasian	No	[25]

Dependence	648	192	Colorado Center on Antisocial Drug Dependence	Suggestive	Chr9:128 to 138 Mb	305	Mixed	No	[18]
Withdrawal	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Strong evidence	Chr9: 132 to 144 Mb	229	Caucasian	No	[25]
Dependence	1364	143	Collaborative Study on the Genetics of Alcoholism	Suggestive	Chr14:89 to 103 Mb	311	Caucasian	No	[20]
Craving	1647	713	University of California San Francisco (UCSF) Family Alcoholism Study	Suggestive	Chr15: 85 to 96 Mb	147	Caucasian	No	[25]
Initiation and frequency of use	5600	2352	Australian Twin Registry	Suggestive	Chr18: 56 to 80 Mb	219	Caucasian	No	[24]

fied on Chr 8 in the FLS. A SNP (rs17664708) located in a candidate gene for schizophrenia risk, *NRG1*, was modestly associated with CD in both AAs and EAs. The association between CD and genotype at rs17664708 was replicated in an independent sample of AAs (758 dependent cases and 280 controls) but was not able to be replicated in an independent sample of EAs (568 dependent cases and 318 controls). Of interest, the *NRG1* variant is common in EA samples but rare in AA samples and was primarily associated with CD in AA samples. The variant in *NRG1* may also be associated with drug dependence in general as it was also associated with opioid dependence in the original AA study cohort ascertained for cocaine and opioid dependence.

Some of the first genetic studies for CUD and related traits involved FLS that included less than 10,000 individuals, less than 3000 families, and less than 2000 markers. These studies were not well-powered to identify individual candidate genes. However, they were able to identify genomic regions that might harbor genes related to initial cannabis use and dependence. The one exception was the multistage analysis performed by Han and colleagues [26] which used FLS, GWAS, and independent replication cohorts to nominate variants in *NRG1* as possible genetic risk factor for cannabis and opioid dependence, particularly in AAs. Taken together, these FLS were able to demonstrate that the high heritability for CUDs ascertained from twin studies translated into the detection of large linkage regions possibly harboring gene variants mediating CUD and cannabis use traits. Few of these regions pass stringent genome-wide correction and most were not replicated in separate cohorts. Thus, we cannot exclude the possibility that some of these suggestive loci represent false positives. Of note, some of these suggestive regions overlap among studies. Genetic maps cannot be translated directly into physical maps, but in general 1 cM is roughly equivalent to 1.2 Mb. Approximate linkage regions were determined for studies that provided marker information and/or the cM interval for linkage regions by using the physical marker locations as an anchor on the physical map (human genome assembly GRCh/hg38).

Genetic distances in cM were then converted into physical map distances. Approximate linkage regions are provided in Table 7.2. Although these intervals are a rough estimate, they provide support for possible overlapping linkage regions containing gene variants that may modulate CUDs. This includes a region on Chr 9 from 128 to 144 Mb that is associated with both dependence and withdrawal in two different cohorts [18, 25]. A region on Chr 7 from ~38 to 75 Mb was associated with cannabis withdrawal and craving. However, both traits were collected from the same cohort [25]. It is important to note that comparing overlapping linkage intervals is not a robust comparison method across studies. The appropriate comparison between studies would be a meta-analysis using summary scores associated with linkage between markers and traits in both studies. However, this information was not provided in some of the FLSs. Another possible reason for the general lack of replication across studies could arise from differences in how each cohort was ascertained (*e.g.* adolescent cannabis dependence versus lifetime use or genetic risk for alcohol or nicotine dependence) and these loci may confer risk of CUD or related traits only during specific developmental periods or populations (e.g. *NRG1*).

7.2.3 Genome-Wide Association Studies

Although the number of cannabis use and related GWAS is still relatively small, several large studies have identified genome-wide significant and suggestive loci containing candidate genes (Tables 7.3 and 7.4). These studies provide evidence that CUD is a polygenic disease and that increasing the GWAS sample sizes should increase the number of candidate genes and generate better genetic predictors to evaluate risk, the relationship to other diseases or behavior traits, and the role of environmental factors.

The first GWAS for CD based on DSM-IV criteria was published by Agrawal and colleagues [27]. A panel of 948,142 SNP markers was genotyped in a case-control study design with 708

Table 7.3 Results from genome-wide association studies

Phenotype	Type	Number of individuals	Number of independent genome-wide significant markers	Replication cohort/ significance	Study design	Study cohort	Ethnicity	Study reference
Dependence	GWAS	3054	0	No	Case-control: 708 cases and 2346 controls	Study of Addiction: Genes and Environment (SAGE)	EA and AA	[27]
Initiation of use	metaGWAS	10,091	0	No	Family based: 7175 Australian twins and families and 2916 UK twins and families	Australia and UK twin registries	European Ancestry	[33]
Lifetime use	metaGWAS	32,330	0	Yes/N.S.	Discovery cohort with replication (5627 individuals)	International Cannabis Consortium samples	European (discovery cohorts) and Mixed European and AA (replication cohorts)	[32]
Age at initiation	metaGWAS	24,953	5	Yes/N.S.	Discovery cohort with replication (3735 individuals)	International Cannabis Consortium (ICC)	European Ancestry	[30]
Lifetime use	metaGWAS	184,765	8 SNPS, 35 genes	No	Discovery cohort	ICC; UK Biobank; 23andMe	Primarily Caucasian	[29]
Dependence severity	metaGWAS	14,754	3	Yes/partial	Case-control discovery and replication cohorts	Yale-Penn Study, SAGE, International Consortium on the Genetics of Heroin Dependence	EA and AA	[31]
Dependence	metaGWAS	8515	1	Yes/partial	Case-control discovery and replication cohorts	COGA, SAGE and the Comorbidity and Trauma Study	European Discovery Cohort and AA and EA discovery cohort	[28]
Dependence	metaGWAS	51,372	1	Yes/yes	Case-control discovery and replication cohorts	Lundbeck Foundation Initiative for Integrative Psychiatric Research, deCODE	European discovery and repliocation cohort	[39]

