

Genetic Association Studies of Alcohol Dependence

Gregory John Lydall

This thesis is submitted for the degree of Doctor of Medicine
(MD) at

University College London

2015

Statement of Conjoint Work

My work was very much part of a team effort. My specific contributions were:

1. UCL alcohol dependence (AD) sample collection recruitment, diagnostic interviewing, blood sampling and data management of 100 alcohol dependent volunteers and 50 control volunteers.
2. UK-wide coordination of AD recruitment through the UK-COGA and the UK Clinical Research Network. This included recruiter training for over 20 NHS collection sites, advice, support and coordination nationwide, and initial UCL alcohol dependence database management, and UK-COGA website development. This effort resulted in over 2000 AD samples at time of writing.
3. DNA preparation - extraction, quantification and plating out for analysis.
4. Marker selection - from literature and using bioinformatic tools.
5. Literature search to build a database of gene and SNP associations with AD and related phenotypes.
6. Statistical analysis of data.
7. Lead author on two linked publications and co-author on a further 7 genetics publications (see Appendix 18.3).

AUTHOR NOTE

Sources of support: This work was funded by the UK Medical Research Council, The Brain Damage Research Trust, the Neuroscience Research Charitable Trust, the Stanley Foundation and the Stanley Psychiatric Research Center at the Broad Institute. Genotyping was carried out at the Broad Institute.

During my training I was employed by Barnet Enfield and Haringey NHS Trust, and the Camden and Islington NHS Foundation Trust who supported my research sessions.

ABSTRACT

Objectives: Alcohol dependence (AD) commonly co-occurs with bipolar disorder (BP) and schizophrenia (SZ). Together these heritable (and cross-heritable disorders) account for significant morbidity and mortality. The genetic comorbidity between AD, BP, and SZ was investigated to find genetic factors affecting risk for developing AD.

Methods: Subjects were from the UCL AD (n=586), BP, and SZ samples and supernormal controls (SNC; up to 603 subjects). The BP and SCZ subjects were further categorized as having comorbid AD (BPALC n=143; SZALC n=77) or without AD comorbidity (BPnonALC n=367; SZnonALC n=384). The following single gene association study or Genome-Wide Association Studies (GWAS) of the above phenotypes were performed:

- i) Seven SNPs previously associated with AD in the GABRA2 gene were genotyped in the AD cases and SNCs.
- ii) GWAS of BPALC vs SNC; and BPALC vs BPnonALC.
- iii) GWAS of SZALC vs SNC; and SZALC vs SZnonALC.
- iv) Meta-analysis of BPALC and SZALC GWAS data.

Single marker tests, and gene-based permutation tests on all SNPs within a 50kb region flanking each gene, were performed. Genes previously implicated with AD and related phenotypes were tested for association in the datasets. Pathways analysis were performed on all genes uncorrected $p < 0.01$.

Results: i) None of the GABRA2 SNPs showed association with AD. ii-iv) No GWAS SNP met the genome-wide significance threshold. Several gene wide tests with CNS genes suggested replication of prior AD findings. AD implicated pathways included neuronal generation; inflammation; reaction to inorganic substance; transportation; and tyrosine metabolism.

Conclusions: No significant single marker associations were detected in the AD sample or comorbid AD groups. Potential candidate genes were suggestively implicated by gene-based analysis and literature replication. Possible mechanisms for AD susceptibility genes in affective/psychotic disorders are discussed. This exploratory study was underpowered to detect genome wide significance and larger studies are needed.

TABLE OF CONTENTS

AUTHOR NOTE	3
ABSTRACT	4
DEDICATION	10
ACKNOWLEDGEMENTS	11
1 CHAPTER 1: INTRODUCTION	12
2 CHAPTER 2: WHAT IS ALCOHOL DEPENDENCE ?	16
2.1 Historical perspectives.....	16
2.2 Physiology of alcohol	18
2.3 Effects of alcohol on the body.....	19
2.4 Diagnostic criteria and classification	21
2.5 Epidemiology- why are alcohol use disorders important?.....	25
2.6 AD as a complex psychiatric disorder	27
2.7 Molecular genetic studies	28
2.7.1 Challenges of genetic studies in general	28
2.7.2 Alcoholism-related endophenotypes.....	30
2.7.3 Low response to alcohol	32
2.7.4 MaxDrinks	32
2.7.5 Electrophysiological phenotypes	33
2.7.6 Alcohol expectancies.....	33
2.7.7 Craving.....	33
2.7.8 Personality/temperamental factors.....	34
2.7.9 Psychiatric comorbidity.....	38
2.8 Conclusions.....	38
3 CHAPTER 3: HERITABILITY - IS ALCOHOLISM A “GENETIC” DISORDER?	39
3.1 What is heritability?	39
3.2 Animal studies	40
3.3 Family studies	41
3.4 Twin studies.....	41
3.5 Adoption studies.....	43
3.6 Heritability of co-occurring psychiatric disorders	45
3.7 Genes and environment.....	46
3.8 A note on epigenetics	48
4 CHAPTER 4: MAPPING AD & ALCOHOLISM: LINKAGE STUDIES.....	50

4.1	GENETIC MARKERS	50
4.2	Linkage studies	51
4.2.1	Advantages and disadvantages of linkage studies.....	51
4.3	Linkage study findings in AD.....	52
5	CHAPTER 5: MAPPING ALCOHOLISM: ASSOCIATION STUDIES	56
5.1	Approach	56
5.2	Candidate gene study findings in AD.....	57
5.2.1	The GABA system	58
5.2.2	Alcohol metabolism genes	61
5.2.3	The Dopamine system and COMT.....	63
5.2.4	The Serotonin system.....	65
5.2.5	The opioid system	65
5.2.6	Taste receptors.....	67
5.2.7	The endocannabinoid system	68
5.2.8	Miscellaneous and meta-analyses	68
5.2.9	Summary	69
5.3	Genome wide association studies	70
5.3.1	GWAS methodology	70
5.3.2	Limitations of GWAS	73
5.3.3	GWAS findings in AD and related alcohol use disorder phenotypes	75
5.3.4	CNVs in AD.....	78
5.4	Summary.....	80
6	CHAPTER 6: BIPOLAR DISORDER.....	81
6.1	Clinical features	81
6.2	Family, adoption and twin studies	82
6.3	Linkage Studies	82
6.4	Association Studies.....	83
6.4.1	Candidate genes.....	83
6.4.2	GWAS	87
6.4.3	CNVs in BP	90
6.5	Sequencing	91
6.6	BP co-morbidity with alcoholism.....	92
6.6.1	Heritability.....	92
6.6.2	Clinical implications of co-morbidity.....	94

6.6.3	Dopamine in bipolar and alcoholism	95
7	CHAPTER 7: SCHIZOPHRENIA	97
7.1	Definitions	97
7.2	Family, Adoption and twin studies	97
7.3	Molecular genetic studies	97
7.3.1	Linkage studies	97
7.3.2	Association studies.....	99
7.3.3	GWAS of SZ and overlap with BP	100
7.3.4	The Major Histocompatibility Complex (MHC) Locus	104
7.3.5	Rare CNVs and Schizophrenia	105
7.4	Co-morbidity with BP and AD	106
7.5	Summary and conclusions	107
8	CHAPTER 8: STUDY AIMS	109
9	CHAPTER 9: MATERIALS AND METHODS.....	110
9.1	Ethical approvals.....	110
9.2	Power considerations	110
9.3	UCL case-control samples.....	111
9.3.1	Ancestry.....	111
9.3.2	UCL alcohol dependence research sample	112
9.3.3	UCL bipolar research sample.....	113
9.3.4	UCL schizophrenia research sample.....	114
9.4	Genotyping	115
9.4.1	AD sample	115
9.4.2	BP sample	116
9.4.3	SZ sample	116
9.5	Data analysis and quality control (QC)	116
9.5.1	Population Stratification and final dataset	117
9.6	Association Analyses.....	118
9.6.1	Primary Whole Genome Association Analysis	118
9.6.2	Testing of allelic association.....	118
9.7	Use of control groups	118
9.8	Gene-based testing.....	120
9.9	Correction for multiple testing	121
9.10	Pathway-based analysis with VEGAS2 and PANTHER	122

9.11	The literature as control	123
9.12	Graphical techniques	124
10	CHAPTER 10. RESULTS OF GABRA2 CANDIDATE GENE ANALYSIS IN AD	125
10.1	GABRA2 SNP study discussion	125
11	CHAPTER 11. RESULTS OF BPALC VS BPNONALC GWAS	129
11.1	Discussion	131
12	CHAPTER 12. RESULTS OF BPALC VS SNC GWAS.....	149
12.1	Discussion	152
12.2	BPALC vs SNC GWAS conclusions	158
13	CHAPTER 13: RESULTS OF SZALC VS SZNONALC	173
13.1	Discussion	175
14	CHAPTER 14: RESULTS OF SZALC VS NORMAL CONTROLS	187
14.1	Discussion	189
14.2	SZALC GWAS conclusions	190
15	CHAPTER 15: RESULTS OF META-ANALYSIS: BPALC + SZALC VS BPNONALC + SZNONALC.....	201
15.1	Discussion	203
16	CHAPTER 16: RESULTS OF META-ANALYSIS (COMORBIDITY VS SUPERNORMAL CONTROLS): BPALC+SZALC VS SNC	218
16.1	Discussion	222
17	CHAPTER 17. DISCUSSION AND CONCLUSIONS.....	242
17.1	General comments on these GWAS	242
17.2	Gene-based analyses	245
17.3	Pathway analyses	247
17.4	“Missing heritability”	249
17.5	Future directions	253
17.6	Final remarks	255
	REFERENCES	256
18	APPENDICES.....	281
18.1	List of tables.....	281
18.2	List of figures	282
19.3	List of published papers (where I am first author in bold):	284

DEDICATION

To the late Prof Hugh Gurling whose passionate genius for the subject and care for the mentally ill inspired me into the field of psychiatric genetics.

To my long-suffering wife, Petra, and my three children, Peter, Philip and Joseph who were born during this part time MD, and my garden: all have been neglected during this work.

My parents who contributed (apparently) low-risk genes and a loving environment to their offspring, despite their innate temperaments.

ACKNOWLEDGEMENTS

I should like to thank the following people for their advice, patience, and support with this project:

Professor Hugh Gurling

Professor David Curtis

Dr Sally Sharp

Dr Marsha Morgan

Dr Irene Guerrini

Dr Alan Thompson

Dr Petra Du Plessis

I should give a special mention to Dr. Andrew McQuillin for his support and unflappable niceness, and for so ably leading the Department. I am hugely grateful to the NHS Trusts who employed me during my research; the NHS Consultants who supported me; the staff including the UK Clinical Research Network who helped with recruitment; and all the patients who so kindly volunteered for this study.

CHAPTER 1: INTRODUCTION

“When I had my first drink, I was thirteen... something just clicked in my brain... it just felt right”

- UCL AD sample, male subject (45)

Humans have a long and complicated relationship with alcohol.

While the majority of humans drink alcohol within lower risk limits, a minority develop problems with ethanol use. These harms can be in the domains of physical health, mental health, social, legal or economic. Ongoing problematic alcohol use has attracted various terms including “alcohol abuse”, “harmful use” and “alcohol misuse”. When over time physical or psychological dependence develops, this is termed “alcohol dependence” (AD). AD is often used interchangeably with the term “alcoholism” in the literature, and I am using the term AD in this thesis as the main phenotype of interest. Where another term is used, it is because the linked literature has used that term. Alcohol use disorders (AUD) under the most modern diagnostic system refers to a quantitative spectrum of severity including harmful use and dependence.

AD is a complex disorder affected by genetic and environmental factors (Schuckit 2014). In this thesis I hope to contribute to the understanding of the genetic risk factors for developing alcoholism both as a standalone disorder, and also in the context of two commonly co-occurring major psychiatric disorders, i.e. bipolar affective disorder (BP), and schizophrenia (SZ). By having either of these severe mental illnesses, a person’s risk of developing alcoholism (or indeed any addictive disorder) is significantly

increased. Having a comorbid disorder (“dual diagnosis”) is common in mental health services, and appears to affect clinical features, response to treatment, and outcomes. Understanding more fully the risk factors for developing alcohol dependence in the context of two severe and enduring mental illnesses may better enable researchers and clinicians to diagnose and treat these disorders.

In this thesis I will review the epidemiology and global impact of alcoholism to justify why this is a public health problem worth addressing. I will then explore the rationale for studying the genetics of alcoholism by summarizing a) the evidence for a genetic contribution to alcoholism, and b) the current literature on alcoholism genetics. I will also discuss issues around phenotype. I will give a brief overview of bipolar disorder and schizophrenia from a clinical and genetic perspective, and where they overlap clinically and genetically both with each other, and with alcoholism.

I will approach the new research component from three angles using an association study method from the University College London (UCL) psychiatric genetics samples. This collection comprises subjects with one or more well-defined diagnoses: alcohol dependence/alcoholism, bipolar disorder, and/or schizophrenia.

I will firstly use a candidate gene approach on the UCL Alcohol Dependence (AD) sample, which I collected in part. The GABA beta receptor 2 gene (GABRB2) has previously been implicated in alcoholism aetiology, biology and pharmacology. Involvement of the GABA cluster in general, and the GABRB2 gene, makes plausible biological sense for contributing to alcoholism genetic risk, yet GABRB2 has a number of seemingly contradictory positive and negative published studies of varying strengths.

Secondly, I will perform a series of Genome Wide Association Studies (GWAS) on subjects with either BP or SZ and comorbid alcoholism diagnoses, and then a meta-analysis. In each case, the controls will be firstly unaffected individuals (normal controls); and a second analysis will be done by using the remaining BP/SZ individuals who have no diagnosed AD comorbidity. In this way, I will look at the genetics of comorbid disorders vs normal controls; and vs the single disorders to attempt to tease apart different genetic effects.

Specifically I will compare the genetic signals of patients with comorbid bipolar alcoholism (BPALC) with:

- patients with non-alcoholic bipolar disorder (BPnonALC), and
- supernormal controls.

I will also examine the genetics of patients with schizophrenia and comorbid alcoholism (SZALC) vs:

- patients with schizophrenia without alcoholism (SZnonALC), and
- supernormal controls.

To increase power, I will perform a meta-analysis of (BPALC + SZALC) vs:

- (BPnonALC + SZnonALC), and
- supernormal controls.

Then, for each GWAS dataset, I will perform a series of gene-wise tests in order to evaluate significance at the level of the gene. Significant findings will be compared with the literature as a replication analysis. Then I will perform a pathway analysis of the whole genome data.

Finally I will discuss the findings with respect to the literature, the implications, strengths and limitations of these approaches, and aims for future work.

During this piece of work I have been fortunate to publish the first GWAS approach using samples of bipolar alcoholics, and bipolar non-alcoholics, compared with normal controls (screened for major mental illness and substance misuse in the subjects and their first degree relatives) (Lydall et al. 2011a).

I have also published a negative replication study on GABRA2, a candidate gene for alcoholism (Lydall et al. 2011b).

CHAPTER 2: WHAT IS ALCOHOL DEPENDENCE?

“But the physical influence of custom, confirmed into habit interwoven with the actions of our sentient system, and reacting on our mental part have been entirely forgotten. The perfect knowledge of those ... causes which first induced the propensity to vinous liquors, whether they sprung from situation in life, or depended on any peculiar temperament of body, is also necessary for conducting the cure”

- Trotter (1813), “On Drunkenness”

Alcohol is a commonly used psychoactive and addictive drug. Along with caffeine and nicotine, it has a high population prevalence of use (Fitt et al. 2013; HSCIC 2015). Although alcohol is freely available in most countries, problematic alcohol use appears only to affect a minority of the population. Not all people who drink alcohol do so excessively, and not all of those who enjoy it in excess develop harmful use of alcohol; a further minority develop alcohol dependence (AD).

2.1 HISTORICAL PERSPECTIVES

Early hominids are thought to have evolved alcohol dehydrogenase class IV (ADH4), the first enzyme exposed to ethanol in the digestive tract that is capable of metabolizing ethanol, around the time that they began using the forest floor about 10 million years ago. The ADH4 enzyme however in our more ancient and arboreal ancestors did not efficiently oxidize ethanol according to the growing field of palaeogenetics (Carrigan et al. 2015).

The production of an ethanol beverage requires only a very simple organic process – that of fermentation, by which sugar is transformed by the action of yeast to produce ethyl alcohol. Homo sapiens have been fermenting produce containing starch into various forms of ethanol, and imbibing this for thousands of years at least. The right conditions for harnessing the fermentation process are thought to have first existed only among the early Bronze Age cultures of the Eastern Mediterranean and

Mesopotamia between the eighth and fourth millennia BC. On the basis of archaeological records, it is thought that beer production provided the earliest source of widely available alcohol (Rudgley 1994). The first documentary evidence of alcoholic beverages was written in either Egyptian or Sumerian around 3200 BC, and included a recipe for making beer (Katz & Voigt 1986). Some 1500 years later is the first record of prohibitionist teaching, by an Egyptian priest, who wrote to his pupil: “I, thy superior, forbid thee to go to the taverns. Thou art degraded like beasts” (Crafts et al. 1900). Archaeobotanical studies have revealed evidence of beer-brewing in Mediterranean France and the consumption of alcoholic beverages during the 5th century BC (Bouby et al. 2011). Wine, and the potential problems related to too much wine, were written about extensively by the ancient Greeks and Romans. The Roman philosopher Seneca, for example, classified entrenched drunkenness as a form of insanity, linking the concept to a diseased state.

Libations have been used by virtually all cultures through most of their recorded history, and the commonality across societies suggests a role for positive, culturally adaptive mechanisms. Where alcohol was traditionally consumed, production of alcoholic beverages commonly occurred on a small household scale. Drinking alcohol was thus often an occasional and communal activity, associated with particular communal festivals. In societies where there was no indigenous alcohol consumption, their modern age encounter with alcoholic beverages was often problematic (McGovern 2009). Dietary alcohol has played and continues to play important social roles in human society. Alcohol, and other psychoactive chemicals, has played a role in cultural rituals, and in modern times drinking alcohol is largely a social activity in most countries.

In Europe in the sixteenth century, alcohol (called “spirits”) was used largely for medicinal purposes. At the beginning of the eighteenth century, the English Parliament passed a law encouraging the use of grain for distilling spirits. Cheap spirits flooded the market and reached a peak in the mid-eighteenth

century. In Britain, gin consumption reached 18 million gallons and alcohol-related problems became widespread, inspiring Hogarth's 1751 classic etching, "Gin Lane" which vividly depicts London's poor grappling with gin misuse. The years of the Industrial Revolution then brought a marked reduction in alcohol consumption, followed by a renewed upturn towards the end of the 19th century and another decline continuing into the second half of the 20th. From the 1950s England witnessed its most recent increase, particularly in per-capita wine consumption (Withington & McShane 2009).

Before the ready availability of clean water, the antiseptic and acidic power of alcohol in beverages may have killed pathogens and thus been a safer alternative to drinking water. This view is by no means uncontroversial and for brevity will not be explored further in this work (Phillips 2014).

2.2 PHYSIOLOGY OF ALCOHOL

Ethanol (alcohol) is a colourless liquid found in all alcoholic beverages. Besides ethanol, other fermentation products are found in alcoholic beverages, and these affect the taste of different beverages. Examples include sugars, aldehydes, histamine, phenols methanol, butanol, iron, lead, and cobalt. The chemical equation for ethyl alcohol (ethanol) is C_2H_5OH . Ethanol is a simple, highly water- and lipid-soluble molecule that can easily move across cell membranes, and quickly reaches equilibrium between blood and tissue. Ethanol penetrates the blood-brain barrier and inhibits central nervous system (CNS) functions; it is directly toxic to the brain.

Ethanol is mainly absorbed in the proximal small intestine, but also in the colon, and stomach, and also minimally in the mucous membranes of the mouth and oesophagus. Absorption is increased in the presence of carbonation, and reduced in the presence of proteins, fats, carbohydrates and other products of the fermentation of alcohol.

Metabolism of ethanol occurs mainly in the hepatic cytosol and mitochondria and is enzymatically mediated. Some 2% to 10% may be excreted directly by the lungs, or through urine or perspiration. Here alcohol dehydrogenase (ADH) produces acetaldehyde, which is rapidly broken down by aldehyde dehydrogenase (ALDH). In high doses, aldehyde may produce histamines, which may reduce blood pressure, and lead to nausea and vomiting.

When there is a high blood-ethanol concentration, a second metabolic pathway activates in the smooth endoplasmic reticulum of the microsomal ethanol oxidizing system (MEOS), which is responsible for approximately 10% of the oxidation of ethanol.

2.3 EFFECTS OF ALCOHOL ON THE BODY

Lower risk alcohol consumption (one to two drinks per day) causes peripheral vasodilation and decreases contractility of the heart, resulting in a mild decrease in blood pressure (Krenz & Korthuis 2012). Increases in high-density lipoprotein and clotting mechanisms alterations have been reported in people who typically have one or more drinks per day. This may even confer a cardio-protective effect (Roerecke & Rehm 2012) although the 2016 UK Department of Health guidelines suggest that these have been overestimated in the past and may now apply only in older age groups. Furthermore the guidelines note that any level of drinking increases health related risks (Department of Health 2016). In otherwise healthy people, alcohol is toxic to nearly all the organ systems at doses above one to two drinks per day (Guo & Ren 2010). Cellular toxicity can be initiated by the metabolism of ethanol and subsequent accumulation of acetaldehyde, a metabolite that can damage intracellular proteins and induce cell death through apoptosis, and changes in oxidation–reduction and cellular respiration (Guo & Ren 2010; Manzo-Avalos & Saavedra-Molina 2010).

Chronic intake of alcohol generally increases the risk of damage to the gastrointestinal, cardiovascular, immune, nervous, and other systems. Gastro-oesophageal bleeding, acute pancreatitis, fatty liver, alcohol-induced hepatitis, cirrhosis, hepatic cancer and hepatocellular carcinoma are all more prevalent in alcoholics than in the unaffected population (Nitsche et al. 2011; Gao & Bataller 2011). Depression, epilepsy and haemorrhagic stroke can occur secondary to high alcohol consumption, and alcohol is involved in a high proportion of road accidents and episodes of violence (Rehm et al. 2009). Alcohol and its metabolite acetaldehyde are carcinogens, and excessive alcohol consumption is associated with increased risk for many cancers including oral, oropharyngeal, breast, pancreas, stomach, and liver. Alcohol consumption by pregnant mothers can result in birth defects that comprise the foetal alcohol spectrum syndrome (Sulik 2005).