Table 7.4 Results from genome-wide association studies (top marker shown for each region)

Gene	Gene new	SNP/SNP number	Chr	Mb Start	P value	Phenotype	Study reference
UCHL5	UCHL5	rs9427573	1	193,026,083	4.75E-06	Dependence	[27]
Intergenic	LINC01122	rs17552189	2	58,845,834	4.12E-06	Dependence	[27]
KYNU	KYNU	rs2381356	2	142,929,013	1.70E-05	Dependence	[27]
Intergenic	Intergenic	rs12491921	3	106,570,103	1.03E-06	Dependence	[27]
AIM1	CRYBG1	rs1462146	6	106,538,205	1.16E-05	Dependence	[27]
RPS6KA2	RPS6KA2	rs16900973	6	166,699,507	2.48E-05	Dependence	[27]
Intergenic	Intergenic	rs10858373	9	135,235,645	1.68E-05	Dependence	[27]
Intergenic	Intergenic	rs11007350	10	29,000,102	2.68E-06	Dependence	[27]
STAM	STAM	rs1007264	10	17,668,025	2.34E-05	Dependence	[27]
Intergenic	Intergenic	rs2169487	11	38,182,871	1.41E-05	Dependence	[27]
Intergenic	Intergenic	rs2068909	11	94,724,752	1.96E-05	Dependence	[27]
Intergenic	Intergenic	rs1352414	11	37,989,792	2.07E-05	Dependence	[27]
MICAL2	MICAL2	rs1609930	11	12,231,627	2.77E-05	Dependence	[27]
CHST11	CHST11	rs12811699	12	104,623,386	7.88E-06	Dependence	[27]
LGR5	LGR5	rs1280605	12	71,562,287	1.04E-05	Dependence	[27]
DKFZP779	CCDC91	rs7313862	12	28,458,596	2.61E-05	Dependence	[27]
ACADS	ACADS	rs555404	12	120,737,931	2.66E-05	Dependence	[27]
Intergenic	LINC00362	rs9507041	13	23,170,240	8.47E-06	Dependence	[27]
KATNAL1	KATNAL1	rs7317962	13	30,280,075	2.23E-05	Dependence	[27]
Intergenic	Intergenic	rs17102248	14	34,135,613	2.19E-05	Dependence	[27]
FTO	FTO	rs11865530	16	53,920,866	2.18E-05	Dependence	[27]
ANKFN1	ANKFN1	rs1019238	17	56,201,354	6.12E-07	Dependence	[27]
Intergenic	Intergenic	rs1431318	17	56,115,162	9.14E-07	Dependence	[27]
Intergenic	Intergenic	rs10521290	17	14,947,119	2.69E-05	Dependence	[27]
Intergenic	Intergenic	rs8111749	19	34,548,589	2.05E-05	Dependence	[27]
PIM3	Intergenic	rs28372448	22	49,957,323	8.00E-06	Dependence	[27]
RP11-406O16.1	LINC02549	rs12207424	6	68,275,183	6.94E-05	Initiation of use	[33]
GNG5P5	Intergenic	rs1417205	13	47,898,879	1.62E-07	Initiation of use	[33]
NELL1	Intergenic	rs1573535	11	21,599,402	8.83E-06	Initiation of use	[33]
Intergenic	Intergenic	rs9900808	17	13,159,478	8.82E-05	Initiation of use	[33]
KCNT2	KCNT2	rs73067624	1	196,364,081	6.30E-05	Lifetime use	[32]
Intergenic	Intergenic	rs58691539	2	52,526,521	6.40E-06	Lifetime use	[32]
Intergenic	Intergenic	rs2033867	2	174,323,303	4.20E-06	Lifetime use	[32]

SETD2	rs35053471	3	47,083,021	9.20E-05	Lifetime use	[32]
Intergenic	rs12518098	5	61,568,390	4.70E-05	Lifetime use	[32]
NCAM1	rs4471463	11	113,112,623	9.00E-07	Lifetime use	[32]
Intergenic	rs7107977	11	915,514	6.40E-06	Lifetime use	[32]
Intergenic	rs2099149	12	30,326,175	5.10E-07	Lifetime use	[32]
Intergenic	rs4984460	15	95,880,920	2.20E-06	Lifetime use	[32]
CADM2	NA	3	84,959,005	MTC < 0.05	Lifetime use	[32
SCOC-AS1	NA	4	140,283,724	MTC < 0.05	Lifetime use	[32]
SCOC	NA	4	140,257,286	MTC < 0.05	Lifetime use	[32]
ATP2C2	rs1574587	16	84,419,200	4.00E-09	Age at initiation	[30]
Intergenic	rs4935127	10	54,894,976	4.60E-07	Age at initiation	[30]
Intergenic	rs2249437	6	1,594,731	5.10E-07	Age at initiation	[30]
Intergenic	rs9266245	6	31,357,675	1.60E-06	Age at initiation	[30]
Intergenic	rs28622199	8	5,534,331	2.70E-06	Age at initiation	[30]
ABCC1	rs215069	16	15,997,130	3.80E-06	Age at initiation	[30]
SPINT1-AS1	rs4924506	15	40,837,019	5.50E-06	Age at initiation	[30]
ECT2L	rs7773177	6	138,821,701	8.50E-06	Age at initiation	[30]
CADM2	rs2875907	3	85,469,180	1.59E-19	Lifetime use	[29]
SDK1	rs10085617	7	3,594,829	2.93E-08	Lifetime use	[29]
ZNF704	rs9773390	8	80,653,207	5.66E-09	Lifetime use	[29]
NCAM1	rs9919557	11	113,006,436	9.94E-11	Lifetime use	[29]
ATP2A1	rs10499	16	28,903,956	1.13E-09	Lifetime use	[29]
SMG6	rs17761723	17	2,203,546	3.24E-08	Lifetime use	[29]
KLHL21	96	1	6,590,724	7.65E-07	Lifetime use	[29]
PHF13	84	1	6,613,696	1.99E-06	Lifetime use	[29]
LRRTM4	3621	2	76,747,724	1.02E-07	Lifetime use	[29]
MSANTD1	231	4	3,244,369	2.22E-06	Lifetime use	[29]
HTRA1	64	5	122,461,525	2.41E-06	Lifetime use	[29]
BEND6	252	6	56,955,292	2.60E-08	Lifetime use	[29]
KIAA1586	58	6	57,046,532	1.75E-07	Lifetime use	[29]
RAB23	86	6	57,186,992	2.32E-09	Lifetime use	[29]
REV3L	539	6	111,299,052	1.99E-06	Lifetime use	[29]
ARID1B	1344	6	156,777,930	1.15E-08	Lifetime use	[29]
ADGRB1	275	8	142,464,016	1.21E-06	Lifetime use	[29]

(continued)

Table 7.4 (continued)

Gene	Gene new	SNP/SNP number	Chr	Mb Start	P value	Phenotype	Study reference
	NEURL1	17	10	103,493,979	1.83E-07	Lifetime use	[29]
	BORCS7	87	10	102,854,210	1.19E-06	Lifetime use	[29]
	AS3MT	177	10	102,869,453	1.53E-08	Lifetime use	[29]
	CNNM2	549	10	102,918,293	8.02E-07	Lifetime use	[29]
	NT5C2	389	10	103,088,018	7.64E-07	Lifetime use	[29]
	BRAP	97	12	111,642,146	5.48E-07	Lifetime use	[29]
	ACAD10	141	12	111,754,283	9.89E-08	Lifetime use	[29]
	ALDH2	112	12	111,766,933	3.61E-07	Lifetime use	[29]
	MAPKAPK5	195	12	111,842,626	5.58E-07	Lifetime use	[29]
	TMEM116	222	12	111,931,283	3.96E-07	Lifetime use	[29]
	SBK1	23	16	28,292,519	4.52E-08	Lifetime use	[29]
	NPIPB7	10	16	28,456,372	5.46E-08	Lifetime use	[29]
	CLN3	62	16	28,477,294	2.56E-09	Lifetime use	[29]
	APOBR	49	16	28,494,649	7.56E-09	Lifetime use	[29]
	IL27	57	16	28,499,362	7.48E-09	Lifetime use	[29]
	SGF29	181	16	28,553,926	4.87E-09	Lifetime use	[29]
	SULT1A2	25	16	28,591,943	6.66E-08	Lifetime use	[29]
	SULT1A1	51	16	28,605,597	1.14E-07	Lifetime use	[29]
	CDC37P1	31	16	28,700,294	1.42E-07	Lifetime use	[29]
	EIF3C	14	16	28,688,558	8.08E-08	Lifetime use	[29]
	EIF3CL	23	16	28,379,581	4.55E-08	Lifetime use	[29]
	NPIPB9	8	16	28,751,787	6.29E-08	Lifetime use	[29]
	ATXN2L	89	16	28,823,093	2.50E-09	Lifetime use	[29]
	TUFM	55	16	28,842,411	2.83E-09	Lifetime use	[29]
	SH2B1	71	16	28,863,975	5.46E-09	Lifetime use	[29]
	NFATC2IP	8	16	28,950,937	8.82E-08	Lifetime use	[29]
	RABEP2	71	16	28,904,421	2.84E-08	Lifetime use	[29]
	SRR	121	17	2,303,704	5.03E-08	Lifetime use	[29]
	TSR1	90	17	2,322,503	1.12E-08	Lifetime use	[29]
	NPC1	257	18	23,541,437	5.87E-08	Lifetime use	[29]
	LINC00526/C18orf18	132	18	5,236,724	5.65E-08	Lifetime use	[29]
	PKN2-AS1	rs74823926	1	88,263,750	1.40E-05	Dependence severity	[31]
ARHGEF33	ARHGEF33	rs114383460	2	38,938,782	1.09E-06*	Dependence severity	[31]

Gene	Gene new	SNP/SNP number	Chr	Mb Start	P value	Phenotype	Study reference
Intergenic	Intergenic	rs12621150	2	77,801,462	2.91E-04	Dependence severity	[31]
AFF3	AFF3	rs7586604	2	99,834,964	1.06E-06	Dependence severity	[31]
Intergenic	Intergenic	rs144605126	2	103,147,706	8.67E-07*	Dependence severity	[31]
Intergenic	Intergenic	rs150064803	2	117,733,075	3.41E-07*	Dependence severity	[31]
SCN9A	SCN9A	rs143020225	2	166,357,954	7.19E-07*	Dependence severity	[31]
RP11-206 M11.7	Intergenic	rs143244591	3	149,295,898	2.18E-08*	Dependence severity	[31]
PI4K2B	SEPSECS antisense RNA 1	rs73252553	4	25,199,446–25	1.66E-07	Dependence severity	[31]
SEC24D	SEC24D	rs28595532	4	118,795,545	1.13E-06	Dependence severity	[31]
CTNND2	CTNND2	rs114311699	5	11,892,022	3.78E-07*	Dependence severity	[31]
COL23A1	COL23A1	rs10066744	5	178,319,349	4.82E-07*	Dependence severity	[31]
Intergenic	Intergenic	rs17665889	6	51,356,409	2.58E-04	Dependence severity	[31]
Intergenic	Intergenic	rs12534830	7	85,323,065	4.52E-07	Dependence severity	[31]
CSMD1	CSMD1	rs77378271	8	3,215,717	4.60E-08	Dependence severity	[31]
Intergenic	Intergenic	rs10969106	9	29,364,079	7.39E-08#	Dependence severity	[31]
Intergenic	Intergenic	rs115553536	10	31,692,207	6.46E-07*	Dependence severity	[31]
RET	RET	rs74400468	10	43,097,111	6.46E-07*	Dependence severity	[31]
CCAR1	CCAR1	rs12218439	10	68,730,099	1.13E-04	Dependence severity	[31]
SLC35G1	SLC35G1	rs146091982	10	93,899,951	1.95E-07*	Dependence severity	[31]
	Intergenic	rs73443003	11	20,539,214	1.31E-06	Dependence severity	[31]
	Intergenic	rs200453611	11	81,721,913	6.81E-04	Dependence severity	[31]
	Intergenic	rs200391037	11	109,028,447	1.32E-06	Dependence severity	[31]
	Intergenic	rs193047854	12	55,880,121	7.06E-07*	Dependence severity	[31]
	Intergenic	rs199783889	20	21,725,717	3.32E-07*	Dependence severity	[31]
	Intergenic	rs78068107	21	16,646,750	1.02E-06*	Dependence severity	[31]
	Intergenic	rs186825689	21	46,585,890	1.86E-08*	Dependence severity	[31]
Intergenic	Intergenic	rs1409568	10	118,871,023	3.95E-08	Dependence	[28]
ACYP2	ACYP2	rs2287641	2	54,115,568	9.00E-07	Dependence	[28]
CHRNA2	CHRNA2	rs56372821	8	27,459,761	9.31E-12	Dependence	[39]