Consuming three or more drinks per day is a factor in mild-to-moderate hypertension and heavy drinkers are at increased risk for coronary artery disease and cardiomyopathy. The effects of heavy drinking can range from left ventricular impairment and arrhythmia to heart failure as a result of limited contractility of heart muscle. Binge drinking (e.g. a single exposure to 90 mL of 80-proof whiskey) can produce atrial or ventricular arrhythmias, even in individuals who have no other evidence of heart disease (Greenspon 1983).

Alcohol-dependent individuals may experience peripheral neuropathy characterized by tingling or numbness, especially in the hands and feet. A progressive neurologic syndrome that affects gait and stance, often accompanied by nystagmus, can result from atrophy of the cerebellum due to alcohol toxicity (Chopra & Tiwari 2012). Less common are neurologic syndromes that result from thiamine deficiency secondary to heavy drinking: Wernicke's syndrome consists of encephalopathy, uncoordinated muscle movement, and eye muscle weakness; and Korsakoff's syndrome is characterized by amnesia.

Molecular determinants of excessive alcohol consumption are difficult to study in human brains for obvious reasons, and most work has focused on animal brains (e.g. Mulligan et al. 2006). Neural cell adhesion molecules (NCAMs) have been implicated in rodent studies of in utero ethanol exposure (Miñana et al. 2002). Studies of human post-mortem (PM) brain samples of people with AD or alcohol misuse diagnoses have reported altered gene, microRNA and transcriptome profiles. Specific reported genes were involved in myelination, ubiquitination, apoptosis, cell adhesion, neurogenesis, and neural degeneration (Sokolov et al. 2003; Liu et al. 2006). Similar findings of altered expression have been found in PM samples of people with BP and SZ. (Iwamoto et al. 2004; Moreau et al. 2011; Mirnics et al. 2001).

2.4 DIAGNOSTIC CRITERIA AND CLASSIFICATION

The term 'alcoholism' was first coined by Magnus Huss in 1849 to describe the persistence of drinking despite adverse effects on the drinker's health. Huss described alcoholism as, "as a conjunction of pathological manifestations of the central nervous system, in psychic, sensory, and motor spheres". Huss noted that alcoholism was found in people who consumed alcoholic beverages in a continuous manner, in excess, and over a longer period than others.

In 1960, Morton Jellinek added the concept of alcoholism causing personal harm to the drinker and to wider society. He categorized alcoholism as a disease based on the quantities of alcohol consumed (Jellinek 1960). Today the World Health Organization (WHO) defines an "alcoholic" as an excessive drinker, whose dependence on alcohol is accompanied by mental disturbance, poor physical health, poor social relations, as well as poor social and economic behaviours (WHO 2014).

Edwards & Gross in their classic 1976 paper on the “alcohol dependence syndrome” (ADS) which informed modern diagnostic criteria included a number of elements: narrowing of repertoire; increased salience of alcohol over other things; tolerance; withdrawal (which was relieved by drinking); awareness of compulsion to drink; and reinstatement after abstinence (Edwards & Gross 1976).

The two main diagnostic systems used by psychiatry during my research for this thesis were the American Psychiatric Association’s (APA) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (APA, 2003) and the World Health Organization’s (WHO) International Classification of Diseases, Tenth Revision (ICD-10) (WHO, 1992). These provide accepted diagnostic criteria for the phenotypes “alcohol abuse/harmful use” and “alcohol dependence” under the category of Addictive Disorders. These two distinct categories have replaced the term, “alcoholism,” which was first discarded in DSM-III (APA, 2003). The ICD-10 definition of AD is broadly similar to that of DSM-IV. “Harmful use” (ICD-10) and “abuse” (DSM-IV) refer to substance use resulting in physical or mental harm in the absence of dependence.

Alcohol dependence (syndrome) is commonly defined as a disorder characterized by physiological, behavioural and cognitive effects in individuals who consume large amounts of alcohol. AD is diagnosed in ICD-10 as follows:

“A definite diagnosis of dependence should usually be made only if three or more of the following seven criteria have been present together at some time during the previous year:

- (a) a strong desire or sense of compulsion to take the substance;*
- (b) difficulties in controlling substance-taking behaviour in terms of its onset, termination, or levels of use;*

- (c) *physiological withdrawal state when substance use has ceased or been reduced, as evidenced by: the characteristic withdrawal syndrome for the substance; or use of the same (or a closely related) substance with the intention of relieving or avoiding withdrawal symptoms;*
- (d) *evidence of tolerance, such that increased doses of the psychoactive substances are required in order to achieve effects originally produced by lower doses (clear examples of this are found in alcohol and opiate-dependent individuals who may take daily doses sufficient to incapacitate or kill nontolerant users);*
- (e) *progressive neglect of alternative pleasures or interests because of psychoactive substance use, increased amount of time necessary to obtain or take the substance or to recover from its effects;*
persisting with substance use despite clear evidence of overtly harmful consequences, such as harm to the liver through excessive drinking, depressive mood states consequent to periods of heavy substance use, or drug-related impairment of cognitive functioning;

Narrowing of the personal repertoire of patterns of psychoactive substance use has also been described as a characteristic feature (e.g. a tendency to drink alcoholic drinks in the same way on weekdays and weekends, regardless of social constraints that determine appropriate drinking behaviour). ” (WHO, 2014).

Alcohol dependence was until recently classified separately in the DSM-IV, which was the prevailing North American system in use at the time of this research. The DSM-IV criteria for AD are:

“A maladaptive pattern of alcohol use, leading to clinically significant impairment or distress, as manifested by three (or more) of the following, occurring at any time in the same 12-month period:

1. *Tolerance, as defined by either of the following:*

- a. *A need for markedly increased amounts of alcohol to achieve Intoxication or desired effect*
 - b. *Markedly diminished effect with continued use of the same amount of alcohol*
2. *Withdrawal, as manifested by either of the following:*
 - a. *The characteristic withdrawal syndrome for alcohol*
 - b. *Alcohol (or a closely related drug such as valium) is used to relieve or avoid withdrawal symptoms*
3. *Alcohol is often used in larger amounts or over a longer period than was intended*
4. *There is a persistent desire or unsuccessful efforts to cut down or control alcohol use*
5. *A great deal of time is spent in activities necessary to obtain alcohol, use alcohol, or recover from its effects*
6. *Important social, occupational, or recreational activities are given up or reduced because of alcohol use*
7. *Alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by alcohol (e.g. continued drinking despite recognition that an ulcer was made worse by alcohol consumption)."*

The current fifth version of the DSM (DSM-5) has consolidated the separate DSM-IV (substance-specific) categories of dependence and abuse into a single Substance Use Disorder (SUD). The criteria for SUD merge the previous lists of 7 criteria for dependence and 4 criteria for abuse into a single list of 11 criteria. SUD is now additionally defined by the presence of craving, while the criterion of recurrent legal problems has been discarded, largely because legal and cultural differences made the criteria difficult to apply internationally. Severity of SUD is graded by the number of criteria met: 0–1 (no diagnosis); 2–3 (mild SUD); 4–5 (moderate SUD); 6 or more (severe SUD). Thus, a quantitative

spectrum disorder has replaced the dichotomous harmful use/abuse vs dependence classification of ICD-10/DSM-IV in DSM-5 (APA, 2013). Interestingly some 35 years prior, Edwards and Gross had also described alcohol dependence as a graduated syndrome (Edwards & Gross 1976).

The DSM and ICD definitions of AD/SUD are broad and potentially include diverse sub-phenotypes, all of which may contribute to the range of presentations observed in clinical practice. For example, a patient may have alcohol dependence without significant tolerance or withdrawals by fulfilling sufficient other criteria.

Alcohol dependence (Online Mendelian Inheritance in Man reference #103780) is considered the most severe psychiatric alcohol-use phenotype, and has been the focus of most of the studies considered in this chapter. Earlier studies may have used other/current at the time definitions of alcoholism. Many source papers refer to ‘alcoholism’, ‘alcohol dependence’ and ‘alcohol dependence syndrome’ interchangeably. In this thesis, I have used AD, ADS and ‘alcoholism’ interchangeably. I have used DSM-IV and ICD-10 definitions of alcohol dependence (AD) as the main phenotype. This is because these were the diagnostic systems in use during the bulk of the research project and writing and they allow meaningful comparisons with the bulk of the literature.

2.5 EPIDEMIOLOGY- WHY ARE ALCOHOL USE DISORDERS IMPORTANT?

Alcohol use disorder (AUD) is a severe and common disorder occurring worldwide. It is responsible for significant global morbidity, mortality and economic loss, and has contributed directly and indirectly to immeasurable human suffering. Excessive alcohol consumption is associated with violent crime, aggression and increased risk of accidents, injury and death.

AUD is a serious global public health concern (Room et al. 2005). According to the World Health Organization (WHO) 2 billion individuals worldwide consume or misuse alcohol. Alcohol misuse causes up to 2.5 million deaths per year globally. WHO has also estimated that approximately 20%–30% of oesophageal cancer, liver cancer, cirrhosis of the liver, homicide, epilepsy, and road traffic accidents worldwide result from the harmful use of alcohol. Alcohol consumption has been identified as an important risk factor for more than 200 different disorders and injuries (World Health Organization 2014). An estimated 3.8% of all global deaths and 4.6% of the global burden of disease (measured in disability-adjusted life-years) are attributable to alcohol (Rehm et al. 2009). UK data suggests that there were a million alcohol-related hospital admissions in 2013 and over 6 thousand alcohol-related deaths in England alone (Health and Social Care Information Centre 2014).

Point prevalence of alcohol use disorders (AUDs, harmful use and dependence) in adults has been estimated to be around 1.7% globally according to WHO Global Burden of Disease 2000 analysis. The rates are 2.8% for men and 0.5% for women. The prevalence of alcohol use disorders varies widely across different regions of the world, ranging from very low levels in some Middle Eastern countries to over 5% in North America and parts of Eastern Europe (WHO 2014).

Twin data suggest that more than 40% of the societal burden of brain disorders is likely to be genetically mediated. Most of this disease burden arises from complex multi-gene genetics as well as from environmental influences (Uhl & Grow 2004).

Despite the known harmful effects of chronic alcohol consumption, our understanding of the aetiology of alcohol misuse is incomplete. This limits opportunities for development of preventative and therapeutic strategies.

2.6 AD AS A COMPLEX PSYCHIATRIC DISORDER

Alcohol dependence represents a complex and heterogeneous phenotype, with behavioural, as well as psychological, pharmacological, and medical components. AD belongs to a group of conditions known as complex genetic disorders which include BP and SZ (Heath et al. 2012). By way of illustration, the genes contributing to another complex disorder like diabetes are likely to vary across families through various vulnerability and protective factors, for example insulin tolerance, weight gain, hypertension, cholesterol, etc.

In AD, as with other complex disorders, it is likely that multiple genes influence a range of intermediate characteristics that subsequently interact with the environment to produce the condition (McGue 1999; Hyman & Malenka 2001; Kendler 2008; Edenberg et al. 2010a; Dick et al. 2011). Therefore, any gene that affects a specific intermediate characteristic or phenotype, such as insulin sensitivity, is likely to contribute to the vulnerability toward diabetes, but may explain only a small proportion of the vulnerability. AD can thus be thought of as a clinical outcome that has been generated by a combination of many genetic and environmental risk and protective factors. It is commonly thought that many genes are involved in each disorder with each gene conferring only a small effect on the phenotype. The individual risk variants are thus without diagnostic predictive value, and any estimations of risk are probably going to change in the future as large epidemiological samples become available for analysis (Gejman et al. 2010).

Our understanding is now shifting from single gene models to a polygenic model of alcoholism, but despite the wealth of research, its genetic architecture still remains largely unknown (Kraft & Hunter 2009). The concept that complex disorders result from dysfunction of entire molecular networks, i.e. system disorder, as opposed to single genes, is gaining traction in the literature (Schadt 2009). Whether this applies to AD is still an empirical question that remains to be addressed.

Hopefully this work will contribute meaningfully to that debate.

2.7 MOLECULAR GENETIC STUDIES

Prior to discussing the findings from the genetic studies of AD, I will discuss some of the components contributing to the complexity of psychiatric disorders in general, using AD and other illnesses as illustrations. Although numerous studies have been published on genetic associations with several mental disorders, there is evident inconsistency of the results among these results. Outcome divergence in genetic studies, particularly of complex phenotypes, may be largely ascribed to inadequate sample sizes, phenotypic heterogeneity, locus heterogeneity (Schulze & McMahon 2003) and pleiotropy (Solovieff et al. 2013). As the late Prof. Hugh Gurling said in 1986: “Specific strategies will need to be adopted to overcome uncertain modes of inheritance, incomplete or non-penetrance of disease alleles and disease heterogeneity” (Gurling 1986).

2.7.1 Challenges of genetic studies in general

Sample size

Until recently a significant challenge in both linkage and association studies of complex disorders has been inadequate sample size for statistical power to detect genetic differences. Several international meta-analyses of genome-wide scans have been published in the last decade, resulting in the detection of weak, and at times inconsistent, signals across studies (Craddock & Sklar 2009). This may be explained by the work of Suarez & Washington (1994) who found that for an oligogenic trait simulated under reasonable parameters, when sampling families sequentially, the first true linkage to be detected is unlikely to be replicated until many generational samples have been studied. Thus the power to replicate the first linkage is achieved only when multiple samples several times larger than the initial sample have been collected and analysed (Badner & Gershon 2002).

Consortia combining genome wide data for meta-analysis, or replication studies, have come a long way in increasing sample sizes. The largest group is the Psychiatric Genetics Consortium (PGC, <http://www.med.unc.edu/pgc>) of which UCL is a member. This group has published on BP, SZ, major depression, and cross-disorder studies, with sample sizes in the tens of thousands. The largest AD GWAS contained over 16,000 subjects. As a result, novel markers have revealed new pathways and understanding of the biology of these complex disorders.

Phenotypic heterogeneity

Psychiatric disorders do not yet have reliable biomarkers, so we use diagnostic or research criteria typically based on interview, clinical record and collateral history. Diagnosis in psychiatry, despite our best efforts at improving validity and reliability, is still subject to human bias and error. At the present time, a clear phenotype that best captures the underlying mechanism of AD or other psychiatric disorders has not yet been defined. Identification of the affected phenotype for AD is however considered accurate and reliable (Bucholz et al. 1994; Üstün et al. 1997; Hasin et al. 2006). I discuss one way of approaching this for AD by using endophenotypes in section 2.7 below.

Locus heterogeneity

Locus heterogeneity describes a state when mutations at more than one gene locus cause the same phenotype. Allelic heterogeneity is a phenomenon when more than one allele at a specific gene locus results in the same phenotype (McClellan & King 2010). While linkage analysis is immune to allelic heterogeneity, locus heterogeneity reduces the power of association analysis in detecting disease variants. Ascertaining study samples from genetically isolated populations and using strict reliable phenotyping (or sub-phenotyping) could help alleviate the issue with locus heterogeneity to some extent, by reducing the number of different susceptibility genes (Heutink 2002). The main downside is that

variation found within one population may not apply to other humans – for example Chapter 6.2.6 on taste receptors.

Cross phenotype associations and pleiotropy

Pleiotropy occurs when a genetic locus genuinely affects more than one trait and is one possible underlying cause for an observed cross-phenotype (CP) association (Solovieff et al. 2013). A recent evaluation of genome-wide-significant SNPs listed in the National Human Genome Research Institute (NHGRI) Catalogue of Published Genome-Wide Association Studies found that 4.6% of SNPs and 16.9% of genes have CP effects (Sivakumaran et al. 2011). Pleiotropy can occur at the gene (or region) level, at which multiple variants in the same gene (or region) are associated with different phenotypes; or the allelic level, where a single causal variant is related to multiple phenotypes. Distinguishing the associations that represent genuine shared effects of single variants from those that represent the effects of co-localizing but independent variants is key. Statistical and fine mapping approaches to discover and address CP associations as well as ensuring the highest standards for GWAS studies are adhered to should reduce the bias introduced by these effects (Solovieff et al. 2013).

2.7.2 Alcoholism-related endophenotypes

Although the diagnoses for alcohol use disorders are based on a range of reported symptoms, they are typically treated as a binary outcome (affected or unaffected). As early as the 1960s, it was conceptualized that alcoholism was not a single entity and that various “types” of alcoholism were noted clinically. Jellinek for example originally identified five “species” of alcoholism characterized by psychological and physiological dependence (Jellinek 1960).

An endophenotype is a biological or a psychological characteristic transmitted with a disease, with a strong biological substrate and a higher heritability than the disease itself. Researchers have utilized and refined such endophenotypes order to identify more aetiologically homogeneous subtypes as a means for

studying, diagnosing, and treating alcoholism, for example Cloninger's classic paper on predisposition to petty criminality in Swedish adoptees (Sigvardsson et al. 1982). With reference to schizophrenia research genetic theories, Gottesman and Shields originally described endophenotypes as internal phenotypes, not obvious to the unaided eye, which can fill the gap between the gene and the available descriptors of disease (Gottesman & Shields 1973). Subsequently, Tsuang et al suggested the following criteria for evaluating endophenotypes in psychiatric genetics research: biological plausibility, specificity, state independence (stable over time), heritability, familiarity, and cosegregation (Tsuang et al. 1993). The main benefit of using endophenotypes is that the number of genetic and environmental factors that contribute to these should be reduced compared to the full clinical syndrome and thus easier to classify (Carlson et al. 2004).

A number of endophenotypic traits that influence susceptibility to AD have been investigated. These include variation in hangover symptoms (Piasecki et al. 2010), alcohol consumption, level of response (LR) to alcohol, endocrine measures, electrophysiology, personality traits, and drinking behaviour (Hines et al. 2005). Schuckit proposed that the majority of genetically related markers of AD risk were represented by five relatively independent main endophenotypes, including: metabolizing enzymes; LR; major psychiatric disorders (e.g. BP, SZ); the opioid system; and neuronal or behavioural disinhibition (Schuckit 2002).

Alcohol-related phenotypes are typical quantitative traits, with population variation attributable to multiple segregating loci with effects that are sensitive to environmental exposures. Below I discuss the more relevant endophenotypes.

2.7.3 Low response to alcohol

A low level of response (LR) to alcohol refers to a relatively mild objective or subjective state of intoxication following alcohol challenge (assessed by self-report, including subjective “high”/“buzz” effects and objective body sway (Schuckit 1994). An estimated 40% of children of alcoholics have a LR to alcohol, and prospective studies have shown that it may be a predictor of future development of alcohol use disorders among offspring (Schuckit et al. 1996; Pollock 1992). In a longitudinal study of young men, the presence of the LR trait at age 20 years was associated with a fourfold increase in the risk of AD in later life (Schuckit et al. 1997). People who are at high risk of developing AD (e.g. having a parent with AD) report having to consume greater amounts of alcohol to get the same effects as other people (Schuckit et al. 2001). Offspring of parents with alcohol problems generally report greater hangover symptoms than do those of non-alcoholics in some groups e.g. (Slutske et al. 2003) although differences were not found by other groups, e.g. (Span & Earleywine 1999).

Candidate genes for alcohol response include the opioid and serotonin systems, cyclic AMP (cAMP), and the alcohol metabolism enzymes. For example, Ray and Hutchison found an association between sensitivity to the effects of alcohol and the A118G allele of the μ -opioid receptor gene (Ray & Hutchison 2004). A polymorphism in the promoter region of the serotonin transporter gene (*5'HTLPR*, locus ID *SLC6A4*) has been associated with self-reported feelings of intoxication during an alcohol challenge protocol in a non-psychiatric sample (Fromme et al. 2004). Further information is needed regarding familial association and cosegregation applied to alcohol response endophenotype.

2.7.4 MaxDrinks

Twin studies have that demonstrated a moderate genetic influence on the maximum number of drinks a person consumes in 24 hours (MaxDrinks). The COGA studies reported that the rate of alcoholism increased with increasing MaxDrinks. For instance, among participants who reported drinking 9 or more

drinks in a 24-hour period, 65 percent of men and 53 percent of women were diagnosed with alcohol dependence (Saccone et al. 2000).

2.7.5 Electrophysiological phenotypes

Several brain electrophysiological measures, such as EEGs and event related potentials (ERPs), are altered in people with various psychiatric disorders, including AD (Begleiter et al. 1998). For example, the P300 ERP brain wave has a lower amplitude in people with AD than those without. Research also suggests that the variations in brain electrophysiological activity predate the development of alcoholism. For example, compared with people without a family history of alcoholism, alterations in EEGs and ERPs exist both in alcohol-dependent people and in non-alcoholic relatives of alcoholics – i.e. those at risk for alcoholism because they are genetically related to the person with AD. These electrophysiological measures therefore represent biological markers that are related to a predisposition for developing AD and have been used to identify two genes in the COGA study [GABRA2 (Edenberg et al. 2004) and CHRM2 (Wang et al. 2004)].

2.7.6 Alcohol expectancies

Alcohol expectancies (AE) —i.e. Beliefs about a drug's impact on behaviour—and subjective experience of alcohol's effects have been shown to be a genetically influenced characteristic. AEs have estimated heritability factor between 0.4 and 0.6 (Heath et al. 1999; Schuckit et al. 2001) with greater alcohol consumption found for high-risk (e.g. sons of alcoholics) compared to low-risk families (Newlin & Thomson 1999).

2.7.7 Craving

Alcohol craving has been defined as a strong desire to consume alcohol and has been associated with loss of control over drinking, which is part of the alcohol dependence syndrome, as defined in the DSM-

IV. Although there has been some disagreement about the definition of the term, the endophenotype of craving is a construct that is central to alcohol dependence and is often a target of treatment.