dependent cases and 2346 non-dependent controls (66% EA and 34% AA). Similar to many of the FLS, the population of cases and controls for which CD was assessed were originally ascertained for alcohol dependence. A caveat of measuring cannabis or other drug-related traits in these populations is highly comorbid polydrug dependence. However, at the time no populations recruited exclusively for cannabis-related traits existed and populations ascertained for alcohol dependence were readily available. As expected based on the small sample size (<10,000 individuals), no markers met genome-wide significance. However, the large number of genotyped markers resulted in suggestive associations for markers tagging individual genes or intergenic regions. These suggestive associations were located on Chrs 1 (*UCHL5*), 2 (*LINC01122*, *KYNU*), 3 (intergenic), 6 (*CRYBG1*, *RPS6KA2*), 9 (intergenic), 10 (*STAM*), 11 (*MICAL2*), 12 (*CHST11*, *LGR5*, *CCDC91*, *ACADS*), 13 (*LINC00362*, *KATNAL1*), 14 (intergenic), 16 (*FTO*), 17 (*ANKFN1*), 19 (intergenic), and 22 (intergenic) (Table 7.4). The first GWAS for CUDs lacked statistical power and did not follow-up their suggestive loci in an independent replication cohort. However, it identified the first putative candidate genes for CD.

Later GWAS [28–33] were able to increase the number of candidate gene associations for CUDs and related traits by increasing the sample size of the discovery cohort. This was achieved primarily by combining results from smaller GWAS studies using meta-analysis (metaGWAS). Individually, each study may be underpowered to detect small effect alleles due to small sample sizes. However, when each study is combined the detection of small effect loci becomes possible due to the increased sample size. In metaGWAS summary statistics (effect size, standard error, and/or p-values) for associations between SNPs and phenotypes from multiple population studies comprised of unique individuals are combined to generate new association scores, effect estimates, and evaluate data set heterogeneity (differences in methodology between studies that could impact results). For a review of the metaGWAS approach, see [34]. As a note of

caution, metaGWAS can increase sample size and power, but inclusion of samples ascertained for substance dependence other than cannabis can introduce heterogeneity and has the potential to confound results or limit reproducibility.

Several groups, starting with Verweij and colleagues [33], were able to increase the number of subjects beyond 10,000 through the use of metaGWAS and by selecting a dichotomous cannabis-related trait (yes or no to cannabis use) that could easily be assessed on a large-scale. However, few markers passed the criterion for genome-wide significance at sample sizes of 20,000 to 30,000 individuals. For example, Verweij and colleagues densely genotyped over two million SNPs from families in Australia and the United Kingdom (10,091 related-individuals) that were part of the Australia and UK twin registries (Spector & Williams 2006). Associations between SNPs and initiation of cannabis use were assessed in the Australian and UK cohort separately using family-based association tests followed by meta-analysis. Suggestive associations were observed for markers on Chrs 6, 13, 11, and 17, but no SNPs reached genome-wide significance (Table 7.4). Likewise, Stringer and colleagues [32] examined 32,330 subjects (European ancestry) for lifetime cannabis use and failed to identify any SNP associations reaching genome-wide significance. This was despite tripling the sample size used by Verweij and colleagues [33] by combining 13 discovery samples collected from around the world (International Cannabis Consortium data sets) and performing meta-analysis. Nevertheless, suggestive associations were identified on Chrs 1, 2, 3, 5, 11, 12, 15 (Table 7.4) and a less stringent gene-based analysis of 24,576 genes/genetic regions identified significant associations for the genes *NCAM1* (neural cell adhesion molecule 1, Chr11), *CADM2* (cell adhesion molecule 2, Chr3), *SCOC-AS1* (short coiled-coil protein anti-sense RNA 1, Chr4), *SCOC* (short coiled-coil protein, Chr4), and *KCNT* (Chr1) following multiple test correction (Table 7.4). The top SNP and gene associations identified in the discovery cohorts failed to replicate in an independent samples consisting of 5627 individuals (53% European and

47% AA), with the exception of suggestive associations for *SCOC-AS1* and *SCOC* in one of the 4 replication samples (AA). SNP heritability based on common SNPs in the Stringer study was estimated at 13–20% for lifetime cannabis use, which was an improvement over the 6% SNP heritability computed for Verweij and colleagues [33]. Both studies included discovery cohorts with different recruitment strategies and subsequent wide variation in the prevalence of lifetime cannabis use among cohorts may have deflated heritability estimates in both studies. Nevertheless, improvements in SNP heritability with larger samples sizes in the Verweij study confirmed that lifetime use of cannabis is a heritable trait contributed to by many loci of small effect. Thus, further increases in sample size should result in identification of more significant loci.

As proof of this concept, Pasman and colleagues [29] published the largest metaGWAS of lifetime cannabis use (184,765 individuals) and identified eight genome-wide significant independent SNPs in six regions (Chrs 3, 7, 8, 11, 16, and 17). Altogether, the identified SNPs accounted for 11% of the individual variance in lifetime use of cannabis. Using gene-based tests they identified 35 genes significantly associated with lifetime cannabis use (Table 7.4). Replication in an independent cohort was not performed, likely because the replication cohort would be much smaller and less well-powered than the discovery cohort. There was also substantial heterogeneity among cohorts used in the meta-analysis that might have limited power in some analyses and/or reproducibility or generalizability. Despite some limitations of the study, Pasman and colleagues were able to identify multiple significant loci and genes for lifetime cannabis use using a massive cohort of nearly 200,000 individuals. This study provides more evidence that the genetic architecture of lifetime cannabis use is complex and involves many small effect genes. Importantly, most of the loci identified in the study were novel and had not been identified previously.

The GWAS discussed thus far took advantage of samples recruited based on different criteria to identify loci associated with lifetime cannabis use. However, there is some debate over how lifetime use is related to development of problematic use and dependence. Early use has been associated with progression to problematic cannabis use and susceptibility for other substance use disorders [1, 35–37]. Early use may also interact with environmental and social factors. For example, the age at which individuals begin to use cannabis may depend on the overall prevalence of use within a country. Higher prevalence has been related to younger ages of initiation [38].To begin to address this issue, Minică and colleagues [30] used GWAS to identify loci associated with early cannabis use. The authors performed metaGWAS on a discovery cohort of 24,953 individuals with replication in a sample of 3735 individuals. This study also estimated heritability for age of initiation at 39% based on three cohorts consisting of 8055 twins (European descent). SNPs in the ATP2C2 gene reached genome-wide significance (Table 7.4). However, they failed to replicate in the independent cohort, and SNP-based heritability for age of initiation was not significant. Note that in both metaGWAS studies with replication examined thus far [30, 32] the replication cohort was much smaller than the discovery cohort which may have limited the power for replication in the discovery cohort.

Only three metaGWAS studies [28, 31, 39] examined CUD directly. In the first study, Sherva and colleagues [31] identified loci associated within CD severity based on DSM-IV criteria using metaGWAS and replication across three independent cohorts consisting of 14,754 individuals (AA and EA). Each cohort was ascertained separately for drug dependence as part of the Yale-Penn Study on the genetics of substance use [40], the SAGE Study on the genetics of alcohol, nicotine, and cocaine use [41], and the International Consortium on the Genetics of Heroin Dependence [42]. SNPs tagging several independent loci met the criteria for genome-wide significance ($P < E^{-7}$) in the AA samples alone or in the combined metaGWAS (Table 7.4). These SNPs were upstream of the gene for S1000 calcium binding protein (*S100B*) and within the gene for CUB and Sushi multiple domains 1

(*CSMD1*). Secondary analysis using a replication cohort found additional support for the dependence severity association score and SNPs in CSMD1 and the drug/metabolite transporter superfamily gene solute carrier family 35 member G1 (SLC35G1). Potential limitations of the study were that CD severity was significantly correlated with dependence for other drugs of abuse (alcohol, nicotine, opioids, and cocaine) and there was high heterogeneity among the sample cohorts included in metaGWAS and used for replication.

The second metaGWAS for CD relied on 8515 individuals of European descent and was drawn from five different cohorts, four of which were ascertained for substance use, including COGA, SAGE, and the Comorbidity and Trauma Study [43]. Agrawal and colleagues [28] analyzed 2080 dependent and 6435 non-dependent cannabis-using controls from this cohort using metaG-WAS. The selection of non-dependent controls (based on DSM-IV criterion) with at least one reported use of cannabis was a unique aspect of the study. SNPs on Chr 10 were identified as genome-wide significant and there was modest evidence for replication of this association in AA (but not EA) individuals in a small independent replication sample (896 AA cases and 1591 controls). These SNPs were not associated with genes but the authors provided some evidence that one SNP in the Chr 10 region (rs1409568) may be located within an active enhancer. A suggestive association between dependence severity (cannabis dependence symptoms counts based on DSM-IV criteria) and SNPs on Chr 2 around marker rs2287641 was also identified but did not replicate in the independent cohort.

Finally, Demontis and colleagues [39] identified an association between SNPs located in a cluster on Chr 8 (rs56372821 index SNP) and CD using a data set consisting of 2387 dependent cases and 48,985 controls. The cohort used in this analysis differed from most of the previous studies in that it was ascertained for major mental illness (schizophrenia, bipolar disorder, attention deficit hyperactivity disorder, anorexia nervosa and autism spectrum disorder) and not drug use or dependence. All individuals were part of the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) Danish nation-wide cohort [44]. Significant replication was observed in a replication sample of 5501 cases and 301,041 controls. Of note, expression of the neuronal acetylcholine receptor (nAChR) alpha-2 subunit gene, *CHRNA2*, in human cerebellum was found to be controlled by the variants at the Chr 8 locus using the Genotype-Tissue Expression (GTEx) dataset [45]. Taken together these results provide a possible causal mechanism by which variants at the Chr8 locus regulate *CHRNA2* brain expression and possible risk of CD.

7.3 Limitations and Future Directions

Increasing the sample size of GWAS for cannabinoid use has dramatically increased the number of markers that pass the criteria for genome-wide significance. This was clearly demonstrated by Pasman and colleagues [29] following identification of 8 independent genome-wide significant markers on Chrs 3, 7, 8, 11, 16, and 17 for lifetime cannabis use in a cohort of 184,765 individuals. Another large case-control study of 51,372 individuals identified one genome-wide significant loci on Chr 8 that replicated in an equally sized replication cohort [39]. However, these successes are modest compared with recent large GWAS studies for other diseases. For example, over 100 risk loci for Schizophrenia have been identified in GWAS combining ~50,000 individuals [6].