Measuring subjective “craving” is challenging in humans, however craving and loss of controlled drinking have been biologically linked to the actions of alcohol on the brain mesolimbic and mesocortical dopamine pathways. These are the neural pathways that may underlie the attribution of incentive salience to alcohol. Individual variation in the ability to stop drinking and loss of control may be related to genetic factors that influence the effects of alcohol on craving and the linked neural substrates. For example, a study by Hutchison et al found that individuals with the long variant of the D4 dopamine receptor gene (*DRD4 VNTR*) showed higher craving after consumption of alcohol, compared with a placebo drink (Hutchison et al. 2002). This same variant affects response to the atypical antipsychotic olanzapine in reducing alcohol cravings (Hutchison et al. 2003).

2.7.8 Personality/temperamental factors

Personality or temperament is a complex multi-dimensional construct and not the primary focus of this work. However it would be remiss not to mention personality factors concisely. Personality is often described as is the characteristic manner in which one thinks, feels, behaves, and relates to others, and it tends to remain constant throughout life.

Several studies suggest that genetic susceptibility to AD may be expressed, at least in part, through an individual's temperament. Tarter and Vanyukov (1994), for example, proposed a temperament model of alcoholism risk based on five traits that increase an individual's AD risk burden . These heritable traits include behavioural activity level, sociability, attention span/persistence, emotionality, and "soothability." An individual's liability is altered by the deviation of each trait from the population norm.

Each of these trait clusters that may represent a "difficult" temperament relate to an increased risk for developing a problem with substance use (Tarter et al. 1990; Ohannessian & Hesselbrock 2008).

Personality traits can also be used as genetic endophenotypes of the risk for SUD (Belcher et al. 2014). Measures of risk-taking, novelty seeking, reward dependence, and low conscientiousness are all associated with increased risk for developing a SUD (see Widiger 2011 for a review). For example, self-reported impulsivity, a form of novelty-seeking was associated with more pleasurable drug effects and consequent increased use in a cross-sectional sample of adolescents (Bidwell et al. 2015). People with personality traits which are closer to the population norm are thought to have more control over their own behaviour; conversely people with behavioural and emotional regulation challenges may be at higher risk of AD. These traits may in turn affect environmental influences and stressors, for example seeking out environments conducive to alcohol and drug use.

Antisocial behaviour, and the closely related personality traits, behavioural disinhibition and emotional dysregulation, are closely linked to development of all SUDs (Baker et al. 2004; Iacono et al. 2008; Tarter et al. 2003). The combination of antisocial behaviour and disinhibited personality traits (termed "externalizing" or "behavioural disinhibition") is associated with a higher familial loading, earlier onset, and a more severe and persistent course in adulthood (Hicks et al. 2010; Iacono et al. 2008). Diverse personality traits may also predict different patterns of substance use (Conrod et al. 2013; Elkins et al. 2006).

One example of the interplay between personality factors and substance misuse comes from a longitudinal twin study by Hicks et al (2013). The authors found that genetic risk in socialization contributed to substance misuse indirectly via increased exposure to contextual risk. In addition, the

“boldness“ trait appeared to index an independent and direct genetic risk factor for adolescent substance misuse (Hicks et al. 2013).

Variation in personality traits is attributed to genetic influences in the region of 30–60% (Power & Pluess 2015). Identification of associated genetic variants has however remained elusive. Consequently, knowledge regarding the molecular genetic architecture of personality and to what extent it is shared across the different personality traits is limited.

Verweij et al in the first genome-wide association study of Cloninger's temperament scales in a sample of 5117 individuals, found no association between subjects' scores on harm avoidance, novelty seeking, reward dependence, and persistence and testing of over a million SNPs, as well as gene-based association tests and biological pathway analyses (Verweij et al. 2010).

In contrast, another measure of personality – the Big Five Inventory, consisting of metrics named openness to experience, conscientiousness, extraversion, agreeableness and neuroticism – was recently explored using genomic-relatedness-matrix residual maximum likelihood analysis (GREML) in >5000 people by Power and Pluess (Power & Pluess 2015). They found significant and substantial heritability estimates for neuroticism and openness, but not for extraversion, agreeableness and conscientiousness.

Closely related to personality traits are affective temperaments. Affective temperaments are also a set of biologically determined and heritable characteristics thought to shape the way a person views and acts by interacting with environmental factors to shape personality (Moore et al. 2005). Sometimes the terms personality traits and affective temperaments are used interchangeably in the literature (Greenwood et al. 2013). Specific affective temperaments may be characteristic of certain personalities, or serve as a starting point for an episodic affective illness (Blöink et al. 2005). Affective temperaments are often

considered along a spectrum at the extremes of which are diagnosable mental disorders (Rihmer et al. 2010; Harnic et al 2013).

Affective temperaments may also refer to subaffective trait expressions representing the earliest subclinical phenotypes of affective disorders which are present during euthymic periods of affective disorders (Maremmani et al. 2005). They are thought to increase risk of mental illness or SUD (or pleiotropically, both). For example, high trait anxiety and hyperthymic temperament are commonly found in people with BP and AD (Levander et al. 2007; Azorin et al. 2010; Harnic et al.)

Affective temperaments are also increasingly being recognized as important risk factors for developing SUDs. For example, Pacini et al (2009) found higher scores for depressive, cyclothymic and irritable temperaments in patients with AD compared to controls. Irritable temperament may also be a risk factor for drug misuse disorders in patients treated for alcohol dependence (Khazaal et al. 2013). Vyssoki et al. (2011) found that higher cyclothymic scores influenced the age of onset of alcohol misuse and AD. Maremmani and colleagues (2009) found that methadone-treated individuals with heroin dependence had significantly higher cyclothymic and irritability scores compared to unaffected controls (Maremmani et al. 2009; Vyssoki et al. 2011). Maremmani and colleagues further commented that these results, “could represent the temperamental profile of heroin addicts, largely irrespective of comorbidity, and tend to cohere with previous conceptualizations hypothesizing "sensation-seeking" (and "novelty-seeking") as the main personality characteristics of addiction.”

Most recently, Singh et al (2015) in a study of subjects from the Bipolar Disorder Research Network (BDRN) compared measures of temperament in subjects with and without alcohol misuse diagnoses. They found that mean scores for four affective temperaments (hyperthymic, cyclothymic, depressive and irritable) were higher in cases (BP+AD) than controls (BP only) ($p < 0.001$). Hyperthymic and irritable

temperaments in particular significantly increased the odds of concomitant alcohol misuse within the BP sample after adjustment for potential confounders.

2.7.9 Psychiatric comorbidity

Psychiatric comorbidity is a commonly occurring feature that is not consistently considered an endophenotype but is an established risk factor for developing AD and other SUDs. In this thesis I have examined two sets of patients with psychiatric comorbidity and I have attempted to separate the genetic influences of the co-morbid psychiatric illness and the alcoholism. Psychiatric disorders have generally higher heritability estimates than AD, although these do vary. Heritability of BP is discussed in Chapter 7 and of SZ in Chapter 8. Both chapters include discussion on overlap with AD.

2.8 CONCLUSIONS

This chapter highlights the complexity of phenotype classification in psychiatric genetic research, which I take further in the Discussion section. In the next session, I explore the evidence for a genetic influence in the development of alcoholism.

CHAPTER 3: HERITABILITY - IS ALCOHOLISM A “GENETIC” DISORDER?

“Ebrii gignunt ebrios” (Drunks beget drunkards)

—attributed to Plutarch, c110AD

It has long been recognized that a tendency towards AD runs in families, with the quote above being one of the better known early records. Whether this is nature or nurture has been the subject of much debate. Genetic effects in alcoholism were first established using classical quantitative genetic family, twin and adoption studies (Kaij, 1960; Bohman, 1978; Sigvardsson et al. 1996; Sigvardsson et al. 1982; Ball 2005). These methods each have their strengths and limitations. I have summarized these approaches and the evidence supporting the heritability of alcoholism in this chapter.

3.1 WHAT IS HERITABILITY?

Heritability is a measure of genetic influence. Specifically it is the degree of variation in a population attributable to genetic variation. The heritability of alcohol dependence is well recognized, with higher alcoholism rates occurring among blood relatives of affected individuals compared to non-relatives. AD heritability estimates range between 40 and 70%. Research establishing the heritability of AUDs primarily utilizes three research designs: twin studies, adoption studies, and family studies.

3.2 ANIMAL STUDIES

Before considering human family and related studies, I will briefly mention the importance of animal models in understanding AD. Much data supporting significant genetic contribution to alcohol dependence comes from animal research. Although many of the genes/regions of interest have also been studied in animal models, these will not be extensively reviewed here except where relevant to this thesis.

In selectively bred rodent lines, a considerable proportion of heritability in the development of addiction-related behaviour has been reported, including preference for alcohol over water, willingness to work for alcohol, sensitivity to the hypnotic or activating effects of alcohol and alcohol withdrawal, and demonstrations that alcohol is rewarding even in the presence of food and water (Buck et al. 1997; Crabbe et al. 1999; McBride & Li 1998; Murphy et al. 2002; Whatley et al. 1999).

In a fascinating UCL mouse study, Anstee et al found that alcohol consumption increases following mutations to the murine γ -aminobutyric acid A receptor (GABAAR) β 1 subunit gene (*Gabrb1*). The mutation causes spontaneous GABA ion channel opening and increases GABA receptor sensitivity linked to increased tonic currents in the nucleus accumbens, a region long-associated with alcohol reward. Mutant mice also work harder to obtain ethanol, and are more sensitive to alcohol intoxication (Anstee et al. 2013). Murine genes are largely homologous with human genes, and thus allow for meaningful comparisons across species.

Animal models have thus been useful in characterizing alcohol related behaviours, the effects of alcohol on organisms, and in understanding biological pathways for the development of AUDs, although clearly they cannot model the entire psychobiological diagnostic construct of alcoholism seen in humans (Lovinger & Crabbe 2005).

3.3 FAMILY STUDIES

Family studies of alcohol use disorders have clearly demonstrated familial aggregation for alcohol dependence and for comorbid disorders such as substance use disorders and mood and anxiety disorders (Bucholz et al. 1996; Nurnberger et al. 2004; Slutske et al. 2008; Todd et al. 1996; Davis et al. 2008; Winokur et al. 1993; Cotton 1979). Alcoholism is strongly recurrent in both the first and second degree relatives of alcoholic probands (Cotton, 1979). Although family studies have demonstrated that AD runs in families, this approach cannot tell us whether familial transmission is caused by environmental or genetic factors, since family members share some of their genes as well as important aspects of their environment. To distinguish between environmental and genetic factors, adoption and twin studies are conducted.

3.4 TWIN STUDIES

Twin studies have also been performed to investigate the extent to which shared environmental and nonshared environmental factors contribute to the development of AD. Monozygotic (MZ) twins share 100% of the same DNA, whereas dizygotic twins (DZ) have approximately 50% of the same DNA. Differences in prevalence between MZ and DZ twins can help us estimate the heritability of a phenotype. Twin studies also take advantage of the fact that pairs of twins are born at the same time and usually raised together, subsequently experiencing major life events in childhood at the same ages.

Heritability estimates for human diseases have inherent limitations but one twin study of alcohol dependence estimated heritability to be between 50% and 61% (Hansell et al. 2008; Kendler et al. 1997; McGue et al. 1992). The largest twin study, of 9,000 male pairs, found that alcoholism was 53% heritable (Kendler et al. 1997). However, twin studies do not always show substantial heritability or increased concordance for alcoholism in MZ compared to DZ female twins (Gurling et al. 1984, Pickens

et al. 1991, Prescott et al. 2007, Prescott et al. 2005). In the Australian twin sample study, the estimated contribution to the variance of liability to AD was 1% for shared environmental factors and 35% for non-shared environmental factors (Heath et al. 1997). Stronger and more consistent evidence supporting genetic influences for AUDs has been found for males, but several twin studies support similar conclusions for females (Cadoret et al. 1987; Cadoret et al. 1995; Kendler et al. 1992).

Longitudinal formal genetic studies have shown that the relative influence of genetic and environmental risk factors on AD and alcohol-related phenotypes fluctuates over time. Research has revealed alternation between periods in which genetic influences predominate and periods in which environmental influences are more dominant. A detailed longitudinal study by Kendler et al. (see Kuo et al. 2006; Kendler et al. 2011) showed that environmental influences on alcohol consumption were highest in adolescence. This finding suggests that adolescence may be the optimal time point for educational interventions.

The early twin and adoption studies can be interpreted as suggesting that there is a higher heritability for alcoholism combined with criminality than for other subtypes of alcoholism (Kaij, 1960, Gurling et al. 1984, Gurling and Cook, 1999). Men with alcohol dependence have much higher rates of dissocial personality disorder and criminality than women with alcoholism (Cloninger et al. 1978, Cloninger et al. 1986).

Women with alcohol dependence have higher rates of depression and anxiety than men, and the heritability of AD may be higher in women (Bierut et al. 1999; Gurling and Cook 1999; Klerman et al. 1996; Kuo et al. 2006a; Merikangas et al. 1985; Merikangas et al. 1994). The relationship between anxiety disorders and alcoholism seems to be very similar to the relationship between alcoholism and depression (Merikangas et al. 1994; Mullan et al. 1986).

Twin studies have demonstrated that the amount of alcohol one consumes has a genetic influence (Kendler et al. 2011). Age at first drink appears to be associated with alcohol-related problem behaviour, but progression to alcoholism is under stronger genetic control than initiation, and the effect of early exposure to predict outcome is genetically mediated (Prescott & Kendler 1999).

3.5 ADOPTION STUDIES

The second approach for separating the impact of genes from that of the environment uses adoption studies. Unlike twin studies, in adoption studies individuals are exposed to a different environment to their biological or genetic relatives. Both a study of half siblings as well as evaluations of adopted-away children of people with AD support the conclusion that genes impact the risk for alcoholism (Goodwin et al. 1973; Schuckit et al. 1972). An early comparison of the drinking patterns of sons of alcoholic men raised by their fathers versus those of their full brothers who had been adopted away indicated that genes may be a more robust predictor of alcoholism than being raised by an alcoholic (Goodwin 1974).

Based on adoption studies, Cloninger et al. (Cloninger et al. 1981) suggested the existence of two types of alcoholism: type 1 was a mostly environmentally triggered “milieu limited”, late-onset type. Type 2 was male-limited type associated with a high genetic loading, early onset, higher risk of legal problems and moderate alcohol consumption. Type one alcoholism showed lesser genetic transmission and a greater environmental component and is described as “milieu limited”. Type one alcoholism was associated with depression and anxiety in both men and women (Bohman, 1978; Sigvardsson et al. 1996; Sigvardsson et al. 1982). This dichotomy has been critiqued on the basis that the distinction between type 1 and type 2 alcoholics no longer offered clinical subtypes with distinct severity once patients with comorbid antisocial personality disorder were excluded (Irwin et al. 1990). An alternative sub-categorization was suggested based on age of onset, the presence of childhood risk factors such as hyperactivity, and severity of alcoholism (Schuckit et al. 1995; Johnson et al. 1996). Alcoholism types

may thus vary on a continuum of severity, rather than represent distinctly different disease entities (Bucholz et al. 1996).

Two important adoption studies from Sweden found replicated evidence for genetic effects associated with criminality and male alcoholism, known as type two alcoholism, which had an early age of onset and strong genetic transmission (Sigvardsson et al. 1982). The largest published adoption study of alcoholism, the Stockholm Adoption Study (Sigvardsson et al. 1996), illustrates these findings. The study included 862 men born in Stockholm, who had been adopted in infancy and both their adoptive and biological parents. Alcohol misuse among the study participants was assessed by the number of registrations with local Temperance Boards — local agencies that document how often a person has been treated or cited for alcohol misuse. In this study, the rate of alcohol misuse among the adoptees was proportional to the number of alcoholic parents: alcohol misuse was 33.3% if both biological parents misused alcohol; 26.0%/22.4% if only the biological mother/father misused alcohol respectively; and 14.7 percent if neither biological parent misused alcohol. Thus, adoptees with one or two alcohol-misusing parents had a significantly greater risk of alcohol misuse than did adoptees with no alcohol-misusing parents.

Family, twin, and adoption studies do not indicate that simple Mendelian dominant or recessive factors are major contributors to the AUD risk. Rather, similar to what was described regarding heart attacks, alcoholism might be best viewed as a disorder influenced by multiple genes (i.e. it is polygenic), reflecting multiple characteristics (i.e. the causes are heterogeneous), and whose causes include both genes and environment (i.e. it is multifactorial).

3.6 HERITABILITY OF CO-OCCURRING PSYCHIATRIC DISORDERS

Emil Kraepelin noted a co-occurrence between alcoholism and bipolar disorder in his classic phenomenological work which also introduced the “Kraepelinian dichotomy” that SZ and BP were separate disorders (Kraepelin 1921). Psychotic disorders including SZ and BP co-occur commonly with alcohol and other substance use disorders (Buckley 1998; Dixon 1999; Clark 2011).

Bipolar disorder and schizophrenia for example carry a two- to four-fold higher prevalence of number of substance use disorders, including alcoholism. Bipolar disorder is sometimes grouped into Bipolar 1 (BP1, the more severe phenotype) and Bipolar 2 (BP2, a less severe phenotype) (APA 2000). The prevalence rate of bipolar illness and concurrent substance use disorders, by naturalistic studies, has been estimated to be 21–58% (Brady & Lydiard 1992), and the ECA Study reported a 60.7% lifetime prevalence rate for substance use disorders in BP1, with alcohol as the most common substance abused (Regier et al. 1990). In this study, people with BP1 and 2 had the highest lifetime prevalence rate of alcohol abuse or dependence (46.2% and 39.2%, respectively) followed by individuals with schizophrenia (33.7%), compared to the general population without a major mental disorder (13.8%). In addition, patients with mania were six times more likely to have an alcohol use disorder compared to those without a mental disorder. Similarly, in the National Comorbidity Survey, respondents with a lifetime diagnosis of alcohol dependence had a significantly increased likelihood of having co-occurring lifetime diagnosis of mania, which was much higher for men [odds ratio (OR) -12.03] than for women (OR -5.3) (Kessler et al. 1997).

There is also considerable cross-heritability between many psychiatric and addictive disorders. Genetic studies have suggested the existence of a common genetic influence for alcohol dependence/consumption and several diagnostic categories. For example, Preissig & Rigby reported the

risks of comorbidity of alcoholism in relatives with mood disorders including BP (Table 3.1, (Preissig & Rigby 2009)).

Table 3.1. Comorbidity of Alcoholism and Mood Disorders in relatives: Adjusted Odds Ratios (95% Confidence Limits) Controlled for Sex, Age, and Interview Status

Relatives	Alcoholism	
	Dependence	Abuse
Mood disorders (all)	4.6 (3.1-6.9)	1.9 (1.1-3.2)
Bipolar	8.1 (2.7-24.0)	14.5 (3.7-57.6)
Major depressive disorder	3.2 (2.0-5.0)	1.7 (0.9-3.3)

(From Preissig & Rigby 2009)

3.7 GENES AND ENVIRONMENT

Environmental factors and Gene x Environment (GxE) interactions are not the focus of this thesis, but a discussion would be incomplete without mentioning these.

Environmental factors include cohort of birth and ethnicity that affect the risk for developing AD.

Recent birth cohorts have higher prevalence rates of alcoholism than birth cohorts from the earlier part of the twentieth century (Reich et al. 1988; Grant & Dawson 1997). The differences in prevalence rates are thought to be due to variations in the availability of beverage alcohol resulting from Prohibition, economic depression, or wartime shortages, rather than significant changes in the human genome.

Differences in prevalence rates and the course of alcohol dependence have also been noted in relation to ethnicity. For example, there appears to be little difference among ethnic groups with regard to first age of the early stages of alcoholism, but differences in first age begin to appear with later stages of the disorder (Ulloa et al. 2014; Chartier et al. 2009; Hesselbrock et al. 2003).

The variability in apparent vulnerability to alcoholism is thought to be due to biological differences as well as social factors (e.g. involvement in religion, family relations, social acceptability of alcohol, poverty, and so on), and the effect sizes of these vary from adolescence to adulthood. Kendler and colleagues in a landmark study of Virginia male twins demonstrated variation at different time points for psychoactive substance use. They showed that at age 14 years, genetic factors have no influence on the level of alcohol consumption. From ages 15 to 23 years, we see a steady rise in the importance of genetic influences, which reach a heritability level of approximately 40% through age 35 years with the exception of a “dip” in the mid-30s that reverses itself by age 40 years. For familial environment, at age 14 years, the familial resemblance for alcohol consumption results entirely from familial environmental influences, accounting for 40% of the variance. The importance of the familial environment declines steadily until age 23 years. From then until age 35 years, familial environmental influences on alcohol consumption are modest except for a brief “burst” in the mid-30s that disappears by age 40 years. (Kendler et al. 2008). Furthermore, Kendler’s group reported that age at first drink appears to be associated with alcohol-related behaviour, that the effect of early exposure is genetically mediated, and that progression to alcoholism is under stronger genetic control than initiation (Kendler et al. 2011).

While quantitative genetic studies implicate genetic factors in the development of alcoholism, their findings also indicate that the genetic factors are not deterministic. If certain polymorphisms were highly penetrant then all alcoholics would have a genetically close relative with alcoholism; the concordance rates for monozygotic twins would approach 100%; and almost all offspring of an alcoholic parent—including adoptees—would eventually develop an alcohol use disorder. Since this clearly is not the case, environmental and other biological factors must also play an important role in alcoholism susceptibility.

Many researchers have noted the importance of the interaction between an inherited biological vulnerability and environmental risk factors for developing alcohol use disorders (Kendler et al. 1995; Slutske et al. 1998; Cadoret et al. 1995). This assumes a synergy between genetic and environment factors that may contribute either to an increased susceptibility for developing an alcohol use disorder or attenuate possible genetic risk by producing a level of protection for vulnerable individuals, and is termed the gene-environment interaction (G x E) model (Dick et al. 2006). A striking example of gene x environment interaction are the Sigvardsson Swedish adoption studies, referenced above, which showed that high genetic risk alone is insufficient to activate expression of the alcoholism phenotype. Environmental effects (for example, coming from parenting style, adverse childhood experiences, etc) in children with genetic loading (as defined by a having a first degree relative with alcoholism) are necessary to develop alcoholism. This work suggests that without the genetic risk factors, the phenotype will not present any more than the population rate, despite adverse life experiences usually associated with developing addictions.

I have thus far focused on genetic risk factors, as the building blocks of biological susceptibility. While again not within the scope of this work, I have commented briefly on epigenetic factors as an important concept in many psychiatric disorders.