There are several possible reasons for the paucity of strong candidates in human association studies of cannabis use and dependence. The first reason is that the genetic architecture of cannabis use and dependence may be difference than that of other substance use disorders and psychiatric diseases. Most GWAS for substance use disorders and psychiatric diseases demonstrate a large genomic target associated with disease risk. In other words, many biological pathways and genes contribute a small amount to overall risk. Thus, many genes of small effect combine in an

additive fashion to influence disease risk. This type of genetic architecture typically results in a positive linear relationship between sample size and identification of genome-wide significant associations. However, it is possible that the genetic architecture of cannabis use and dependence is different and that rare variants and non-additive interactions or environmental interactions drive the disease more than the combined actions of many small effect loci. However, this hypothesis seems at odds with recent large GWAS for cannabis use and dependence [29, 39] and for GWAS results for other drugs of abuse, which likely share some common underlying biological and genetic mechanisms.

Perhaps a more likely explanation for the small number of high-confidence candidates is the type of data sets used in recent GWAS. For cannabis research, there is a lack of large population-based data sets for which individuals were ascertained primarily for cannabinoid or cannabis-related traits. Combining data sets ascertained for psychiatric disease or dependence for other drugs of abuse can lead to heterogeneity among samples that can actually decrease power. This may explain why so few loci were identified as genome-wide significant level for association with lifetime cannabis use in a massive data set of nearly 200,000 individuals [29].

Yet another issue with association studies over the past two decades is the lack of replication among data sets. Only a single linkage region overlapped among FLS (Table 7.2). Of all the suggestive and significant associations identified in GWAS (Table 7.4), only two genes (*CADM2* and *NCAM1*) were identified by different studies [29, 32]. However, it is important to note that both studies that identified *NCAM1* and *CADM2* included the same set of ~30,000 individuals from the International Cannabis Consortium.

Association studies in humans have the capability to identify genes and risk alleles. As the sample sizes for GWAS studies increase, so does the number of associations. If these alleles can be identified directly in humans, why the need for testing in preclinical animal models? The answer depends on biological systems and causality. Preclinical genetic animal models (specifically rodents) offer the ability to directly test the role of genes in CUDs and explore the underlying biology in ways that would be impossible in humans. Environmental factors can also be controlled and manipulated in preclinical studies in ways that are not possible in studies involving human subjects. Bi-directional translation between association studies in humans and preclinical models is essential for identifying the environmental, genetic, and molecular mechanisms contributing to disease and for design of effective therapeutics.

One of the simplest ways in which genetic preclinical models support association studies is through reverse genetic engineering. In this case, candidate genes are manipulated in the preclinical model to evaluate their role in disease. It is even possible to introduce the precise human genetic variant into a preclinical model to evaluate its impact. Such humanized mice have been used to evaluate the role of common functional variants in the catechol-O-methyltyransferase gene [46] and to model the role of alleles involved in risk of familial Alzheimer's disease [47]. Thus, genetic engineering approaches in preclinical rodent models can be used to directly evaluate the role of candidate genes evaluated in human association studies. However, relatively few genes have been evaluated for a role in cannabis or cannabinoid-related traits in rodent models. The only gene identified from human association studies that has also been independently evaluated for cannabinoid-related traits is the *NRG1* gene. Of interest, mice heterozygous for deletion of murine *Nrg1* show enhanced sensitivity to the main psychoactive cannabinoid in cannabis, Δ9-tetrahydrocannabinol or THC [48–51]. Recent advances in genetic engineering, including CRISPR/Cas9 mediated genetic engineering [52] should facilitate functional evaluation of genes and gene variants such as *CADM2* and the schizophrenia susceptibility gene, *NCAM1*, associated with cannabis use, dependence, and/or withdrawal in humans.

Preclinical genetic models can also be used for unbiased genome-wide linkage or association studies to identify genes and gene variants that contribute to disease variation. Examples of these

include rodent genetic models in which two or more inbred progenitor strains are crossed repeatedly and then inbred (recombinant inbred lines such as the BXD panel or collaborative cross panel in mice) or outcrossed repeatedly (diversity outcross and heterogenous stock mice) to create genetic panels segregating millions of variants [53–56]. For a review see [57]. As of 2018, traits related to CUD in humans (initial sensitivity, tolerance and dependence, withdrawal severity, and/ or self-administration) have not been profiled in genetic rodent populations in order to identify candidate genes. A single study attempted a short-term selection in an F2 cross between C57BL/6J and DBA/2J inbred strains in order to determine if locomotor sensitivity to THC was heritable and could be selected for in order to produce progeny that carry sensitive or resistant alleles for later genetic dissection [58]. As in all systems, preclinical models have some advantages and disadvantages. The clear advantage is the ability to manipulate all aspects of preclinical studies and derive causality from these controlled manipulations. The main disadvantage is that preclinical models are not identical to humans at all levels of behavior and physiology and, as a result, there will always be some controversy regarding translatability.

7.4   Conclusions

Association studies for cannabis use and dependence over the past two decades have identified candidate linkage regions (FLS, Table 7.2) and genes (primarily through GWAS, Table 7.4). In contrast, CGAS have yielded inconsistent results. Over the next two decades, it is likely that more GWAS containing 50,000 to 1000,000 individuals will be performed for cannabis use, dependence, and withdrawal. Recruiting samples directly for these traits along with other methods to reduce heterogeneity among cohorts can be expected to increase the number of genome-wide significant associations. This should lead to a larger and more reproducible list of candidates and a better assessment of polygenic risk and genetic architecture. It is also important to

remember that, despite some of the current issues with power and reproducibility, human association studies have identified candidate genes and mechanisms that should be evaluated to determine how and how much they contribute to disease risk. The stage is already set for this type of translational research in preclinical animal models.