3.8 A NOTE ON EPIGENETICS

In addition to environmental effects, recent work has highlighted an additional role for epigenetic processes in mediating susceptibility to major mental illness. Epigenetics refers to heritable changes in gene expression that do not involve changes to the underlying DNA sequence; a change in phenotype without a change in genotype. Epigenetic change is a common and natural process and can also be influenced by several factors including age, environment, nutrition, substance use, and disease state. I will briefly mention a relevant example.

Since monozygotic (MZ) twins share a common DNA sequence, their study represents an ideal design for investigating the contribution of epigenetic factors to disease aetiology. Georgiades and colleagues (2011) performed an innovative genome-wide analysis of DNA methylation on peripheral blood DNA samples obtained from a unique sample of MZ twin pairs discordant for major psychosis. Numerous loci demonstrated disease-associated DNA methylation differences between twins discordant for SZ and BD individually, and together as a combined major psychosis group. Pathway analysis of their top loci highlighted a significant enrichment of epigenetic changes in biological networks and pathways directly relevant to psychiatric disorder and neurodevelopment. Their top psychosis-associated, differentially methylated region, which was significantly hypomethylated in affected twins, was located in the promoter of *ST6GALNAC1*, an area overlapping a previously reported rare genomic duplication observed in SZ. The mean DNA methylation difference at this locus was 6%, but there was considerable heterogeneity between families, with some twin pairs showing a 20% difference in methylation (Georgiades et al. 2011).

In the next chapter I approach the commonest method for tackling AD genetics in the 1990s-2000s: linkage studies.

CHAPTER 4: MAPPING AD & ALCOHOLISM: LINKAGE STUDIES

This chapter aims to summarize the state of knowledge of linkage studies with reference to “alcoholism”.

4.1 GENETIC MARKERS

The genome among humans is approximately 99.5% identical, with the remaining difference among people due to polymorphisms and sequence variations – and a key mission in genetic research is to differentiate normal variation from disease-related genetic change. Genetic markers are an essential component used in any disease gene mapping study. Single nucleotide polymorphisms (SNPs) are most commonly used at present. SNPs are bi-allelic markers that are densely spread throughout the human genome, occurring approximately once every 1000 base pairs (bp) (Levy et al. 2007). SNPs have been widely used for fine mapping of positional and functional candidate genes. Sets of over a million of SNPs on a single array chip have been used for GWAS to screen the whole human genome and for phenotypes of interest. Copy number variations (CNVs) are being increasingly recognized and identified as both genetic markers and in psychiatric illness. I discuss CNVs in more detail for BP in Chapter 6 and in SZ in Chapter 7.

Other markers now used less commonly include Short tandem repeats (STRs) or microsatellites, which are short di-, tri-, or tetra-nucleotide tandem repeat sequences widely distributed in the genome, and are highly polymorphic among individuals. Microsatellites were initially described by Weber and May (Weber & May 1989) and Litt and Luty (Litt & Luty 1989), and have traditionally been the most utilized markers for linkage studies although latterly SNPs are also being used (Venken et al. 2008; Han et al. 2013). SNPs have the disadvantage of being bi-allelic and are thus not as highly polymorphic as STRs. However, they represent the most frequent type of polymorphism and their detection can be automated (Wang 1998).

4.2 LINKAGE STUDIES

Genetic linkage analysis is a useful technique to detect the chromosomal location of disease –related genes. It is based on the observation that genes that reside physically close on a chromosome tend to remain linked during meiosis i.e. that traits did not assort or segregate independently, but that traits encoded by these linked genes were inherited together.

A measure for the likelihood of linkage is the logarithm of the odds (lod) score. The lod score Z is the logarithm of the odds that the loci are linked divided by the odds that the loci are unlinked (Morton 1955). Expression of the likelihood as a logarithm allows summation of likelihoods observed in different pedigrees. Since the true genetic distance between two loci is frequently unknown, the lod score is calculated for different recombination fractions providing a maximum likelihood estimate for the recombination fraction (θ_{max}) at which the greatest lod score (Z_{max}) is observed.

A lod score higher than 3.0 is generally accepted as evidence for linkage, when the human genome is screened with 100 to 300 markers. A lod score lower than -2 is accepted as evidence against linkage. Lod score analysis requires the assumption of precise genetic models, including penetrance, disease gene frequency, and the clear classification of individuals as affected or unaffected.

Linkage studies are a type of genome wide analysis used to identify specific chromosomal regions implicated in a disease or trait of interest (Pulst 1999).

4.2.1 Advantages and disadvantages of linkage studies.

Linkage studies are useful for studying multiple markers simultaneously and for localising areas of disease risk across the genome. Disadvantages include needing a large number of families and family members containing several affected generations. If the disease is of late-onset with a high mortality,

finding families with more than one affected generation will be difficult. Linkage studies are considered less helpful for complex traits, such as diabetes where multiple genes are important in disease causation. For polygenic disorders it is highly unlikely that every family will have the same collection of underlying genetic factors and so linkage results will conflict and the LOD scores will be reduced.

4.3 LINKAGE STUDY FINDINGS IN AD

Genetic linkage studies on large samples of families have been carried out by several research groups in European and Native American populations (Cook et al. 1996; Guerrini et al. 2005; Kendler et al. 2006; Kuo et al. 2006b; Hesselbrock et al. 2001; Ehlers et al. 2004). The most important have been those from the USA Consortium on the Genetics of Alcoholism (COGA) (Foroud et al. 2000). For example a lod above 3.00 for linkage with alcohol dependence syndrome (ADS) was found on chromosome 3. Another lod above 3.00 with ADS was found on chromosome 6 (Cantor and Lanning 1999). Lods above 3.00 were found with alcoholism on chromosome 11 and 4 in Native American families (Long et al. 1998). Some of these positive findings were replicated in a linkage study in Ireland (Prescott et al. 2006).

In the USA COGA study a combined ADS and conduct disorder (Dick et al. 2004) phenotype showed positive lod scores over a region of chromosome 2 that includes the Tachykinin A Receptor 1 (TACR1) locus. The TACR1 locus was later reported to be associated with bipolar disorder and attention deficit hyperactivity disorder (ADHD) (Perlis et al. 2008; Ferreira et al. 2008; Yan et al. 2009; Sharp et al. 2014). Lod scores greater than 3.0 were found on chromosomes 7, 2, 3, 5, 9, and 14 when antisocial families were selected to achieve greater homogeneity (Jacobson et al. 2008). Replication of the linkage to alcoholism on chromosome 1 originally found by the USA COGA study (Dick et al. 2002) has been achieved in Finland (Lappalainen et al. 2004) with further support from a linkage study in the UK (Guerrini et al. 2005). This locus appears to be showing linkage to a combined depression and alcoholism phenotype (Nurnberger et al. 2001).

Several studies have reported evidence for linkage of substance dependence to a broad area on distal chromosome 1 (Reich et al. 1998a; Foroud et al. 2000; Nurnberger Jr. et al. 2001; Dick et al. 2002; Hill et al. 2004; Lappalainen et al. 2004; Guerrini et al. 2005; Ehlers & Wilhelmsen 2006).

A site for AD on chromosome 4 near a GABAA receptor gene cluster was found by the COGA team (Reich et al. 1998b and Long & Goldman 1998) in a southwest Indian tribe. This chromosomal region, which also contains genes coding for receptor subunits GABRG1, GABRA2, GABRA4, and GABRB1, was also linked to an electrophysiological phenotype (EEG beta activity) that characterizes individuals at risk for alcoholism in the COGA study (Porjesz et al. 2002), heavy drinking in the Framingham study (Wyszynski et al. 2003), cannabis dependence in the nicotine addiction genetics project (Agrawal et al. 2008), and “addiction phenotype” haplotypes in substance misusers (Drgon et al. 2006). A site on chromosome 7 has also been found by the COGA study but is yet to be duplicated (see Saccone & Rice 2005; Dick et al. 2008).

More recently, Han and colleagues performed ordered subset linkage analysis in 384 African American families using “admixture proportion” as a covariate to identify a more homogeneous subset of families and determine whether there is increased evidence for linkage with AD. Statistically significant increases in lod scores in subsets relative to the overall sample were identified on chromosomes 4, 12, 15 and 22. In a subset of 44 families with African ancestry proportions ranging from 0.858 to 0.996, Han et al observed a genomewide significant linkage at 180 cM on chromosome 4 (lod = 4.24) near a candidate gene GLRA3, which encodes a subunit of the glycine neurotransmitter receptor (Han et al. 2013).

Pertinent linkage findings are summarized in Table 4.1 which is adapted from Gizer’s summary (Gizer et al. 2011).

In summary, linkage approaches have found replicated evidence for numerous regions linked to AD and related phenotypes in different ancestral groups. The next stage is to map these regions more finely using association studies, which is the theme of Chapter 5.

Table 4.1 Chromosomal Regions with Evidence of Linkage to AD and related phenotypes (adapted from Gizer et al. 2011).

Phenotype	Chromosomal Region	cM	LOD	Nearest Markers	Previous evidence for linkage to alcohol dependence (within 40 cM)
Alcohol Dependence with clustering	1p36.31–p36.22	11	1.4	D1S214/D1S450	-
	2q37.3	287	1.2	D2S140	Prescott et al. 2006 (265 cM) ^{*,2} ; Kuo et al. 2005 (250 cM) ^{*,3}
		180	4.24	-	Han 2013
	8q24.3	163	1.5	D8S1743/D8S1836	-
	10q23.3 – 24.1	172	3.32	-	Gelernter 2009
	18p11.21–p11.2	48	1.3	D18S453/D18S1107	Kuo et al. 2005 (83 cM) ⁴ ; Schuckit et al. 2005 (21 cM) ⁵ ; Hill et al. 2004 (41 cM) ¹ ; Wilhelmsen et al. 2003 (70 cM) ⁵
Alcohol Dependence without clustering	1p22.3	114	1.2	D1S2766	Reich et al. 1998 (121 cM) ^{*,1} ; Corbett et al. 2005 (99cM) ^{*,1} ; Foroud et al. 2000 (120 cM) ^{*,1} ; Guerrini et al. 2005 (121 cM) ¹ ; Kuo et al. 2006 (141 cM) ⁶ ; Hill et al. 2004 (90 cM) ¹
	2p24.3–p24.1	32	2.4	D2S149/D2S305	Wilhelmsen et al. 2005 (43 cM) ⁷
	9p24.1–p23	18	1.7	D9S286/D9S168	Long et al. 1998 (42 cM) ¹
	22q12.3–q13.1	36	1.8	D22S283/D22S423	Prescott et al. 2006 (8 cM) ^{*,1} ; Kuo et al. 2005 (8 cM) ^{*,8} ; Bergen et al. 2003 (33 cM) ⁴ ; Foroud et al. 2000 (58 cM) ¹ ; Wilhelmsen et al. 2003 (25 cM) ⁵

Abbreviations: cM – centimorgans

Notes:

*studies indicated by an asterisk within a chromosomal region were conducted using overlapping datasets, numbered superscript indicate the trait studied:

¹Alcohol Dependence/Alcoholism, ²Alcohol Dependence symptom count, ³withdrawal symptoms, ⁴maximum drinks in 24 hours, ⁵body sway following exposure to alcohol, ⁶initial sensitivity to alcohol, ⁷average drinks/week, ⁸tolerance.

CHAPTER 5: MAPPING ALCOHOLISM: ASSOCIATION STUDIES

5.1 APPROACH

Whereas linkage analyses can identify a broad chromosomal region that is likely to contain a gene contributing to an endophenotype or disease, association studies can locate the gene or genes that influence an outcome with higher accuracy (Lewis & Knight 2012).

One of the most commonly used experimental designs to identify genetic variation contributing to a disease is that of candidate gene analysis. This approach seeks to test the association between a particular allele of the candidate gene and a specific outcome by comparing the frequencies of alleles in a population of unrelated cases (e.g. AD) and a population of control subjects (e.g. non-AD).

A candidate gene normally is chosen based in a biological hypothesis involving gene function or because the region has been implicated in prior linkage studies. Many candidate gene studies use population-based association methods which compare the genes of groups of people, cases and controls. Ideally, the two groups would be matched with respect to as many factors as possible (e.g. age and ethnicity) so that they differ only in disease status. The investigators would then compare the frequencies of various alleles of a marker (e.g. a SNP) within or near the candidate gene. Differences in allele frequencies between the two groups may indicate that the genetic variant is important in disease susceptibility. It may also represent population stratification, Type 1 error, or LD between the candidate gene and the true disease susceptibility gene, which is located within the chromosomal neighbourhood, but is discrete from the candidate gene. This means that the candidate gene is not actually the causative mutation, but is sufficiently close to the disease-causing mutation that the candidate gene and the disease mutation tend to be inherited together during gamete formation.

Population-based association studies of candidate genes have been widely used but the approach has drawbacks: there is a growing number of numerous biochemical pathways being categorized which means candidate genes numbers increase rapidly; in polygenic disorders such as AD, the effects of each gene are probably small and large sample sizes in multiple populations are often required to detect such genes. Spurious associations between the disorder and certain alleles and the disorder are likely to show up if the cases and controls are not well matched with respect to important genetically linked characteristics such as ethnicity.

Family-Based Association Studies (FBAT) were thus developed to address some of these pitfalls. A family-based association test—the transmission disequilibrium test (TDT) – was published by Spielman and Ewens (Spielman et al. 1993). The TDT tests the hypothesis that a particular marker allele is more frequently transmitted to affected offspring from heterozygous parents and analyses a nuclear trio consisting of an affected person and their parents. Using the information on the transmission of various alleles from many families, investigators can conduct statistical analyses to determine if a particular allele of the marker being tested is associated with disease development. Using a “within-family” design, the control sample of alleles is perfectly matched to the affected sample of alleles because both samples are transmitted from the same two parents, and no matched unrelated control sample is needed. Scientists have built upon the TDT using other designs, that include data from both affected and unaffected siblings or from even more extended pedigrees, and quantitative traits in addition to qualitative traits (Laird 2001).

5.2 CANDIDATE GENE STUDY FINDINGS IN AD

Numerous genes involved in the metabolism of alcohol, brain learning and reward systems, and ethanol's varied effects on the brain have been studied as candidate genes for AD. However association

studies have not often given consistent results. For example, I noted in the GABRB2 paper that there were 5 positive and 5 negative association studies of the gene.

According to the HUGENavigator, at time of writing, 724 genes have been implicated with the “alcoholism” phenotype (<https://phgkb.cdc.gov/HuGENavigator/home.do>). The following section offers a summary of selected positive and replicated associations with AD and related phenotypes.

5.2.1 The GABA system

The gamma aminobutyric (GABA) system has been implicated in alcoholism, substance misuse, externalizing behaviours and electroencephalogram (EEG) beta-activity. GABA is the principal inhibitory neurotransmitter in the central nervous system (CNS). A number of biological and genetic findings in human and animal models support GABA receptor genes as good candidates.

In rats, GABA-A antagonists decrease ethanol intake (Boyle et al. 1993; Nowak et al. 1998; Tomkins and Fletcher, 1996) and GABA agonists increase ethanol intake. GABA alpha receptor subtype 2 (GABRA2) has been associated with alcoholism and ethanol-conditioned taste aversion in rodents (Hood and Buck, 2000).

Linkage was found by Goldman and colleagues in a sample of US Indians for D4S3242 ($P = 0.0002$, $\text{lod} = 2.8$) which is located near GABRB1 (Long et al. 1998). The Collaborative Study on the Genetics of Alcoholism (COGA), using linkage scans implicated markers D4S3242, D4S2393 which are near GABRB1 on chromosome 4p. A linkage disequilibrium analysis of 69 SNPs across the GABA gene cluster in the family-based COGA sample found allelic and haplotypic association between alcohol dependence and GABRA2.

In humans, several alcohol effects, including anxiolysis, sedation, motor incoordination, some withdrawal symptoms, and ethanol preference, are mediated by GABA transmission (Buck, 1996; Grobin et al. 1998; Korpi et al. 1998). Cross tolerance is shown between alcohol and GABA-A receptor-modulating drugs like benzodiazepines which are used to treat alcohol withdrawal (Kumar et al. 2009).

GABAA receptors are involved in mediating both the acute and chronic effects of alcohol, and markers in GABA system genes have been associated in a number of studies with AD and other alcohol-related phenotypes (Cui et al. 2012), the most replicated being the GABAA receptor $\alpha 2$ subunit gene GABRA2. GABRA2 is thought to contribute to sensitivity to alcohol intoxication (Haughey et al. 2008; Pierucci-Lagha et al. 2005) and to a common genetic vulnerability to alcohol and drug misuse (Agrawal et al. 2006). The SNP rs279858, though lying in exon 5 of the gene, is a synonymous substitution (Covault et al. 2004). It has been associated, either individually or as a member of multi-SNP haplotypes, with AD and other alcohol related phenotypes, such as alcohol sensitivity, more often than any other marker.

Reanalysis of the COGA sample revealed that the association between GABRA2 and AD was strongly dependent on illicit drug dependence (Agrawal et al. 2006). The opposite effect was noted in the Covault study (Covault et al. 2004) where removal of drug dependent subjects increased the association between GABRA2 and AD. GABRA2 may also contribute to alcoholism through effects on psychiatric co-morbidity (Dick et al. 2006; Enoch et al. 2006) although other studies contradict this: one study found a strong GABRA2 association in an AD sample without depression (Covault et al. 2004) while another observed no association in a sample with low psychiatric and substance misuse co-morbidity (Matthews et al. 2007).

Bierut et al's 2010 AD GWA of predominantly European- and African-Americans also evaluated GABRA2 independently as a candidate gene. Five SNPs in this GWA at GABRA2 yielded a nominal (uncorrected) $P < 0.05$ (Bierut et al. 2010). GABRA2 and GABRA6 have, separately, also been associated with alcoholic Korsakoff's psychosis (Loh et al. 1999).

At the time of writing (final literature search 1 November 2015) five association studies have implicated GABRA2 in alcoholism samples (Bierut et al. 2010; Covault et al. 2004; Fehr et al. 2006; Lappalainen et al. 2005; Soyka et al. 2008) while five studies, excluding ours, have found no association (Drgon et al. 2006; Enoch et al. 2010; Lind et al. 2008; Matthews et al. 2007; Onori et al.).

To reconcile these conflicting data, Li and colleagues performed a meta-analysis of variants in the GABAA receptor genes (*GABRB2*, *GABRA6*, *GABRA1*, and *GABRG2* on chromosome 5q and *GABRA2* on chromosome 4p12). Using a Bonferroni-corrected threshold of 0.007, they found strong associations between *GABRA2* and AD, rs567926, and rs279858; and between *GABRG2* and AD and heroin dependence. Significant association was also observed between *GABRA6* rs3219151 and AD (Li et al. 2014).

Kosobud et al recently reported an association between the GABRA2 SNP rs279858 and higher recent drinking history ($p = 0.01$) in a group of non-AD moderate to heavy drinkers. A recent drinking history \times genotype interaction ($p = 0.01$) were significantly associated with acute adaptation of the subjective responses to alcohol. Higher recent drinking was found to be associated with reduced acute tolerance to the stimulating effects of alcohol in carriers of the rs279858 risk allele. The authors postulated that GABRA2 effect on alcohol dependence may, in part, be due to its effect on subjective responses to alcohol (Kosobud et al. 2015).

5.2.2 Alcohol metabolism genes

In *Homo sapiens*, alcohol is metabolized by the liver enzymes alcohol dehydrogenase (ADH) into acetaldehyde, and aldehyde dehydrogenase (ALDH) into acetate, at the rate of about one standard drink per hour (Li 2000). At the usual blood alcohol concentrations, the acetaldehyde is then quickly broken down by the mitochondrial form of the enzyme aldehyde dehydrogenase (ALDH2). Genes coding for these enzymes are thus of great interest as candidate genes for AD.

Genetic polymorphisms in the genes encoding ADH and ALDH are associated with alterations in enzyme kinetics, which determine the rates of manufacture and elimination of the toxic intermediate metabolite acetaldehyde. This has led to the widely accepted hypothesis that genetic factors which alter the metabolism of ethanol are important aetiological risk factors for ADS (Quertemont 2004; Schuckit et al. 1996; Whitfield et al. 2001).

Genetic variation among alcohol-metabolizing genes has been well studied with respect to their role in affecting predisposition to alcohol dependence (Way et al. 2015). The ALDH2*2 functional variant in aldehyde dehydrogenase type 2 (ALDH2), predominantly found among Asian populations, produces a reduced capacity to metabolize acetaldehyde and a physiologic flushing response, nausea and tachycardia are may create an aversion to alcohol consumption.

The Australian Alcohol Challenge Twin Study, started over 20 years ago, has provided impressive contributions to understanding the genetics of alcohol metabolism in relation to alcohol use disorders, such as pharmacokinetic measures, alcohol consumption, and heritability of various alcohol-related traits (Martin, Perl, et al. 1985; Martin, Oakeshott, et al. 1985). Studies have been published on genetic influences on peak blood alcohol concentrations, rate of decrease in blood alcohol concentration, and AD. (Whitfield & Martin 1994). Whitfield et al observed a two- to threefold increased AD risk in individuals who demonstrated blood or breath alcohol concentrations after a challenge in the highest quartile of values compared with those in the lowest at 10 year follow-up (Whitfield et al. 2001). These

adverse effects significantly affect drinking behaviour; in doing so they protect against the development of alcohol misuse/dependence and the medical consequences of harmful drinking, including alcohol-related cirrhosis and pancreatitis (Cichoz-Lach et al. 2007).

The involvement of the alcohol dehydrogenase (ADH) gene cluster has been reinforced by linkage studies repeatedly implicating chromosome 4 (Luczak et al. 2006; Edenberg et al. 2006) and by several allelic association studies (Kuo et al. 2008; Edenberg et al. 2006; Ball 2005). In a GWAS of German men, only one SNP, rs1789891 (located between the ADH1B and ADH1C genes), achieved genome-wide significance for AD (Frank et al. 2012). The major allele of rs1789891 was found, based on imputed genotypes, to be in LD ($D'=1.0$, $r^2=0.27$) with the major allele of rs1693482, encoding the high activity (Arg272) variant in ADH1C (Edenberg 2012; Lee 2014). However, comparisons with the results of three other GWAS undertaken in individuals of European origin or ancestry have produced conflicting results (Macgregor et al. 2008; Frank et al. 2012; Edenberg et al. 2010b). Thus, in one GWAS rs1789891 showed nominal but not genome-wide significance with ADS (Edenberg et al. 2010b) while no significant association was found in the other two studies. Further, conditional analysis of the German GWAS data indicated that partly independent signals from other markers also contributed to the observed associations supporting the role of other variants in the ADH cluster affecting ADS risk.