References

1. Agrawal A, Lynskey MT (2006) The genetic epidemiology of cannabis use, abuse and dependence. Addiction 101(6):801–812
2. Kendler KS, Prescott CA (1998) Cannabis use, abuse, and dependence in a population-based sample of female twins. Am J Psychiatry 155(8):1016–1022
3. American Psychiatric Association (2013) American Psychiatric Association. DSM-5 Task Force. Diagnostic and statistical manual of mental disorders: DSM-5. 5th edn. American Psychiatric Association, Washington, DC, xliv, 947 p
4. (UNODC) UNOoDaC (2010) World drug report 2010. United Nations, New York
5. Hasin DS, Kerridge BT, Saha TD, Huang B, Pickering R, Smith SM et al (2016) 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 173(6):588–599
6. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA et al (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101(1):5–22
7. Agrawal A, Wetherill L, Dick DM, Xuei X, Hinrichs A, Hesselbrock V et al (2009) Evidence for association between polymorphisms in the cannabinoid receptor 1 (CNR1) gene and cannabis dependence. Am J Med Genet B Neuropsychiatr Genet 150B(5):736–740
8. Hopfer CJ, Young SE, Purcell S, Crowley TJ, Stallings MC, Corley RP et al (2006) Cannabis receptor haplotype associated with fewer cannabis dependence symptoms in adolescents. Am J Med Genet B Neuropsychiatr Genet 141B(8):895–901
9. Chiang KP, Gerber AL, Sipe JC, Cravatt BF (2004) Reduced cellular expression and activity of the P129T mutant of human fatty acid amide hydrolase: evidence for a link between defects in the endocannabinoid system and problem drug use. Hum Mol Genet 13(18):2113–2119
10. Sipe JC, Chiang K, Gerber AL, Beutler E, Cravatt BF (2002) A missense mutation in human fatty acid amide hydrolase associated with problem drug use. Proc Natl Acad Sci U S A 99(12):8394–8399
11. Buhler KM, Gine E, Echeverry-Alzate V, Calleja-Conde J, de Fonseca FR, Lopez-Moreno JA (2015)

Common single nucleotide variants underlying drug addiction: more than a decade of research. Addict Biol 20(5):845–871

12. Tyndale RF, Payne JI, Gerber AL, Sipe JC (2007) The fatty acid amide hydrolase C385A (P129T) missense variant in cannabis users: studies of drug use and dependence in Caucasians. Am J Med Genet B Neuropsychiatr Genet 144B(5):660–666

13. Haughey HM, Marshall E, Schacht JP, Louis A, Hutchison KE (2008) Marijuana withdrawal and craving: influence of the cannabinoid receptor 1 (CNR1) and fatty acid amide hydrolase (FAAH) genes. Addiction 103(10):1678–1686

14. Schacht JP, Selling RE, Hutchison KE (2009) Intermediate cannabis dependence phenotypes and the FAAH C385A variant: an exploratory analysis. Psychopharmacology 203(3):511–517

15. Verweij KJ, Zietsch BP, Liu JZ, Medland SE, Lynskey MT, Madden PA et al (2012) No association of candidate genes with cannabis use in a large sample of Australian twin families. Addict Biol 17(3):687–690

16. Tawa EA, Hall SD, Lohoff FW (2016) Overview of the genetics of alcohol use disorder. Alcohol Alcohol 51(5):507–514

17. Patton AL, Seely KA, Yarbrough AL, Fantegrossi W, James LP, McCain KR et al (2018) Altered metabolism of synthetic cannabinoid JWH-018 by human cytochrome P450 2C9 and variants. Biochem Biophys Res Commun 498(3):597–602

18. Hopfer CJ, Lessem JM, Hartman CA, Stallings MC, Cherny SS, Corley RP et al (2007) A genome-wide scan for loci influencing adolescent cannabis dependence symptoms: evidence for linkage on chromosomes 3 and 9. Drug Alcohol Depend 89(1):34–41

19. Stallings MC, Corley RP, Hewitt JK, Krauter KS, Lessem JM, Mikulich SK et al (2003) A genome-wide search for quantitative trait loci influencing substance dependence vulnerability in adolescence. Drug Alcohol Depend 70(3):295–307

20. Agrawal A, Hinrichs AL, Dunn G, Bertelsen S, Dick DM, Saccone SF et al (2008) Linkage scan for quantitative traits identifies new regions of interest for substance dependence in the collaborative study on the genetics of alcoholism (COGA) sample. Drug Alcohol Depend 93(1–2):12–20

21. Schuckit MA, Hesselbrock V, Tipp J, Anthenelli R, Bucholz K, Radziminski S (1994) A comparison of DSM-III-R, DSM-IV and ICD-10 substance use disorders diagnoses in 1922 men and women subjects in the COGA study. Collaborative Study on the Genetics of Alcoholism. Addiction 89(12):1629–1638

22. Agrawal A, Pergadia ML, Saccone SF, Lynskey MT, Wang JC, Martin NG et al (2008) An autosomal linkage scan for cannabis use disorders in the nicotine addiction genetics project. Arch Gen Psychiatry 65(6):713–721

23. Saccone SF, Pergadia ML, Loukola A, Broms U, Montgomery GW, Wang JC et al (2007) Genetic linkage to chromosome 22q12 for a heavy-smoking quan-

titative trait in two independent samples. Am J Hum Genet 80(5):856–866

24. Agrawal A, Morley KI, Hansell NK, Pergadia ML, Montgomery GW, Statham DJ et al (2008) Autosomal linkage analysis for cannabis use behaviors in Australian adults. Drug Alcohol Depend 98(3):185–190