Recently, Holmes and colleagues reported on a protective association of the ADH1B rs1229984 A-allele variant in reducing weekly alcohol intake, and better cardiovascular measures, suggesting that reducing alcohol even for light to moderate drinkers is beneficial for cardiovascular health (Holmes et al. 2014).

In 2015 I was part of the UCL group which found at least two genetic effects from the ADH cluster on chromosome 4 which impact on ADS risk in a British and Irish population. Way and colleagues from our group observed a highly significant protective association with the rare missense Arg48His mutation, rs1229984, in ADH1B and AD risk (allelic P 8.4×10^{-6} , odds ratio [OR] 0.26, 95% confidence interval [CI] 0.14, 0.49). A significant independent association with AD risk was also observed with the more

common intergenic variant, rs1789891, (allelic P 7.2×10^{-5} , OR 1.4, 95% CI 1.2, 1.6) and with three non-synonymous SNPs in ADH1C, rs698, rs1693482, rs283413 and ADS risk although their effects were not completely independent from those of rs1229984 and rs1789891 (Way et al. 2015).

5.2.3 The Dopamine system and COMT

Dopamine (DA) signals via D1-like (D1, D5) and D2-like (D2, D3, D4) receptors, which have opposite effects at the cellular level, respectively stimulating or inhibiting, adenylate cyclase (Caspi 2002) (Girault & Greengard 2004). Thus, by differentially affecting D1-like versus D2-like receptors, DA can activate or inhibit regional brain activity. Alcohol and drugs of abuse, which stimulate brain DA signalling (Koob & Bloom 1988), could therefore result in inhibition or activation of regions of the mesolimbic and mesocortical DA pathways, depending on the balance of D1-like versus D2-like receptors. This in turn enhances neurotransmission in the nucleus accumbens (Di Chiara & Imperato 1988; Weiss et al. 1993; Anon n.d.; Tanda et al. 1997) and affects the motivational and reward properties of addictive drugs (Koob 1997; Tanda et al. 1997). Deficiencies in dopaminergic tone have also been linked to addictions and termed the “reward deficiency syndrome” (Koob & Volkow 2010) .

The Taq1A polymorphism (located in the gene cluster ANKK1/DRD2) in AD has been the focus of much research and AD. In 1990 Blum and colleagues first reported that the DRD2 receptor A1 allele was associated with an OR of 8.7 for developing AD compared to controls (Blum et al. 1990), a replicated association. Others have shown that gene polymorphisms that inhibit the expression of the D2 receptor are associated with increased risk of AD (Berggren, 2006; Krachewski, 2009). In support of this hypothesis is a study which found increased alcohol intake among D2 receptor knock-out mice (Bulwa et al. 2011). In contrast, other studies failed to find any association between the D2 receptor and AD (Gelernter & Kranzler 1999; Blomqvist et al. 2000). A large-scale meta-analysis confirmed both allelic and genotypic associations between the Taq1A polymorphism and the risk for AD in over 18,000 subjects included in 61 case–control studies that were published up to August 2012 (Wang, 2013).

Dopamine is linked to cravings (see endophenotypes, Section 2.7.8 above). Reduced prefrontal cortical dopamine transmission has been reported in AD, among many other neurobiological findings involving dopamine (Narendran et al. 2014). Dopamine is also linked to the opioid system (see below) and many other addictions- and psychiatric disorder-related brain circuits beyond the scope of this thesis.

Dopamine metabolism is also of interest in AD risk. Catechol-O-methyl transferase (COMT) is an enzyme involved in the breakdown of dopamine (Axelrod & Tomchick 1958) and is encoded by the *COMT* gene, located on chromosome 22q11 - a region that has been strongly implicated in schizophrenia, anxiety and mood disorders (Bassett & Chow 1999; Bassett et al. 1998; Gothelf et al. 1997; Abdolmaleky et al. 2006; de Chaldée et al. 1999; Egan et al. 2004; Goghari & Sponheim 2008; Kremer et al. 2003; Li et al. 1996; Wonodi et al. 2003; Hetteema et al. 2008; Hamilton et al. 2002; Massat et al. 2005; Mandelli et al. 2007). Rs4680 is an intriguing COMT polymorphism that encodes either methionine or valine at amino acid position 158, which modifies COMT enzyme activity (Chen et al. 2004). Other *COMT* SNPs, rs165774 and rs6267 have also been associated with SZ (Hong et al. 2008; Lee et al. 2005). In addictive behaviours, rs4680 has been found to contribute to the development of AD (Voisey et al. 2011), late-onset alcohol dependence (Tiihonen et al. 1999), alcohol intake in alcoholics and the general male population (Kauhanen et al. 2000), smoking cessation, (Omidvar et al. 2009), and opiate dependence (Oosterhuis et al. 2008). However there have been conflicting studies for *COMT* including a study that did not observe association with alcohol dependence or nicotine dependence (Foroud et al. 2007). Genome-wide association studies of nicotine and alcohol dependence have not as yet found association with *COMT* (Caporaso et al. 2009; Edenberg et al. 2010a; Lind et al. 2010; Treutlein et al. 2009).

5.2.4 The Serotonin system

Serotonin (5-HT) is a key neurotransmitter in the regulation of reward-related behaviours, from eating and drinking to sexual activity, and is strongly implicated in mood and anxiety disorders. Serotonin-containing neurons project diffusely to various areas known to be involved in the reward system, including the nucleus accumbens, ventral tegmental area, substantia nigra, hippocampus, amygdala, and prefrontal cortex. Nearly all of the serotonin receptor subtypes have been linked to areas in the brain associated with addiction and reward (Hayes & Greenshaw 2011).

Presynaptic serotonin transporter (5-HTT) polymorphisms have been investigated as an important regulator of addiction risk and drinking behaviour (Heinz et al. 2000). The serotonin transporter, encoded by the SLC6A4 gene, influences the synaptic actions of serotonin and is responsive to stress hormones. The 5-HTT-linked polymorphic region [5'-HTTLPR; long (L) and short (S) alleles] is a functional SLC6A4 promoter polymorphism and widely studied. Ait-Daoud et al have shown that cue-induced alcohol craving and propensity for higher drinking are modulated by allelic differences in SLC6A4 and by the serotonin transporter gene polymorphism rs1042173 (Ait-Daoud et al. 2012). The same authors reported on genotype combinations that may predict the outcome of the treatment of alcohol dependence using the 5-HT(3) antagonist ondansetron (Johnson et al. 2013), and also interactive effects among variants in SLC6A4-HTR3A-HTR3B affecting AD risk (Seneviratne et al. 2013). The serotonin receptor genes HTR2A, HTR1A and TPH2 have also been associated with suicide attempts in alcohol dependence (Wrzosek et al. 2011).

5.2.5 The opioid system

The opioid system is composed of numerous receptors expressed mainly in the central nervous system. Three types of opioid receptors (mu, delta and kappa) represent the respective targets of the major opioid peptides (β -endorphin, enkephalins and dynorphins, respectively). As discussed above, the mesolimbic dopamine “reward” system ascends from the ventral tegmentum of the midbrain (VTA) to the nucleus

accumbens (NAC). One model proposes that reward results from activation of mu-opioid receptors in the VTA and/or delta receptors in the NAC; both these nuclei are targets of endogenous β -endorphin. In contrast, activation of κ -receptors appears to cause dysphoria.

Models of high preference for alcohol also support the view that alcohol induces opioidergic activity (and thus reward) and/or that ethanol intake relates to the tone of the endogenous opioid reward system – so that alcohol consumption may serve to compensate for inherent deficits in this system (Giannoulakis 1996). One biomarker study found lower cerebrospinal fluid (CSF) levels of β -endorphin in AD subjects than those of the controls (Genezzani, 1981). Whether this is causal or effect is unclear. It is suggested that alcohol itself interferes with this reward pathway either directly or indirectly by affecting binding properties of opioid receptors, as well as modulation of opioid peptide synthesis and secretion (Herz, 1997).

Opioid polymorphisms have been linked to including alcohol use (Bart et al. 2005; Francès et al. 2015), alcohol cravings (van den Wildenberg et al. 2007), and response to naltrexone (a mu opioid receptor antagonist and a medication used in relapse prevention for AD) (Chamorro et al. 2012; Rösner et al. 2010). For example, upon treatment with naltrexone, 87% of rs1799971(G) carriers had fewer heavy drinking days and more days abstinent, than 55% of individuals with the (A;A) genotype (Anton et al. 2008). This SNP may also influence the response (analgesia, nausea and vomiting) to opioids such as morphine, codeine or heroin. A 2015 meta-analysis (totalling 5,902 patients) concluded that the carriers of a rs1799971(G) allele consumed more opioids for analgesia (Lu 2015).

Opioids themselves are potentially addictive in susceptible individuals, and cross-addiction between opioids, alcohol and other substances is well described clinically (Laudet, 2008; Langas, 2011). Several studies have found mixed evidence for an association between receptor polymorphisms and addiction vulnerability (e.g Hoehe 2000; Crowley, 2003). For example, a study of 200 Chinese heroin dependent people reported a higher frequency of the rs1799971(G) allele compared to non-dependent people (40%

vs 29%) (Szeto, 2013), but this was not replicated in another study of the same population (Shi, 2002). Tentative new point mutations have recently been identified in an Iranian opioid-dependent subjects (Dinarvand, 2014).

In summary, the opioid system is closely linked to the reward system, response to alcohol, and risk of AD and related phenotypes including other addictions.

5.2.6 Taste receptors

The flavours in ethanol have sweet and bitter taste components. Variation in bitter and sweet taste receptors is associated with perception of ethanol flavour and consumption of alcoholic drinks as well as food preference (Bachmanov et al 2003). The 25 human genes coding for bitter taste receptors (TAS2Rs) contain high levels of allelic diversity. Since bitter taste is hypothesized to reduce the intake of toxic substances, variation at TAS2R genes could reflect the action of natural selection on the ability to recognize some bitter compounds rather than others. Whether species-specific variation in TAS2R genes is solely the result of genetic drift or whether it may have been influenced by selection due to different feeding behaviours is a matter of debate (Risso et al. 2014).

The most studied taste endophenotype is propylthiouracil (PTC) bitterness, which is mediated primarily by the TAS2R38 gene. In a laboratory-based study, TAS2R polymorphisms appeared to influence the intake of alcohol (Hayes et al 2010). Individuals carrying one or two sensitive alleles of TAS2R38 reported lower yearly consumption of alcoholic beverages than did individuals homozygous for the insensitive allele (Duffy et al 2004). An association has been found between AD and polymorphisms in TAS2R16 (another bitter taste receptor): the less sensitive K172N allele is appears to heighten AD risk (Hinrichs et al 2006). This allele on Chr 7q, a region implicated in ADS by multiple linkage studies (see Linkage section above). However, this risk allele is uncommon in European Americans (minor-allele frequency [MAF] 0.6%), whereas 45% of African Americans carry the allele (MAF 26%), which arguably makes it a more significant risk factor in the African American population.

Bitter taste receptor gene polymorphisms are also an important factor in the development of nicotine dependence (ND) in African Americans in whom reduced bitterness sensitivity represents a risk factor for ND, and increased oral sensitivity confers protection against ND (Mangold, 2008).

5.2.7 The endocannabinoid system

Since the first protean sea slug evolved a cannabis receptor (CB1), the cannabinoid system has been found in many descended animals including humans, who discovered the mind altering properties of herbal CNR1 agonists some 12,000 years ago (McPartland, 2002). In humans the genes coding for cannabinoid receptors are termed CNR1 and CNR2. The endocannabinoid system is the main site of action for the rewarding and pharmacological responses induced by cannabinoids (Ledent, 1999; Lichtman, 2005). However, this system plays an overall modulatory effect on the reward circuitry and also participates in the rewarding and addictive properties of most typical drugs of abuse (Maldonado, 2008).

Alcohol and cannabinoids appear to activate similar reward pathways, and CB1 receptors also seem to regulate the reinforcing properties of alcohol in rodent models although human data are limited. In a study of 550 AD or SUD subjects, Zuo et al found that the risk for each SUD subtype significantly increased with the number of “G” alleles at rs6454674 in European Americans (Zuo, 2012). One recent postmortem brain study of AD subjects demonstrated hyperfunctional CB1 receptors in the caudate nucleus compared to the non-AD groups; and non-suicidal AD subjects showed hypofunctional receptors in the cerebellum. No changes were found in the CB1 protein expression in either region (Erdozain, 2015).

5.2.8 Miscellaneous and meta-analyses

Many small and medium scale case control genetic association studies have already shown that genetic effects on susceptibility to alcoholism are partly derived from neurotransmitter receptor variants (Kohnke, 2008), from alcoholism metabolism genes (Dick et al. 2007, Kuo et al. 2008, Luo et al. 2005)

and from other central nervous system (CNS) related genes. A recent large meta-analysis found no significant associations for SNPs in six otherwise strongly implicated genes (BDNF, DRD1, DRD3, DRD4, GRIN2B and MAOA) (Forero et al. 2015). Carmiol et al. showed that the clock gene D-box binding protein (Dbp) appears to influence the risk for both bipolar disorder and AD (Carmiol et al. 2014).

Coded by Calcium/calmodulin-dependent protein kinase II (CAMK2A) the Alpha Ca(2+)/calmodulin-dependent kinase II (alphaCaMKII) enzyme plays a key role in controlling plasticity in the brain. It is the major synaptic protein in the forebrain and can switch into a state of autonomous activity upon autophosphorylation. It has been proposed that alphaCaMKII autophosphorylation mediates long-term memory (LTM) storage. The autophosphorylation of α CaMKII works as a 'molecular memory' for transient calcium activation, thereby accelerating learning (Irvine et al. 2006). Easton and colleagues in a series of mouse-human translational studies found seven significant associations between CAMK2A SNPs and alcohol dependence, one of which was in an autophosphorylation-related area of the gene (Easton et al. 2013). In another rodent-human study, Easton and colleagues confirmed that human carriers of the CAMK2A rs3776823 TT-allele carriers display a faster transition to severe cocaine use than C-allele carriers, and delayed conditioned place preference in mice given cocaine (Easton et al. 2014). They suggest that CAMK2A autophosphorylation as a facilitating mechanism in the establishment of alcohol drinking and cocaine use behaviour.

5.2.9 Summary

Numerous candidate genes have been proposed, and some fewer replicated, for development of AD or related traits. Biological AD risk appears to be mediated by taste, metabolism, reward, and neurotransmitter pathways.

5.3 GENOME WIDE ASSOCIATION STUDIES

In contrast with candidate gene studies, Genome-wide studies (GWAS) investigate single markers of common variation across the human genome, investigating both known genes and the bulk of the non-genic regions, whether they were previously associated with pathophysiological hypotheses or not, and without needing to understand the functions of associated genes. A number of catalogues of published GWAS are available (for example at <http://www.ebi.ac.uk/gwas/>). Due to their extensive coverage of the human genome, GWAS, in combination with systems biology approaches, have yielded wide-ranging data about complex disorders psychiatric like AD, BP and SZ amongst others. According to the NHGRI-EBI Catalog of published genome-wide association studies (<http://www.ebi.ac.uk/gwas/>) as of 20/11/2015, 183 genes were reported associated to SZ (139), BP (36) and AD (8) at genome-wide significant levels ($p < 5 \times 10^{-8}$) and many of these associations have been replicated or are replications. Additionally another 519 genes were implicated at suggestive p values ($p < 5 \times 10^{-6}$).

5.3.1 GWAS methodology

GWAS are based on the theory of linkage disequilibrium (LD), which is defined as a non-random statistical association of alleles at two or more loci, and which is characteristically associated with short physical distance between genetic markers. A GWAS involves a comprehensive interrogation of the common variations across the entire human genome. In a GWAS, researchers search for statistically significant differences in the frequency of genetic variants between cases and controls at hundreds of thousands (or more) loci (McCarthy et al. 2008). This allows us to identify genetic associations with observable traits, without the need for an *a priori* hypothesis about the location or function of the variants. Due to the extensive correlation or LD between SNPs, the majority of all common genetic variation is captured by genotyping only a small proportion of the total number of known SNPs using this technique. The genetic architecture of psychiatric disorders has proven to be complex resulting in two contrasting hypotheses: the Common Disease Common Variant (CDCV) and Common Disease Rare Variant (CDRV) theories.

According to CDCV, the genetic risk in an individual and in the population is attributable to many high-frequency variants, each conferring modest level of risk (Risch & Merikangas 1996). Alternately, the CDRV hypothesis states that a large number of rare variants with relatively large effects contribute to the disease risk, such that hundreds of genes, each with many rare variants, contribute to the phenotype (Pritchard 2001; Reich & Lander 2001). Based on the concept that psychiatric disorders have a negative effect on fecundity, natural selection would prevent the causal alleles from reaching common frequencies in a population. This assumption would support the CDRV hypothesis (Alaerts & Del-Favero 2009).

GWAS combines the merits of association studies in the power to detect small effects with that of linkage studies because they do not require any knowledge of the disease pathogenesis. In addition, the GWAS approach requires correction for extensive multiple testing and for population stratification. The large samples required for performing GWAS provide power to detect variants with small or large effects.

GWAS was largely made possible by the Human Genome Project (www.ornl.gov/sci/techresources/Human_Genome/home.shtml). Advances in SNP genotyping and DNA sequencing were offshoots of the Human Genome Project, and microarrays made possible rapid and accurate genome-wide genotyping resulting in a map of common genetic variation in a reference set of individuals of European, Asian, and African descent: This map was called the HapMap project (an abbreviation of “haplotype mapping”). HapMap defined “blocks” of the human genome which were relatively free from the linkage. Such blocks were tagged with polymorphic genetic markers – hundreds of thousands of them across the genome. This massive project was followed by translational activities by private sector firms (e.g. Affymetrix Inc. and Illumina Inc.). This output resulted in the introduction of genome-wide scans based on hundreds of thousands of SNPs. These scans were based on common genetic variants in human populations, in which the frequency of minor alleles is still greater than 5% in

the population. The common variants were specifically chosen for scans to "tag" the "blocks" of the genome separated by "recombination hot-spots" (Wellcome Trust Case Control Consortium 2007).

Because most of the markers used for GWAS are tag SNPs, the most significant associated SNP may reflect a direct or an indirect association. The SNP may thus be in LD with a causative variant, and LD can occur hundreds of kb-pair away from the marker. The most commonly used criterion for SNP chip evaluation is global coverage, defined as the fraction of common SNPs that are tagged by the SNPs. For example the Affymetrix 6.0 and Illumina 1M SNP arrays include ~1M common SNPs and probes for analysis of copy number variants (CNVs) reportedly assay some 80% of the common variation in the genome (Li et al. 2008).

The estimated number of common (minor allele frequency, MAF>1%) SNPs is ~10M, but our genotyping capabilities are not (yet) sufficiently developed to genotype every SNP in huge clinical samples (deep re-sequencing technology and new arrays may soon overcome this difficulty). In the meantime, imputation, the computational prediction of genotypes from non-genotyped SNPs, is used to extend GWAS map coverage.

Recent complex disorders GWAS show two main characteristics: First, common loci with small effects are typically reported (ORs=1.1–1.5). This may be an empirical confirmation that a large body of epidemiological studies predicting multiple small common genetic effects for complex disorders were correct since loci with larger effects are rapidly eliminated from the population through natural selection. Second, most studies have tended to detect new susceptibility loci, and only very large samples obtained from combining studies are powered to show robust replication. This is because the power to detect one out of many possible risk loci is much larger than the power to detect specific disorder alleles 21. Furthermore, if only small effects are found, many genes would be predicted to underlie the pathophysiology of most complex genetic disorders.

Meta-analyses of bipolar disorder GWAS (Chen et al. 2011) can focus on lower- or higher- status signals which may suggest risk loci, rather than on the usual 'top hits' of individual studies. Locus and/or allelic heterogeneity will continue to play an important role in diluting true signals, even with larger samples (Pritchard and Cox 2002). A disadvantage of using large samples is that the LD variation between populations could be confounded with disease associations (Frazer et al. 2009). Population structure/stratification is routinely assessed and corrected for in GWAS, using various methods, including principal component analysis methods (Price et al. 2006), logistic regression, identical by state (IBS) clustering (Purcell et al. 2007), and linear mixed models (Alaerts and Del-Favero 2009).

Imputation methods are used in modern GWAS to improve the coverage of the human genome across different genotyping platforms. Imputation methods (Marchini et al. 2007; Purcell et al. 2007) attribute and test ungenotyped SNPs for association in a sample, thus, adding to the information content.

Analysis of GWAS data using more sophisticated statistical methods may provide additional power leading to updated results and a better understanding of the best ways for interpretation of GWAS data. One of the alternative methods previously employed in the analysis of GWAS data include multipoint methods combining information across markers with more powerful tests for indirect association with untyped variants. Other methods include simultaneous analysis of all SNPs to identify the subset that the best predictor for disease outcome, and weighted analyses incorporating prior information and pathway-based approaches, which jointly consider multiple variants in interacting or related genes (Alaerts and Del-Favero 2009).

5.3.2 Limitations of GWAS

While GWAS are useful for signalling candidate loci they are usually unable to ascertain disease causality or to differentiate between the effects of variants in LD. Additionally, some GWAS loci might contain multiple causative variants or genes that contribute to the overall disease susceptibility at a

single locus. However, the majority of current GWAS lack the statistical power to test whether multiple causative genes underlie the same locus. This has been addressed by Flister and colleagues in 2013 by testing multiple GWAS genes empirically, by using gene targeting in a disease-susceptible rat model of genetic hypertension (Flister et al. 2013). There are a number of further important limitations.