25. Ehlers CL, Gizer IR, Vieten C, Wilhelmsen KC (2010) Linkage analyses of cannabis dependence, craving, and withdrawal in the San Francisco family study. Am J Med Genet B Neuropsychiatr Genet 153B(3):802–811

26. Han S, Yang BZ, Kranzler HR, Oslin D, Anton R, Farrer LA et al (2012) Linkage analysis followed by association show NRG1 associated with cannabis dependence in African Americans. Biol Psychiatry 72(8):637–644

27. Agrawal A, Lynskey MT, Hinrichs A, Grucza R, Saccone SF, Krueger R et al (2011) A genome-wide association study of DSM-IV cannabis dependence. Addict Biol 16(3):514–518

28. Agrawal A, Chou YL, Carey CE, Baranger DAA, Zhang B, Sherva R et al (2018) Genome-wide association study identifies a novel locus for cannabis dependence. Mol Psychiatry 23(5):1293–1302

29. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL et al (2018) GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci 21(9):1161–1170

30. Minica CC, Verweij KJH, van der Most PJ, Mbarek H, Bernard M, van Eijk KR et al (2018) Genome-wide association meta-analysis of age at first cannabis use. Addiction 113(11):2073–2086

31. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A et al (2016) Genome-wide association study of Cannabis dependence severity, novel risk variants, and shared genetic risks. JAMA Psychiat 73(5):472–480

32. Stringer S, Minica CC, Verweij KJ, Mbarek H, Bernard M, Derringer J et al (2016) Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Transl Psychiatry 6:e769

33. Verweij KJ, Vinkhuyzen AA, Benyamin B, Lynskey MT, Quaye L, Agrawal A et al (2013) The genetic aetiology of cannabis use initiation: a meta-analysis of genome-wide association studies and a SNP-based heritability estimation. Addict Biol 18(5):846–850

34. Evangelou E, Ioannidis JP (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14(6):379–389

35. King KM, Chassin L (2007) A prospective study of the effects of age of initiation of alcohol and drug use on young adult substance dependence. J Stud Alcohol Drugs 68(2):256–265

36. Grant JD, Lynskey MT, Scherrer JF, Agrawal A, Heath AC, Bucholz KK (2010) A cotwin-control analysis of

drug use and abuse/dependence risk associated with early-onset cannabis use. Addict Behav 35(1):35–41

37. Chen CY, Storr CL, Anthony JC (2009) Early-onset drug use and risk for drug dependence problems. Addict Behav 34(3):319–322

38. Kokkevi A, Nic Gabhainn S, Spyropoulou M (2006) Risk Behaviour Focus Group of the H. Early initiation of cannabis use: a cross-national European perspective. J Adolesc Health 39(5):712–719

39. Demontis D, Rajagopal VM, Als TD, Grove J, Pallesen J, Hjorthøj C et al (2018) Genome-wide association study implicates CHRNA2 in cannabis use disorder. Preprint at bioRxiv

40. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H et al (2014) Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. Biol Psychiatry 76(1):66–74

41. Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB (2008) Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. Drug Alcohol Depend 95(1–2):14–22

42. Maloney E, Degenhardt L, Darke S, Mattick RP, Nelson E (2007) Suicidal behaviour and associated risk factors among opioid-dependent individuals: a case-control study. Addiction 102(12):1933–1941

43. Kristjansson S, McCutcheon VV, Agrawal A, Lynskey MT, Conroy E, Statham DJ et al (2016) The variance shared across forms of childhood trauma is strongly associated with liability for psychiatric and substance use disorders. Brain Behav 6(2):e00432

44. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Baekvad-Hansen M et al (2018) The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry 23(1):6–14

45. Consortium GT (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45(6):580–585

46. Risbrough V, Ji B, Hauger R, Zhou X (2014) Generation and characterization of humanized mice carrying COMT158 Met/Val alleles. Neuropsychopharmacology 39(8):1823–1832

47. Oakley H, Cole SL, Logan S, Maus E, Shao P, Craft J et al (2006) Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. J Neurosci 26(40):10129–10140

48. Long LE, Chesworth R, Huang XF, McGregor IS, Arnold JC, Karl T (2013) Transmembrane domain Nrg1 mutant mice show altered susceptibility to the neurobehavioural actions of repeated THC exposure in adolescence. Int J Neuropsychopharmacol 16(1):163–175

49. Boucher AA, Hunt GE, Micheau J, Huang X, McGregor IS, Karl T et al (2011) The schizophrenia susceptibility gene neuregulin 1 modulates tolerance to the effects of cannabinoids. Int J Neuropsychopharmacol 14(5):631–643

50. Boucher AA, Hunt GE, Karl T, Micheau J, McGregor IS, Arnold JC (2007) Heterozygous neuregulin 1 mice display greater baseline and Delta(9)-tetrahydrocannabinol-induced c-Fos expression. Neuroscience 149(4):861–870

51. Boucher AA, Arnold JC, Duffy L, Schofield PR, Micheau J, Karl T (2007) Heterozygous neuregulin 1 mice are more sensitive to the behavioural effects of Delta9-tetrahydrocannabinol. Psychopharmacology 192(3):325–336

52. Yang H, Wang H, Jaenisch R (2014) Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat Protoc 9(8):1956–1968

53. Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ et al (2016) Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nat Genet 48(8):919–926

54. Chesler EJ (2014) Out of the bottleneck: the diversity outcross and collaborative cross mouse populations in behavioral genetics research. Mamm Genome 25(1–2):3–11

55. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J et al (2004) The collaborative cross, a community resource for the genetic analysis of complex traits. Nat Genet 36(11):1133–1137

56. Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet 5:7

57. Flint J, Valdar W, Shifman S, Mott R (2005) Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet 6(4):271–286

58. Kasten CR, Zhang Y, Mackie K, Boehm SL, 2nd. Short-term genetic selection for adolescent locomotor sensitivity to Delta9-tetrahydrocannabinol (THC). Behav Genet 2018;48(3):224–235