Firstly, the large number of statistical tests makes the method vulnerable to false positive hits (Type 1 error) (George 2013); therefore, the estimation of an appropriate genome-wide significance threshold is fundamental. The genome-wide significance threshold, for a value of 5% significance assuming tests for all common SNPs, has been estimated to be around $p < 5 \times 10^{-8}$ (Pearson & Manolio 2008). Secondly, the statistical power of GWAS to detect an association with rare alleles (i.e. SNPs or CNVs with MAF < 1%) is very limited (Lee et al. 2014). GWAS are designed to identify relatively common polymorphisms associated with the risk for a disease. Rare genetic variants with a potent influence on individual risk might thus be unnoticed by GWAS, but can be still detected by linkage studies (Samochowiec, 2014). For the detection of rare variants re-sequencing is considered more useful than GWAS (Owen, 2014). Third, the study of gene-gene interactions (epistasis) is strictly limited by the statistical power of currently existing samples contrasted to the large number of such tests – despite expectations that epistasis is a significant source of heritability (Hu et al. 2011). Fourthly, large data sets as found in GWAS bring statistical, computational and technical challenges, all of which can contribute to confounding. Factors that may contribute to confounding include: 1) with increased power to detect true effects comes increased potential for the introduction of bias to affect results; 2) non-random genotyping failure where a person's SNP genotype that is either not called or called incorrectly; 3) undocumented relatedness can be another source of bias, although modern whole-genome data analytic tools can be used to detect closely related people (Solovieff et al. 2013).

GWAS for complex disorders have demonstrated that a typical susceptibility locus has an odds ratio (OR) of c1.1 – 1.3, which often means extremely large samples are required for detection. For example,

the Molecular Genetics of Schizophrenia (MGS) EA sample ($N = 5,334$) has adequate statistical power only to detect very common risk alleles (with 30–60% frequency, log additive effects) with genotypic relative risks ~ 1.3 (Shi et al. 2009). To reach sufficient statistical power, the combined analysis of independent datasets is useful, as in the various collaborations including the Psychiatric Genetics Consortium (PGC) of which UCL is a small part. Collaborative joined up datasets have been reasonably successful in the search for a number of complex disorders, including SZ and BP, although the risk is that the diagnostic range is wider (Shi et al. 2009; Purcell et al. 2009; Stefansson et al. 2009).

A further issue posed by GWAS is due to finding of multiple extragenic or poorly characterized loci in strong genome-wide association. Assuming there is no Type 1 error, and accepting that our knowledge base is limited, it can thus be helpful to try and understand the biological processes that underlie the observed associations. One can thus try to characterize the impact of the identified variants at both the molecular and system levels. At a molecular level, expression quantitative trait locus studies allow the identification of both *cis* and *trans* regulatory associations of genetic variations with gene expression levels (Zuo et al. 2011).

Whole genome sequencing (GWS) technology and new arrays may soon overcome many of these limitations, though no doubt they have their own limitations (Ott et al. 2015).

5.3.3 GWAS findings in AD and related alcohol use disorder phenotypes

At time of writing, numerous GWAS on AD have been published. These will be summarized here. The initial GWAS of AD was published by Johnson et al in 2006 from the COGA samples. They used a hundred thousand SNP markers in 120 patients and 160 healthy controls. Fifty-one clusters of polymorphisms were identified including genes involved in brain development, cell adhesion and signaling, and gene regulation. Among these were the genes coding for the cell adhesion molecules cadherin 11 and 13 (CDH11 and CDH13). Following on from this pioneering work are numerous GWAS which have been conducted to identify risk factors for alcohol dependence. Further studies have

examined associations with alcohol-related phenotypes, including behavioural and neurophysiological measures, and also clinical features. Other studies have explored genetic adjusters of alcohol consumption in non-clinical/healthy samples.

Genes coding for alcohol metabolism are the most replicated biological function in AD GWA studies. In summary, Frank et al. reported a genome-wide significance for marker rs1789891 ($P = 1.27E-8$, odds ratio (OR) = 1.46) in the *ADH* gene cluster on chromosome 4q in a sample of 1,333 patients and 2,168 controls (Frank et al. 2012). This region has been replicated in further GWAS by Park et al. Takeuchi et al. and Quillen et al (Park et al. 2013; Takeuchi et al. 2011; Quillen et al. 2014). These findings confirm previous linkage findings discussed above and reviewed recently (Samochowicz et al. 2014).

GWAS of endophenotypes have been published. For example, a GWAS directed at alcohol dependence symptom count, found converging evidence for a role of the thus far poorly characterized *C15ORF53* gene in the aetiology of alcohol dependence-related phenotypes (Wang et al. 2013). Other endophenotypes show statistical, albeit weaker, association with the clinical phenotype of interest. For example, novelty-related brain oscillations in the theta band were associated with variations in the *ARID5A* gene (rs4907240) and the serotonin receptor 7 (*HTR7*, rs7916403) (Zlojutro et al. 2010).

Expression quantitative trait locus (eQTL) studies allow the identification of regulatory associations of genetic variations with gene expression levels. For example, a combined GWAS-eQTL approach on data from 2,090 alcohol-dependent patients and 2,026 controls found associations of variants in the PHD finger protein 3 gene (*PHF3*), protein tyrosine phosphatase type IVA, member 1 gene (*PTP4A1*) region, and *KIAA0040* gene with AD; transcriptome-wide expression analyses confirmed that the transcripts of these genes were involved in regulatory mechanisms of other alcohol-related candidate genes (Zuo et al. 2011; Zuo et al. 2012).

In addition to these expected findings, GWAS have also helped to identify new risk loci for AD and related phenotypes, although with less confidence in the results. None of the SNPs achieved genome wide significance in a Australian/Dutch meta-analysis in the same paper, but a gene network analysis based on the top results revealed in this study overrepresentation of genes coding for ion-channels and cell adhesion molecules (Lind et al. 2010). Additionally a genome-wide association of pooled DNA from alcoholics was compared to quantitative trait loci implicated in mouse addiction phenotypes. This convergent strategy identified brain-expressed susceptibility genes involved in cell adhesion, enzyme activity, protein and transcriptional processes, neuroreceptors, ion channel and transport processes and cell structure. The set of proteins concerned was described as a “connectivity constellation” for susceptibility to addiction (Liu et al. 2006, Uhl et al. 2008a, Liu et al. 2005). Important members of this constellation are the cadherins (CDH), which I discuss later. An association between AD and CDH13 has been reported in a number of GWAS (Treutlein et al. 2009; Johnson et al. 2006). CDH11 has been implicated by an association with rs35164 (just downstream of CDH11) in early onset alcoholism (Edenberg et al. 2010). Cell adhesion molecules (CAMs). CAM gene products are preferentially involved with stereotyped and morphologically identifiable connections between cells (e.g. adherens junctions, gap junctions) and are over-represented in addictions genetic studies (Zhong et al. 2015).

Latterly, meta-analytic approaches have been used to increase power. For example, Wang et al. combined data from the SAGE, the COGA and the OZALC GWAS and found numerous signals for loci which not reported in these samples' initial publications (Wang et al. 2011). A further study by Gelernter et al. (with an impressive sample size of 16,087 subjects) demonstrated strong genome-wide associations with *ADH* gene cluster variants, and additionally identified a possible regulatory variant rs1437396, located between the *MTIF2* and *CCDC88A* genes as a novel risk variant with replicated genome-wide significance (Gelernter et al. 2014). Further work by the Luo's group published this year in a meta-analysis of European American and European Australian cohorts, reported 10 top-ranked single nucleotide polymorphisms (SNPs) ($p < 10^{-6}$) that were associated with alcohol dependence. They

included 6 SNPs at *SERINC2* ($3.1 \times 10^{-8} \leq p \leq 9.6 \times 10^{-8}$), 1 at *STK40* ($p = 1.3 \times 10^{-7}$), 2 at *KIAA0040* ($3.3 \times 10^{-7} \leq p \leq 5.2 \times 10^{-7}$), and 1 at *IPO11* ($p = 6.9 \times 10^{-7}$) (Zuo et al. 2015).

As a side note it is interesting that key SNP rs7590720 implicated in the Beirut GWAS did not replicate in a 2011 study of people who had sent their DNA to the company 23andme.com for genotyping and who completed a self-assessment questionnaire, including an assessment of “alcohol abuse” (Tung et al. 2011). Disappointingly, for our field of all the disease classes reported in this study (cancer, neurological, etc) psychiatric conditions were the least replicated (0/10 replications) compared to cancer (27/64 replications). There may be a number of reasons, but to me this emphasizes the importance of reliable diagnostic and phenotypic classification in psychiatric studies.

Finally, in a much publicized finding this year, Sulovari et al found evidence of association between AD and blue eye colour ($P = 0.0005$ and odds ratio = 1.83 (1.31–2.57)), supporting light eye colour as a risk factor relative to brown eye colour in people of European-ancestry. Their network-based analyses revealed a statistically significant ($P = 0.02$) number of genetic interactions between AD-associated genes and eye-colour genes. They found evidence of LD between an AD-associated GABA receptor gene cluster, *GABRB3/GABRG3*, and eye colour genes, *OCA2/HERC2*, as well as between AD-associated *GRM5* and pigmentation-associated *TYR* (Sulovari et al. 2015). These findings may be of interest to researchers with similar ancestral samples.

5.3.4 CNVs in AD

CNVs are structural variations of the DNA that include insertions, deletions, inversions, and duplications that may vary in length from a few kilobases to several megabases (Lee et al. 2012). Two human chromosomes in a population differ at a rate of 0.1%, according to the early surveys of genetic variation (International Hapmap Consortium 2005; Marian 2012). Although SNPs account for the most numerous variants in the genome, CNVs have been reported to involve up to 12% of the human genome,

contributing to a sizeable amount of phenotypic variation within normal individuals and complex diseases, including psychiatric disorders like schizophrenia and bipolar disorder. CNVs are classified as inherited and de novo. De novo CNVs can occur at a rate of up to four folds greater than single nucleotide substitution rates (Lupski 2007). In comparison to SNPs, the overall genomic change, as measured in nucleotides, is high in the case of CNVs and can be in the order of 1000 bp per generation, leading to a larger functional impact per site (Malhotra and Sebat 2012).

CNV deletions and duplications possibly interfere with the regulatory regions or coding sequences of various genes, thereby altering their genetic makeup and biological functions. Additionally, the presence of CNVs can result in the up- or down-regulation of dosage sensitive genes, thereby contributing to disease susceptibility variation (Lee et al. 2012). Based on the sheer size and potential to impact genes or multiple genes, structural variations are considered more pathogenic on average, and de novo CNVs might be more enriched in variants that have a large effect on disease risk (Malhotra and Sebat 2012). CNV studies are a formal test of the CDRV hypothesis. A widespread distribution of submicroscopic variations (<500 kb in size) in DNA copy numbers in normal human genomes was demonstrated in two landmark studies conducted in 2004 (Iafrate et al. 2004; Sebat et al. 2004). In addition, CNV deletions may account for the variations in intermediate phenotypes within complex neuropsychiatric illnesses, such as cognitive impairment or physiological measures (Friedman et al. 2008).

While the literature on CNVs in AD is somewhat behind that for SZ and BP there are some findings to report. In a sample from the Study of Addiction: Genetics and Environment (SAGE) study, CNVs in 6q14.1 and 5q13.2 were significantly associated with AD after adjusting for multiple tests. On chromosome 5q13.2, there were multiple candidate genes previously associated with various neurological disorders including *SMA4*, *SERF1*, *SERF1B*, *SMN2*, *SMA3*, *NAIP*, *GTF2H2*, *GTF2H2D*, and *OCNL*. The region on chromosome 6q14.1 is a gene desert that has been associated with mental

retardation and language delay but for technical reasons the authors suggest less confidence in the CNV result in this region (Lin et al. 2012). A more recent association was found with CNVs at 11 q14.2 and brain regional volume differences in a group with alcohol use disorder (AUD). Deletion carriers were found to have smaller cerebral cortex, cerebellar, caudate and larger cerebral white matter and 5th ventricle volumes than insertion carriers or subjects with no variation in this region. Similarly, deletion carriers also demonstrated higher AUD severity scores than insertion carriers or subjects with no variation (Boutte et al. 2012).

5.4 SUMMARY

Alcohol dependence is a severe brain-based medical disorder responsible for significant morbidity, mortality and economic loss worldwide. Heritability estimates are in the 50-60% region and environmental influences are complex including availability, price, cultural norms, type of alcohol and age. No single genetic cause has been found and candidate gene studies have been ambiguous with the exception of the alcohol dehydrogenase cluster on chromosome 4. GWAS have confirmed the ADH cluster involvement and identified novel candidate genes for further study. In the next section I cover the genetic background of bipolar disorder, a commonly co-occurring disorder with alcohol dependence.

CHAPTER 6: BIPOLAR DISORDER

6.1 CLINICAL FEATURES

Bipolar disorder, otherwise known as manic depression, is a severe mental illness also thought to have a significant genetic aetiology. Bipolar disorder is characterized by periods of abnormally elated (manic) or depressed mood along with accompanying biological and psychological symptoms. The current diagnostic criteria for bipolar disorder require at least one episode of mania, hypomania, or a mixed episode (with both symptoms of depression and mania occurring within the same episode) and one other separate episode of abnormal mood (either depression or mania). Severe episodes at either pole may include psychotic symptoms. These may be experienced as hallucinations (perceptions in any sense without a stimulus) and delusions (firmly held culturally unacceptable false beliefs). Patients may exhibit risky behaviour including overspending, disinhibition, risk taking, promiscuity, and suicide. Both manic and depressive episodes increase the risk of using drugs or alcohol, and of suicide.

Bipolar disorder generally responds well to medication with mood stabilisers or antipsychotics.

Biological differences are observed in structural and functional brain studies (Arnone et al. 2009) with substantial evidence for a biological aetiology. Details of treatment and brain studies are beyond the scope of this paper except where relevant to the findings.

Again, researchers have used twin, family and adoption studies to unpick genetic from environmental influences. Then they have explored the molecular genetics, which I will cover in this chapter.

6.2 FAMILY, ADOPTION AND TWIN STUDIES

Like alcoholism, BP has been noted to cluster in families. Falret (1854), for example, observed that “circular insanity” seemed to run in families (quoted in Maj et al. 2002). However important environmental risk factors include child abuse, drug and alcohol misuse, and trauma (Alloy et al. 2005).

Traditional family studies conducted during the 20th century based on clinically ascertained samples show substantial familial aggregation, with sibling relative risks of around 8–10 for bipolar disorder (Tsuang & Faraone 1990; Shi et al. 2008; Gershon et al. 1975; Gershon 1982; Tsuang et al. 1980; Maier et al. 1993). Considerably larger studies, based on Scandinavian national population registers, have substantiated the results of the earlier family studies. (Lichtenstein et al. 2009; Laursen et al. 2005; Mortensen et al. 2003; Mortensen et al. 2010; Gottesman et al. 2010).

Family, adoption, twin, linkage and allelic association studies have contributed to the evidence for genetic susceptibility to bipolar affective disorder with markers at several genes showing replicated association with bipolar disorder (Baum et al. 2008; Craddock et al. 2008; Moskvina et al. 2009; Ferreira et al. 2008; Askland 2006; Askland et al. 2009). Heritability estimates for bipolar disorder are between 89% and 93% (McGuffin et al. 2003; Kieseppä et al. 2005). Thus, BP is more heritable than AD and similarly heritable to SZ.

6.3 LINKAGE STUDIES

There have been numerous genome wide linkage scans of BP, and signals from individual studies have been reported throughout the genome, including chromosomes 1, 4, 6, 10, 12, 13, 18, 21, 22, and X

(comprehensively reviewed by Hayden & Nurnberger 2006) . Consistent with other disorders of complex aetiology, independent replication of linkage signals for BP has not been consistent.

McQueen and colleagues conducted a combined analysis using the original genotype data from 11 BP genomewide linkage scans comprising 5,179 individuals from 1,067 families (McQueen et al. 2005). Heterogeneity among studies was minimized in their analyses by using uniform methods of analysis and a common, standardized marker map and was assessed using novel methods developed for meta-analysis of genome scans. This collaboration was the largest and most comprehensive analysis of linkage samples involving a psychiatric disorder at the time (a decade ago), and the collaboration was almost prescient given the evaluation of multi-centre GWAS mega analyses discussed later.

6.4 ASSOCIATION STUDIES

6.4.1 Candidate genes

A number of genetic pathways have been proposed for the pathogenesis of BD. To remain focused, a comprehensive review of only the most replicated and consistent association findings will be discussed in the following sections.

The serotonin (5HT) system is an obvious candidate gene because of the efficacy of serotonergic drugs in treating depression and the evidence implicating 5HT in mood disorders. One of the most widely studied genes in BD is the serotonin transporter, solute carrier 6A4 (SLC6A4), located on 17q, which has been reported significant in two linkage studies. Several association studies and four meta-analyses have also been published with associated findings (Serretti & Mandelli 2008). The tryptophan hydroxylase 2 (TPH2) gene encoding a rate-limiting enzyme in the biosynthesis of CNS serotonin has

been associated with BP in a number of independent studies, including a recent meta-analysis (Gao et al. 2016) although negative associations have also been reported in different populations (Serretti & Mandelli 2008; Barnett & Smoller 2009; Choi et al. 2010; Baum et al. 2008). The serotonin receptor genes HTR1A, HTR2A, HTR2C and HTR7 have also been investigated in several of studies with some evidence of association (Serretti and Mandelli 2008).

Dopamine (DA) is another obvious candidate supported by the involvement of the DA system in BP and SZ, and the response of BP to DA blocking medications (antipsychotics). The role of the dopaminergic pathway in the aetiology of bipolar disorder was first described by Murphy who provoked manic states in healthy subjects with psychoactive substances increasing activity of the dopaminergic system (Murphy et al. 1971).

The dopaminergic genes DRD1, DRD3, DRD4 and SLC6A3 have been investigated in several studies in different populations. Positive reports for a promoter polymorphism and haplotype association along with negative SNP associations have been reported for DRD1 (Severino et al. 2005; Dmitrzak-Weglarz et al. 2006; Nöthen et al. 1992; Cichon et al. 1996). DRD3 has been reported to be positive in only one study. Several studies provide considerable support for the involvement of DRD4 and SLC6A3 genes in bipolar disorder (Serretti and Mandelli 2008).

Catechol-O-methyltransferase (COMT) is another promising candidate gene for BP. COMT encodes for an enzyme catalysing the degradation of numerous key CNS catecholamine neurotransmitters including dopamine, epinephrine, and norepinephrine. The functional Val158Met polymorphism in this gene has been widely investigated with mostly negative results (Shifman et al. 2004; Serretti and Mandelli 2008). More recently, a study has reported interaction between DRD3 Ser9Gly polymorphism and COMT

Val158Met polymorphism in BP I patients (Lee et al. 2011), but this has not been replicated in other studies.

Brain derived neurotrophic factor (BDNF) is a candidate gene selected based on evidence for both positional and functional involvement in BP. This gene is located at 11p13, a region consistently linked to bipolar disorder in previous linkage studies, which also contains the DRD4 locus (Pato et al. 2004; Craddock and Sklar 2009). BDNF codes for a survival factor induced by cortical neurons that is necessary for the continued existence of striatal neurons in the brain. Reports have been published that suggest effects of the Val158Met polymorphism on BDNF secretion and memory (Egan et al. 2001). Variants in BDNF have also been reported to be associated with working memory and schizophrenia (Diaz-Asper et al. 2008). Further, two meta-analyses have confirmed the association between BDNF and bipolar disorder, particularly in the Caucasian population (Serretti and Mandelli 2008; Craddock & Sklar 2009).

D-amino acid oxidase activator (DAOA) is a gene which degrades the gliotransmitter D-serine, a potent activator of N-methyl-D-aspartate (NMDA) type glutamate receptors, DAOA, and the glutamatergic pathway that have been associated with BP and SZ in several studies. Located in a positive linkage region, 13q33.2, DAOA plays an important role in the activation of NMDA receptors, which are implicated in schizophrenia and bipolar disorder (Chumakov et al. 2002; Williams et al. 2006). Several independent studies in the Caucasian population have reported positive associations with different DAOA SNPs in bipolar disorder (Hattori et al. 2003; Chen et al. 2004; Schumacher et al. 2004; Prata et al. 2008; Bass et al. 2009; O'Donovan et al. 2009; Cherlyn et al. 2010; Nothen et al. 2010).

Disrupted in schizophrenia 1 (DISC1) was discovered in 2000 through the molecular cloning of a chromosomal translocation that co-segregated with a spectrum of major mental illnesses in a single

large Scottish family (St Clair et al. 1990). Through *in vitro* experiments and mouse models, *DISC1* has been firmly established as a genetic risk factor for a spectrum of psychiatric illness (Lipina, 2014). *DISC1* is widely expressed not only in the brain, most highly during foetal neurogenesis and in the adult hippocampus, but also in other tissues. As a consequence of its protein scaffold function, the *DISC1* protein impacts on many aspects of brain function, including neurosignaling and neurodevelopment. Since then, multiple studies have confirmed associations between *DISC1* and bipolar disorder; however, a pathogenic variant has yet to be detected (Serretti and Mandelli 2008; Craddock and Sklar 2009; Hennah et al. 2009; Porteous, 2011).

Abnormalities in GABA function have been postulated for many years in bipolar disorder and are supported by the observation of alterations in GABA subunit ratios in the postmortem brain (Fatemi et al. 2013), and decreased CSF GABA in depressed patients where depression is a key “pole” of bipolar disorder (Mann et al. 2014). Benzodiazepines, which are positive allosteric modulators of GABA_A receptors are also used adjunctively in the treatment of mania, a key aspect of BD.

A random-effects meta-analysis of genetic association studies on all polymorphisms of bipolar disorder in three or more case-control studies was published by Seifuddin and colleagues (2012). The results from the meta-analysis were then compared with the findings from a mega-analysis of eleven GWAS in bipolar disorder (Sklar et al. 2011). Nominally significant associations ($P < 0.05$) were found for polymorphisms in Brain derived neurotrophic factor (BDNF), Dopamine Receptor D4 (DRD4), D-amino acid oxidase activator (DAOA) and Neuronal tryptophan hydroxylase (TPH1), although the findings did not survive correction for multiple testing.

Several other genes with replicated positive findings have been identified, including TRPM2, GSK3 β , DTN bipolar disorder1, NRG1, NCAM1, GRIN2B, MOAO, P2RX7, GABA receptor genes, and and

TACR1 (McQuillin et al. 2006; Lachman et al. 2007; Serretti and Mandelli 2008; Craddock and Sklar 2009; (Sharp et al. 2014). Additionally, genes involved in circadian rhythms have been investigated, including CLOCK and BMAL1 (Barnett and Smoller 2009).

The evidence for allelic association with bipolar disorder at various different genes is convincing. The problem of locus heterogeneity leads to a protracted time between initial discovery and replication (Zollner and Pritchard 2007). To alleviate the problem of locus and allelic heterogeneity, extremely large samples are required for robust identification of genes influencing risk and expression of bipolar disorder. Larger sample sizes combined with technological advances ushered in the era of the genome wide association study (GWAS), discussed in the next section.

6.4.2 GWAS

Numerous high quality GWAS on BP have been published in the last seven years revealing numerous interesting loci, some of which have been replicated. The first GWAS on BP was reported by Baum et al. (2008) using 461 BP I patients and 562 healthy controls of European Caucasian background from the NIMH. Over 550,000 SNPs were genotyped in the two samples. Although pooling of DNA samples reduces the power to detect genetic association, this method was used to reduce the cost at that time of genotyping a large number of samples. The strongest association signal was detected within the first intron of DGKH (diacyl glycerol kinase) ($P = 1.5 \times 10^{-8}$, OR=1.59).

The second consortium GWAS was reported by Sklar et al. (2008) and included non-overlapping Caucasian case-control samples from several research groups. The study sample consisted of 1461 BP I cases and 2008 controls from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) study and UCL. A total of 372,193 SNPs were genotyped in the study using Affymetrix

GeneChip Human Mapping 500K Array Set. This same dataset was used in this thesis. Again, none of the SNPs met genome-wide significance on association analysis, but the top SNPs associated with bipolar disorder were in genes MYO5B ($P = 1.66 \times 10^{-7}$), TSPAN8 ($P = 6.11 \times 10^{-7}$) and EGFR ($P = 3.26 \times 10^{-5}$).

There have since been a number of GWAS by larger consortia including the Wellcome Trust Case Control Consortium (WTCCC) (2007); STEP-BD, University of Edinburgh, and Trinity College Dublin (ED-DUB-STEP2) (Ferreira 2008); Baum et al. 2008; Sklar et al. 2008; Cichon et al. 2011; Scott et al. 2009; Smith et al. 2009;). Methodology and sample sizes reflected not only the technology at the time but also the cost of genotyping. The number of markers has increased from 500 thousand to over a million SNPs (and similar numbers of CNVs) with the cost reducing dramatically. This has empowered researchers to perform ever-larger studies of ever more markers.

Our UCL BP sample has been examined in several BP GWAS and have reported some evidence to implicate MYO5B, CACNA1C, and ANK3 (Ferreira et al. 2008; Sklar et al. 2008),

Collaborative GWAS of bipolar disorder have identified genome-wide significant associations with markers in the genes encoding the alpha 1C subunit of the L-type voltage-gated calcium channel (CACNA1C) (12p13) and ankyrin 3 (ANK3) (10q21) (Ferreira et al. 2008) Subsequently, the Psychiatric Genomics Consortium (PGC) published the largest GWAS “mega-analysis” of bipolar disorder to date, involving over 60 000 participants. The CACNA1C association was further substantiated and there was a new genome-wide significant association with a marker in ODZ4 (11q14) (Sklar et al. 2011). These results strengthened the evidence of ion channelopathies being important in the pathogenesis of BP.

A novel bipolar disorder susceptibility candidate gene, Neurocan (NCAN) reached genome-wide significance ($P=3.02 \times 10^{-8}$, $OR=1.31$) in a GWAS conducted by Cichon et al. (2011) on European samples. A significant genome-wide association for NCAN was replicated in two follow-up steps and the combined analysis of all the samples which included 8441 cases and 35,362 controls also confirmed the association ($P=2.14 \times 10^{-9}$, $OR=1.17$).

In other populations, there have been BP GWAS in Han Chinese (Lee et al. 2011a), Norwegian (Djurovic et al. 2010), Japanese (Hattori et al. 2009), and Bulgarian (Yosifova et al. 2011), although none of these studies achieved SNPs with genome-wide level significance.

A meta-analysis of 5 European ancestry major mood disorder case-control samples, including bipolar disorder and MDD found genome-wide significant association at 3p21.1 (rs2251219, $P = 3.63 \times 10^{-8}$, $OR = 0.87$) (McMahon et al. 2010). Another meta-analysis of genome-wide association data of bipolar disorder and MDD was conducted by Liu et al. (2011). This study provided evidence for the role of CACNA1C variants in bipolar and unipolar major mood disorders, but failed to confirm the previous association of bipolar risk locus ANK3 in the meta-analysis. Even larger collections of bipolar disorder GWAS signals were analysed for age of onset and psychotic symptoms and two sub-phenotypes of bipolar disorder. These included GWAS data from GAIN-BP, BiGS, and a German study, but it failed to report any significant genome-wide associations (Belmonte Mahon et al. 2011).

Another study comparing the marker allele frequencies between schizophrenia and bipolar disorder cases rather than normal controls reported a significant association for CACNG5 that supports the role of voltage-dependent calcium channel genes, such as CACNA1C, in the susceptibility to bipolar disorder (Curtis et al. 2011).

The most recently published BP GWAS was a family-based study of 229 small families with association analysis from over 950 cases and 950 ethnicity-matched controls from the UK and Canada, using imputation (Xu et al. 2014). No genome-wide significant markers were identified, however the combined GWAS findings suggested several genes of interest that support findings from other groups, such as at *SYNE1* on 6q25, *PPP2R2C* on 4p16.1, *ZNF659* on 3p24.3, *CNTNAP5* (2q14.3), and *CDH13* (16q23.3). The family-based analysis also implicated regions on 1q21.2 (closest gene: sphingosine-1-phosphate receptor 1 gene, *SIPRI*) and on 1q24.1 near the gene *TMCO1*, and at *CSMD1* on 8p23.2. Pathway analysis suggested association of pathways involved in calcium signalling, neuropathic pain signalling, CREB signalling in neurons, glutamate receptor signalling and axonal guidance signalling. A more recent pathway analysis of NIMH BP GWAS data implicated interleukin pathways: the IL-6, IL-8 and IFN related pathways held twice to thrice the number of expected variants associated with BP (Drago, 2015).

In summary, SNP variants in *ANK3*, *ODZ4*, *SYNE1*, *CACNA1C*, *DGKH*, and *C15ORF53* have consistently been associated with bipolar disorder in different GWAS.

6.4.3 CNVs in BP

A genome-wide survey of large, rare CNVs in a case-control sample from WTCCC (Grozeva et al. 2010) found that the rate of CNVs was higher in controls than in bipolar cases. The absence of rare and large structural deletions associated with bipolar disorder and their presence in 5%-10% of schizophrenia cases suggests that the two disorders are genetically distinct for this type of genetic variation. Similar results were obtained in another study utilizing the UCL bipolar disorder research sample (McQuillin et al. 2011), in which the rare CNV burden was also slightly lower in the cases compared to controls. That

both studies found higher rates of CNVs in controls is difficult to explain, but it could be due to undiagnosed cases in the controls, or the presence of carriers of SZ CNVs.

The effect of de novo CNVs on age-at-onset in bipolar disorder has consistently been nominally significant in several studies (Zhang et al. 2009; Malhotra et al. 2011; Priebe et al. 2012). This suggests that individuals with early-onset bipolar disorder may represent a subclass of the disorder in which there is a greater contribution from rare CNV alleles that have large effect. Support for the notion that segregation of early-onset bipolar disorder in families is consistent with major gene effects. However, familial segregation of bipolar disorder and unipolar disorders shows a wide range of ages of onset and it is not possible to identify clear cut differences in age of onset between families. The potential role of CNVs in BP is complicated with some studies reporting no link and others reporting an overrepresentation of CNVs (Zhang et al. 2009; Malhotra et al. 2011) in patients with bipolar disorder and other studies reporting no associations (Grozeva et al. 2010; McQuillin et al. 2011).

In conclusion, several lines of evidence suggest that the overall rare CNV burden is more modest for bipolar disorder than for schizophrenia, with SZ showing a large increase in the number of CNVs found to be associated compared to controls (Girirajan and Eichler 2011).

6.5 SEQUENCING

Ament and colleagues (2014) sequenced the genomes of 200 BP patients and focused on over 3000 candidate genes informed by previous research. They found that BP pedigrees had an increased burden of rare variants in genes encoding neuronal ion channels, including subunits of GABAA receptors and voltage-gated calcium channels. Targeted sequencing of 26 of these candidate genes confirmed rare

variant associations in *ANK3*, *CACNA1B*, *CACNA1C*, *CACNA1D*, *CACNG2*, *CAMK2A*, and *NGF* (Ament et al. 2015).

6.6 BP CO-MORBIDITY WITH ALCOHOLISM

Alcohol problems are significantly more common in bipolar disorder than the unaffected population (range 13-25%). These studies show that about 30-50% of individuals with bipolar disorder also manifest alcohol misuse or dependence and altogether up to 75% of bipolar patients have any substance use disorder (Regier et al. 1990; Schuckit 2009; Kessler et al. 1997). The lifetime prevalence of alcohol misuse or dependence in bipolar disorder (including types I and II) is 43.6% and over 60% for bipolar type I (Regier et al. 1990).

6.6.1 Heritability

Comorbidity between bipolar disorder (BP) and alcoholism is well established from epidemiological, twin, family and linkage studies (O'Sullivan et al. 1988, Gurling et al. 1984, Merikangas and Gelernter, 1990, Preisig et al. 2001, Sonne and Brady, 2002, Berrettini et al. 1997, Winokur et al. 1996). For both alcoholism and bipolar disorders, co-transmission in families has been reported (Maier and Merikangas, 1996), particularly in subgroups of patients with familial comorbidity (Maier et al. 1995; Saunders et al. ; Schulze, 2006; Winokur et al. 1996) and suicidality (Potash, 2000).

There may also be common neurobiological mechanisms crossing both disorders (Farren et al.,2012). For example, they have shared abnormalities within neuropeptide and neurotransmitter systems (Rakofsky & Dunlop 2013), and both groups have reduced grey matter volumes within the anterior cingulate gyri relative to unaffected controls (Nery et al. 2010).

Familial co-transmission could be either environmental or genetic. Two adoption studies favour the genetic hypothesis. Ingraham showed that substance misuse was more common in the biological relatives of affectively ill adoptees than in controls' relatives (Ingraham and Wender, 1992).

Relatives of early onset bipolar patients (earlier onset is thought to be more genetically influenced) have a prevalence of alcoholism two to three times higher than that reported in controls (Todd et al. 1996). Cadoret showed that, on the basis of his adoptions studies, the depressive spectrum might represent a gene environment interaction involving the genetic diathesis for alcoholism (Cadoret et al. 1995). These findings suggest shared genetic factors being responsible for BP and AD or alcohol misuse.

Several linkage studies have revealed that chromosomal regions associated with bipolar disorder overlap to some degree with those described for alcohol dependence. The overlapping regions involve chromosomes 1p, 6q, 9p, 14q (Nurnberger Jr. et al. 2001; Segurado et al. 2003; Hill et al. 2004; McQueen et al. 2005). A genetic signal of comorbid disorder overlap was in a combined depression-alcoholism phenotype with a linkage peak on chromosome 1p in the Consortium on the Genetics of Alcoholism (COGA) (Dick et al. 2002).

Johnson and colleagues found several overlapping clusters of single nucleotide polymorphisms (SNPs) when comparing the genome-wide findings of four studies on bipolar disorder and one on substance dependence, including alcoholism and illicit drug use (Johnson et al. 2009).

Aside from our group's work, there have been no other genome-wide association studies of bipolar disorder with co-morbid AD.

6.6.2 Clinical implications of co-morbidity

Alcohol misuse in BD is important clinically because comorbid alcohol problems in people with BD are associated with poorer outcomes, treatment non-compliance, an increased suicide risk, and higher rates of other substance misuse problems (Baethge et al. 2008; Oquendo et al. 2010).

Diagnosing bipolar disorder in alcohol-misusing patients can be clinically challenging. Numerous factors affect diagnostic precision, such as features common to both disorders (e.g. excessive involvement in pleasurable undertakings), the underreporting of symptoms (particularly symptoms of mania), and the complex bidirectional effects of alcohol on mood states. Bipolar patients are also likely to use drugs other than alcohol (e.g. stimulant drugs such as cocaine or methamphetamine), further complicating the diagnosis. One study suggested that about 60 percent of patients with both alcoholism and bipolar disorder started using alcohol before the onset of the mood episodes (Strakowski et al. 2000). Diagnostic accuracy can be improved by eliciting a careful history of the chronological order of both illnesses.

Several potential mechanisms might explain the development of alcohol use disorders in people with BP. Alcohol use may also alter the same neurotransmitters or pathways as those involved in bipolar disorder, thereby “prompting” the symptoms of bipolar disorder (Tohen et al. 1998). The self-medication hypothesis suggests that AD develops from the frequent use of alcohol to cope with the symptoms and distress associated with BD (Strakowski et al,2000). Plausible clinical relationships between the disorders have been described by Sonne and Brady and will not be explored further here (Sonne and Brady, 2002).

6.6.3 Dopamine in bipolar and alcoholism

The dopaminergic pathway is implicated in the pathogenesis of SZ, BP and AD as in all three disorders, disturbances in dopaminergic neurotransmission have been observed. Molecular genetic analyses of the dopamine receptor genes revealed polymorphisms that may increase vulnerability to BP as well as alcoholism. The *DRD1* gene -48 G/A polymorphism was found to be associated with sensation seeking in AD men (Limosin et al. 2003). Linkage studies localized the risk locus for bipolar disorder to the 5q chromosome, in the proximity to the *DRD1* gene (Garner et al. 2001); (Schinka et al. 2002) and it was observed that this polymorphism is associated with bipolar disorder in a Sardinian population (Severino et al. 2005) and a Polish sample (Dmitrzak-Weglarz et al. 2006).

A study by Gorwood et al. (2000) investigated possible role of the *DRD2* gene TaqI A1 polymorphism in the shared vulnerability to AD and BP. The researchers analysed four groups (AD, BP, BP+ALC and controls) but no association was found (Gorwood et al. 2000). Studies on the relationship between the Taq1 A1 promoter polymorphism and alcoholism have shown both negative (Sander et al. 1999; Noble et al. 2000) and positive associations (Ishiguro et al. 1998; Samochowiec et al. 2000; Wodarz et al. 2003). For bipolar disorder, a positive Taq1A1 association was found in Chinese bipolar patients (Li et al. 1999), however this result was not replicated in other studies (Furlong et al. 1998; Stöber et al. 1998).

For the *DRD3* gene, several positive association studies were reported for 9 Ser/Gly and AD (Limosin et al. 2005; Sander et al. 1995; Thome et al. 1999). For bipolar disorder, the association studies involving this polymorphism were considered as suggestive (Parsian & Cloninger 1995) or negative (Rietschel et al. 1993; Piccardi et al. 1997; Elvidge et al. 2001).

Novelty seeking, alcohol and drug abuse and the DRD4 –521 C/T promoter polymorphism have been found (Schinka et al. 2002). Indirect support for a role for *DRD4* in alcohol dependence is the location of this gene in one of the linkage regions in chromosome 11 (Long & Goldman 1998). However, these findings were followed by both positive and negative replication studies (Lusher et al. 2001; Preuss et al. 2002). For bipolar disorder, the recent meta-analysis performed by Leon *et al.* (López León et al. 2005) revealed the association of bipolar illness and the DRD4 gene polymorphism. Szczepankiewicz and colleagues did not find an association with BP, AD and the dopamine receptor genes DRD1, DRD2, DRD3, and DRD4 (Szczepankiewicz et al. 2007) .

My colleague Sally Sharp from UCL others published on tachykinin receptor 1 gene (TACR1) in multiple disorders: AD, BP, and attention deficit hyperactivity disorder (ADHD). She found that TACR1 SNP rs3771829 was associated with BP, AD and BPALC compared with controls screened for the absence of mental illness and alcohol dependence. DNA sequencing in selected cases of BPAD and ADHD who had inherited TACR1-susceptibility haplotypes identified 19 SNPs in the promoter region, 5' UTR, exons, intron/exon junctions and 3' UTR of TACR1 that could increase vulnerability to all four phenotypes (Sharp et al. 2014).

In summary, bipolar disorder is a complex illness of mood, with epidemiological, neurobiological and genetic overlaps with alcoholism. In the next chapter, a similar picture is drawn for schizophrenia, which also shares genetic and other factors with AD and BP.

CHAPTER 7: SCHIZOPHRENIA

7.1 DEFINITIONS

Schizophrenia (SZ) is a severe mental disorder marked by hallucinations, delusions, cognitive deficits and apathy with heritability estimated at 73-90% (Sullivan et al. 2003). Inheritance patterns are complex and the number and type of genetic variants involved are not fully understood.

7.2 FAMILY, ADOPTION AND TWIN STUDIES

Adoption studies during the 20th century show familial aggregation of schizophrenia in biological relatives who were separated by adoption early in life, (Heston 1966; Rosenthal et al. 1971; Kety et al. 1994; Wender et al. 1974; Kendler et al. 1994; Tienari et al. 2003) consistent with genetic influences.

7.3 MOLECULAR GENETIC STUDIES

7.3.1 Linkage studies

Over 20 genome linkage scans for schizophrenia have been published, although disappointingly with no strong evidence for a gene of major effect. To date, eight linkage regions have identified promising candidate genes: 22q12–q13, 8p22–p21, 6p24–p22, 13q14–q32, and 6q21–q22, reviewed by Riley and Kendler's comprehensive textbook (Riley & Kendler 2005).

The classic Maryland family sample studies gave the first evidence of linkage to chromosomes 22q12-q13.1 and 8p22-p21 (Pulver et al. 1995; Pulver et al. 1994). Following linkage evidence to 8p in Icelandic families, fine mapping identified two risk haplotypes spanning a region of ~ 1 Mb within the gene for neuregulin 1 (NRG1) which was supported by further work but no linkage to the NRG region was found in an Irish cohort (Stefansson et al. 2002; Thiselton et al. 2004). Neuregulin is expressed in CNS synapses and appears to have a role in the expression and activation of neurotransmitter (including glutamate) receptors. Data from a mixed sample of UK and Japanese families initially suggested linkage to 13q14.1-q32 (Lin et al. 1995) an interesting region which contains the 5HT-2A receptor gene.

Overall, the combined linkage reports are spread over a region of ~ 60 Mb, containing ~ 120 known or putative genes. Initial evidence for a susceptibility locus on 6q21-q22.3 came from a US family sample, and a collaborative study of this region was also significant (Cao et al. 1997; Levinson et al. 2000). Interest in chromosome 1 in schizophrenia began with reports of a balanced 1:11 translocation segregating with serious mental illness in a large pedigree from Scotland (Blackwood et al. 2001). The chromosome 1 breakpoint lies at 1q42.1. A genome scan meta-analysis (GSMA) was carried out on 32 independent genome-wide linkage scan analyses that included 3255 pedigrees with 7413 genotyped cases affected with SZ or related disorders. Suggestive evidence for linkage was observed on chromosomes 5q (142–168 Mb) and 2q (103–134 Mb) (Segurado et al. 2003).

Given the evidence of linkage replication across several groups for regions on 22q, 8p, 6p, 13q, 5q, 10p, 6q, 1q and 15q, and subsequent association replication for a number of genes including DTNBP1, NRG1, it would appear that the risk of consistent false positives is low.

7.3.2 Association studies

For brevity, and because my focus is on AD, I will attempt to summarize the hundreds of studies of SZ candidate genes and GWAS, mentioning only the most relevant. The interested reader may wish to consult the excellent recent review by Cardno and Owen (2014).

Similar to BP, numerous candidate genes related to neurotransmitters and cell communication have been studied in SZ, and new pathways have been implicated by hypothesis-free GWA studies.

In a comprehensive study of some of the most cited candidate genes (e.g. *DISC1*, *DTNBP1*, *NRG1*, *DRD2*, *HTR2A* (5-hydroxytryptamine [serotonin] receptor 2A), and *COMT*), each of 14 genes were tested by genotyping a sample of 1,870 cases and 2,002 screened controls of European ancestry (EA) (Sanders, 2008). A total of 789 (SNPs) including those located in functional domains of genes were genotyped, but no association was found. This appears to contradict ORs predicted from the analysis of smaller samples. The effect size is the strength of the association between a marker and the disorder, and it can be expressed as the odds ratio (OR). OR is the odds for an event, in this case, possessing a risk allele, in the risk group; i.e. cases, divided by the odds in the non-risk group, the controls.

Three recent meta-analyses of *COMT* association with schizophrenia have provided mixed (though generally negative) results (Glatt et al. 2003; Fan et al. 2005; Williams et al. 2005).

Brain-derived neurotrophic factor (BDNF) has been implicated in numerous psychiatric disorders. A recent study found allelic association between two BDNF SNPs (rs6265 and rs7103411) and comorbid AD ($P = 0.044$) in a primary and replication schizophrenia sample. Additionally, they

reported association between both SNPs and risk-taking behaviour after drinking (rs6265, $P = 0.005$; rs7103411, $P = 0.009$). When these SNPs were tested in a non-schizophrenia alcohol-dependent group the researchers were unable to detect association (Cheah et al. 2014).

7.3.3 GWAS of SZ and overlap with BP

The first substantive GWAS result in schizophrenia found association with a marker in the zinc finger-binding protein 804A gene (ZNF804A) on chromosome 2q32 (O'Donovan et al. 2008). A bipolar disorder sample was added to the analysis and the effect size remained similar (OR = 1.12), while the P value became impressively more significant ($P = 9.96 \times 10^{-9}$), consistent with the presence of a genetic variant that has a small effect on the risk of both disorders. These findings for ZNF804A have been further substantiated by larger meta-analysis (schizophrenia: $P = 2.5 \times 10^{-11}$, OR = 1.10; schizophrenia and bipolar disorder combined: $P = 4.1 \times 10^{-13}$, OR = 1.11) (Williams et al. 2011).

As GWAS sample sizes increase, thanks to large-scale international collaborations, genome-wide significant associations have been found with markers in or near the major histocompatibility complex region on chromosome 6 (Shi et al. 2009) and the genes encoding neurogranin (NRGN) (11q24) and transcription factor 4 (TCF4) (18q21) (Stefansson et al. 2009).

The UCL SZ sample has been used in several SZ GWAS and have implicated the human leukocyte antigen (HLA) region (Purcell, 2009). Our UCL BP and SZ samples have been used in association studies and resulted in identification of P2RX7 (McQuillin et al. 2009).for BP and shared risk genes including BRD1, DISC1, and DAOA (Bass et al. 2009; Hennah et al. 2009; Nyegaard et al. 2009)

whereas separately the schizophrenia sample has been reported to show association with *EPIN4* and *PCM1* (Pimm et al. 2005; Datta et al. 2008).

A GWAS mega-analysis of schizophrenia, also from the PGC and involving over 50 000 participants (Ripke et al. 2011) found genome-wide significant associations at 7 loci, 5 of which were new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two of which had been previously implicated (6p21.32-p22.1 and 18q21.2). The strongest of the new findings was for a marker in the gene encoding microRNA 137 (*MIR137*), which has a role in the regulation of neuronal development. The study also reported analyses of combined schizophrenia and bipolar disorder samples, with associations at 3 loci, *CACNA1C*, *ANKK1*, and *ITIH3-ITIH4*, being genome-wide significant and showing increased levels of statistical significance in the combined analysis, consistent with genetic variants in these regions influencing risk of both disorders.

A subsequent review summarized the strongest published GWAS findings for schizophrenia, bipolar disorder, and both disorders combined (Sullivan et al. 2012).). In keeping with the tradition of GWAS findings, the associations each involve a small effect on risk (ORs around 1.1) and are consistent with a partial overlap in genetic influences from commonly occurring genetic variants on the two disorders. There is also evidence that most of these associations occur in and around genes. (Moskvina et al. 2009; Schork et al. 2013).

Recently, the PGC Cross-Disorder Group has published a broader investigation encompassing 5 psychiatric disorders: schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, and attention-deficit hyperactivity disorder (ADHD) (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). In the primary analysis of all 5 disorders combined, markers in 4 regions achieved genome-wide statistical significance. In 2 cases (on chromosomes

3p21 and 10q24), the causal variant behind the marker association could be located in or between a number of genes within the region, while in the other 2 cases, the associations were in calcium channel signalling genes (CACNA1C on chromosome 12 and CACNB2 on chromosome 10), and additional pathway analysis further supported the role of calcium channel activity genes influencing all 5 disorders.

The PGC group also investigated the diagnostic specificity of genome-wide significant associations for schizophrenia and bipolar disorder found in previous PGC analyses. In keeping with previous findings, the results ranged from associations confined to one disorder to associations encompassing all 5 disorders (Table 7.1). Lastly, the PGC CDG analysis confirmed genetic pleiotropy across many disorders and comorbid disorders.

As GWAS sample sizes increase, thanks to large-scale international collaborations, genome-wide significant associations have been found with markers in or near the major histocompatibility complex region on chromosome 6 (Shi et al. 2009) and the genes encoding neurogranin (NRGN) (11q24) and transcription factor 4 (TCF4) (18q21) (Stefansson et al. 2009). Further studies are looking at treatment response among other variables (Ripke et al. 2013; Hamshere et al. 2013; Green et al. 2013; Steinberg et al. 2014; Ruderfer et al. 2013).

There has also been an initial GWAS directly comparing cases of schizophrenia and bipolar disorder, with a view to identifying genetic differences between the 2 disorders. (Ruderfer et al. 2013) The study was underpowered to achieve genome-wide significant results, but further larger studies are likely to follow soon.

Table 7.1. Genome-Wide Association Study Findings for SZ and BP from the review by Sullivan et al, combined with Specificity of Genome-Wide Association Study Findings for SZ and BP in the Psychiatric Genomics Consortium Cross-Disorder Group Study of 5 Psychiatric Disorders

Phenotype	Chromosome Where Marker Is Located	Nearest Gene	Odds ratio (OR)	Disorders Showing Cross-Disorder Association
Schizophrenia	1	<i>MIR137</i>	1.12	SZ, ASD
	2	<i>VRK2</i>	1.09	
	2	<i>ZNF804A</i>	1.1	
	2	<i>PCGEM1</i>	1.2	SZ, ASD
	6	<i>MHC</i>	1.22	SZ
	8	<i>MMP16</i>	1.1	SZ
	8	<i>CSMD1</i>	1.11	SZ
	8	<i>LSM1</i>	1.19	
	10	<i>CNNM2</i>	1.1	SZ, MDD
	10	<i>NT5C2</i>	1.15	SZ, BP, MDD, ASD, ADHD
	11	<i>AMBRA1</i>	1.25	
	11	<i>NRGN</i>	1.12	
	18	<i>CCDC68</i>	1.09	SZ, BP, MDD
	18	<i>TCF4</i>	1.2	SZ, ASD
Bipolar disorder	11	<i>ODZ4</i>	1.14	BP
	10	<i>ANK3</i>		BP
	19	<i>NCAN</i>	1.17	-
	12	<i>CACNA1C</i>	1.14	BP, SZ, MDD
	6	<i>SYNE1</i>		BP
Schizophrenia and bipolar disorder combined	2	<i>ZNF804A</i>	1.11	
	3	<i>ITIH3-ITIH4</i>	1.12	
	10	<i>ANK3</i>	1.22	
	12	<i>CACNA1C</i>	1.11	

Copyright (2012), Macmillan Publishers Ltd and Elsevier (2013)

Key: SZ= schizophrenia; BP = bipolar disorder; MDD = major depressive disorder; ASD = autistic spectrum disorder, ADHD = attention deficit hyperactivity disorder.

Most recently, Johnstone and colleagues carried out CNV analysis of *DISC1* and its binding partners *PAFAH1B1*, *NDE1*, *NDEL1*, *FEZ1*, *MAP1A*, *CIT* and *PDE4B* in Scottish and Northern Swedish population-based samples using multiplex amplicon quantification. They reported finding rare CNVs in *DISC1*, *NDE1* (together with adjacent genes within the 16p13.11 duplication), *NDEL1* (including the overlapping *MYH10* gene) and *CITA* (Johnstone et al. 2015).

7.3.4 The Major Histocompatibility Complex (MHC) Locus

The genes in the MHC region have many different biological functions, but genes with an immune function predominate. Histones regulate DNA transcription by chromatin modification through histone methylation or acetylation (Adegbola et al. 2008; Webb et al. 2008; Shi 2007), and have a role as antimicrobial agents. This raises the possibility that genetic variation in histones might underlie a differential placental susceptibility to infections and in turn they may alter the susceptibility to SZ through this mechanism. Infections during pregnancy have long been postulated in schizophrenia given the seasonal variation of birth and risk of SZ.

Confirmatory evidence of MHC involvement was only recently obtained from a combined analysis of GWAS data of three GWAS studies (Shi et al. 2009; Purcell et al. 2009; Stefansson et al. 2009). This meta-analysis combined the *p*-values for all imputed and genotyped SNPs from the most significant regions of each study and discovered a genome-wide significant association at the MHC region on chromosome 6.

Another new schizophrenia susceptibility gene from schizophrenia GWAS is zinc finger protein 804A (*ZNF804A*), which found an association with SNP rs1344706 (Shi et al. 2009; Purcell et al.

2009; Stefansson et al. 2009). Subsequently, rs1344706 in ZNF804A was reported to be associated with altered neuronal connectivity in the dorsolateral prefrontal cortex in a functional magnetic resonance imaging study of healthy controls (Esslinger et al. 2009).

Other genes in the same region are involved in chromatin structure (i.e. high mobility group nucleosomal binding domain 4, HMGN4), transcriptional regulation (i.e. activator of basal transcription 1, ABT1; zinc finger protein 322A, ZNF322A; zinc finger protein 184, ZNF184), G-protein-coupled receptor signalling (i.e. FKSG83), and the nuclear pore complex (i.e. nuclear pore membrane protein 121-like 2, POM121L2). Other samples and replication samples uncovered associations with neurogranin (NRGN) and with transcription factor 4 (TCF4). NRGN encodes a postsynaptic protein kinase substrate that binds calmodulin, mediating N-methyl-d-aspartate (NMDA) receptor signalling that is important for learning and memory, and relevant to the proposed glutamate pathophysiology of schizophrenia (Wang et al. 2008; Harrison & Weinberger 2005). TCF4 is a neuronal transcriptional factor essential for brain development, specifically neurogenesis (Gulacsi & Anderson 2008).

7.3.5 Rare CNVs and Schizophrenia

A number of rare (<1%) and large (>100kb) CNVs have thus far been implicated in schizophrenia 138–144, as reflected by overall CNV burden and individual CNV loci. Supporting evidence for association of specific rare and large CNVs with schizophrenia is emerging at 1q21.1, 2p16.3 (NRXN1), 15q11.2, 15q13.2, 16p11.2, and 22q11.21 138–144. The 3 Mb deletion at 22q11.21 (22qDS) has been known to cause velocardiofacial syndrome (VCFS), and increases the risk for schizophrenia 145–147. An epidemiological study found that more than 30% of 22qDS carriers develop psychosis, and approximately 80% of this manifesting as schizophrenia 147. This represents

the largest known individual risk factor for the development of schizophrenia, aside from having an identical twin with schizophrenia. The 22q11.21 CNV was the only CNV reaching genome-wide significance in some schizophrenia GWAS 142 and the only one not found in controls (case%/control% was 0.30/0.00). All the CNVs (except for 22q11.21) initially found only in schizophrenia cases were also found in healthy controls in later studies, suggesting that the penetrance of these rare CNVs may be relatively low. With the rest of the rare and large CNVs implicated in schizophrenia spanning multiple genes, specific gene effects, including the possibility of genes presenting pleiotropy, will be difficult to disentangle.

Genes that are involved in the immune system and in brain development and activity – two functions that have evolved rapidly in humans – tend to be enriched in CNVs. By contrast, genes that play a role in early development and some genes involved in cell division – both critical to fundamental biology – tend to be spared.

7.4 CO-MORBIDITY WITH BP AND AD

SZ and BP appear to share common genetic risk factors. This is evidenced by family, twin and adoption studies. The largest of these, a Swedish population register-based study has shown significant familial coaggregation between schizophrenia and bipolar disorder in parent–offspring, sibling–sibling, and biological parent–adopted-away offspring pairs the adopted-away offspring of affected biological parents (Table 7.2, from Lichtenstein 2009).

Several other family studies have supported similar findings. A meta-analysis of family studies published between 1980 and 2006 found evidence of familial overlap, with the first-degree relatives

of probands who had schizophrenia showing a significantly elevated risk of bipolar disorder compared with relatives of controls (OR = 2.08, P = .01) .

Data from risks of disorders in relatives have also been used to model the most likely mode of inheritance of schizophrenia (O'Rourke et al. 1982; Risch 1990) and bipolar disorder (Rice et al. 1987; Craddock et al. 1995). For both disorders, this is likely to be multifactorial in most or all cases, with many genetic and environmental risk factors, each insufficient to cause the disorder on their own, but having a cumulative effect on risk when they occur together in the same individual.

7.5 SUMMARY AND CONCLUSIONS

These findings in the round suggest that schizophrenia, despite the very high reported heritability, is among the most complex of human genetic disorders. Analyses support Gottesman and Shields's polygenic model for schizophrenia susceptibility, involving a set of hundreds of genes, each with unquantified but very small individual effects (Gottesman & Shields 1967). Further evidence suggests that cases have more rare (<1%) and large (>100kb) CNVs than controls. However despite substantial evidence of clinical co-morbidity there is little evidence of cross-heritability or significant common genes between SZ and AD.

Table 7.2 Relative risks for developing SZ and BP with affected probands (from Lichtenstein et al 2009)

Relationship to proband	Relative risk of SZ when proband has SZ (95% CI)	Relative risk of BP when proband has BP (95% CI)	Relative risk of SZ when proband has BP disorder (95% CI)	Relative risk of BP when proband has SZ (95% CI)
Family relationships with shared environment				
Offspring	9.9 (8.5–11.6)	6.4 (5.9–7.1)	2.4 (2.1–2.6)	5.2 (4.4–6.2)
Sibling	9.0 (8.1–9.9)	7.9 (7.1–8.8)	3.9 (3.4–4.4)	3.7 (3.2–4.2)
Adopted-away biological relatives				
Adopted-away offspring	13.7 (6.1–30.8)	4.3 (2.0–9.5)	4.5 (1.8–10.9)	6.0 (2.3–15.2)

CHAPTER 8: STUDY AIMS

The main aim of the thesis was to elucidate the genetic basis of alcoholism by employing two strategies: GWAS and LD mapping. The overall intention of the work was to identify novel AD loci, and to replicate previously implicated markers. Within this framework the aims were sevenfold:

- 1) To replicate the association of GABRA2 with AD in the UCL AD case-control sample.
- 2) To perform a GWAS analysis on the UCL BP+ALC (cases) vs BPNONALC (controls) samples
- 3) To perform a GWAS on BP+ALC vs supernormal controls (SNC)
- 4) To perform a GWAS analysis on the UCL SZ+ALC (cases) vs SZNONALC (controls) samples
- 5) To perform a GWAS SZ+ALC vs supernormal controls
- 6) To perform a meta-analysis of (BP+ALC)+(SZ+ALC) vs BP+SZ
- 7) To perform a meta-analysis of (BP+ALC)+(SZ+ALC) vs supernormal controls

Additionally, I aimed to perform a gene-based permutation test and pathway analysis for each GWA dataset.

CHAPTER 9: MATERIALS AND METHODS

LD between alleles at DNA markers and BP was tested for using a population based case-control design. I will first describe the ethics, power considerations, case-control sample, genotyping, techniques and statistical analysis in a generic sense.

9.1 ETHICAL APPROVALS

U.K. National Health Service (NHS) multi-centre and local research ethics committee approval was obtained (Study title: Genetic case control and brain imaging studies of mental illness, autism and alzheimer's dementia, REC reference:03/11/090). Each volunteer was given an information sheet and description of the study. All volunteers signed an approved consent form.

9.2 POWER CONSIDERATIONS

It is not possible to calculate a priori the size of sample needed to detect association with a given allele, since the relative risk conferred by that allele in the population from which our sample is drawn is not known. This applies to loci in which replication of association is being attempted, and novel loci. Relative risk estimates conferred by a particular allele from previous studies will only apply to the populations which have been studied. Association studies in SZ and BP suggest that considerable variability in risk alleles exists across populations so it would seem imprudent to rely on power calculations based on such assumptions.

It is well established that the larger the sample, the greater the power to detect allelic association (although the use of P-values in densely typed SNPs in genetic association studies may be

conservative if the assumption of independent tests does not hold - (Gao et al. 2010)). Thus the UCL laboratory has tried to maximise the size of the samples. By necessity, the collection of the UCL sample has run parallel to the association testing. Therefore, there is considerable variability in the total number of individuals genotyped for different markers, reflecting the point in time at which the genotyping was performed. However, no selection of cases or controls occurred for the association testing. At the point at which a particular marker was genotyped, the full sample that was available was used. Further controls were later genotyped in relation to an association study in schizophrenia and the control genotypes added into the bipolar analysis.

Power calculation shows that a sample of 600 cases and 600 controls has 89% (2-sided) power to detect LD at the $p < 0.05$ significance level when there is an allele frequency difference of 5% in the context of less common alleles - alleles present in 5% of controls and 10% of cases. To detect a 5% difference in a common marker allele which has a frequency of 45% in controls and 50% in cases, the power is only 38% (2-sided).

9.3 UCL CASE-CONTROL SAMPLES

9.3.1 Ancestry

In order to reduce the chances of population stratification or heterogeneity between the case and control samples it is helpful to derive the case and control sample from the same ancestral population. Ancestry criteria for participation in our study had been set out a priori: Volunteers had to be of British, English, Scots, Welsh or Irish ancestry. This was defined as having at least three of four grandparents of such ancestry. Subjects with a grandparent of non-western European ancestry were excluded. All volunteers were screened with the UCL ancestry questionnaire. The validity of using

this questionnaire had previously been assessed within UCL. This approach applied to cases and controls in AD, SZ and BP samples. Clearly some subjects may be unaware of their lineage, and thus ancestral outliers were removed during quality control.

9.3.2 UCL alcohol dependence research sample

The AD sample consisted of 976 research subjects sampled in two cohorts and all of them gave informed signed consent. The first collection of subjects was recruited from the Centre for Hepatology at the Royal Free Hampstead NHS Trust between 1997 and 2008. The second wave of samples were recruited from several community substance misuse services including the Bexley Substance Misuse Service, South London and Maudsley Alcoholism Service, East Herts Community Drug Action Team and the Max Glatt Unit. We then extended recruitment UK-wide through the UK Clinical Research Network (UKCRN) under the UK-COGA project banner, conceptualized by Prof Hugh Gurling.

All the patients fulfilled the criteria for a diagnosis of alcohol dependence using DSM-IV (APA 1994) or ICD-10 (WHO 1993). None of the patients was currently misusing illegal drugs. The selection of the alcohol dependence subjects was also based on a strict ancestry questionnaire described in section 9.3.1

The control subjects were recruited from London branches of the NHS blood transfusion service, from family doctor clinics and from university student volunteers. All control subjects were interviewed with the Schedule for Affective Disorders and Schizophrenia (SADS-L) (Spitzer & Endicott 1977) to exclude all lifetime psychiatric disorders including alcohol dependence according to Research Diagnostic Criteria (RDC) and Diagnostic and Statistical Manual, III Edition Revised

(DSMIIIIR) criteria. Control subjects were excluded if they reported drinking above the upper limit for safe drinking of 21 units per week for males and 14 units for females as defined by the Royal College of Physicians (Royal College of Physicians, 1987). The control subjects were further selected on the basis of not having a family history of bipolar disorder, schizophrenia or alcoholism. Use of controls is discussed further in Section 9.7 below.

There was a separate EEG-based analysis in the GABRA2 study led by my co-author Sara Montagnese. I did not consider that I contributed sufficiently to that aspect of the work to include it in my thesis.

9.3.3 UCL bipolar research sample

The UCL bipolar research sample consisted of 506 BP I cases and 510 control comparison subjects (Sklar et al. 2008; McQuillin et al. 2011; Dedman et al. 2012). The sample of normal comparison subjects were screened for the absence of psychiatric disorders and with no known first-degree relatives with a psychiatric disorder.

The cases were recruited from National Health Service (NHS) psychiatric services and from volunteers who were members of a support organization for sufferers of bipolar disorder living in South East England. The screened control subjects were recruited from London branches of the National Blood Service, from local NHS family doctor clinics and from university student volunteers. All case and control subjects were interviewed using the Lifetime Version of the Schizophrenia and Affective Disorders Schedule (SADSL) (Spitzer 1977). This information was supplemented by material from case-notes. Diagnoses were assigned using the Research Diagnostic Criteria (RDC) (Spitzer et al. 1978). Diagnoses were always assigned prior to genotyping and were made at the

“probable” level of the RDC. A second psychiatrist reviewed the diagnoses. Most of the cases had BPI (88%). All the bipolar research subjects were also rated with the 90-item OPCRIT checklist (McGuffin et al. 1991). Information regarding family history of mental disorder and drug treatment response was also collected. Phenotypic information was collated in a Microsoft Access database. The sample has been used for several previous allelic and haplotypic association studies as well as GWAS of bipolar disorder (Bass et al. 2009; Kandaswamy et al. 2013; Ferreira et al. 2008; McQuillin et al. 2011; Nyegaard et al. 2010; Dedman et al. 2012).

The use of microarrays with up to 500,000 SNP markers in the GWAS allowed for an analysis of the genetic stratification between UCL cases and controls, which showed that no correction for population stratification was necessary (Ferreira et al. 2008; Sklar et al. 2008).

9.3.4 UCL schizophrenia research sample

The UCL schizophrenia research sample consisted of 617 affected research subjects recruited from NHS psychiatric services. The subjects were selected based on a strict ancestry questionnaire used for selecting UCL BP and AD cases described in section 3.1.1.1. All subjects signed an approved consent after reading an information sheet. All the schizophrenia cases were selected for having an ICD-10 (WHO 1993) diagnosis of schizophrenia recorded in the NHS medical case-notes. Research subjects with short-term drug-induced psychoses, learning disabilities, head injuries and other symptomatic psychoses were excluded. All the affected research subjects were interviewed by a psychiatrist using the Lifetime Version of the SADS-L (Spitzer 1977). The cases were also rated using the 90-item OPCRIT checklist (McGuffin et al. 1991). SZ subjects were then chosen on the basis of having received a diagnosis at the "probable level" of the Research Diagnostic Criteria (RDC) (Spitzer et al. 1978). Patients with schizoaffective bipolar disorder were not included.

9.4 GENOTYPING

DNA was extracted from whole frozen venous blood using a cell lysis, proteinase K, phenol/chloroform method (Sambrook et al. 1989).

9.4.1 AD sample

For the AD sample, we genotyped two batches (total 586) of AD cases and 603 controls (Case Batch 1 = 526, Control 1 = 590; Case Batch 2 = 586, Control 2 = 603). Genotyping success rates were 90% or better. For the GABRA2 study (results in Chapter 11), seven SNPs were selected from the literature as having previously been associated with alcoholism in GABRA2 and genotyped (Table 9.1.)

Table 9.1 GABRA2 SNPs genotyped

#	Marker	Edenberg et al 2004 Alcohol Dependence (p)	Edenberg et al 2004 EEG-b (p)	Covault et al 2004 Alcohol dependence (p)	Soyka et al 2007 Alcohol dependence (p)
1	rs548583	.012*	.028*	-	-
2	rs279871	.0004*	.049*	-	-
3	rs279863	.017*	.011*	-	-
4	rs279858	.0087*	0.22	0.009*	0.79
5	rs279836	.0071*	.0066*	-	-
6	rs279841	.038*	.018*	-	-
7	rs9291283	-	-	0.06	0.006*

9.4.2 BP sample

Genotyping was performed using the Affymetrix Gene Chip Human Mapping 500K Array Set, a first generation genotyping platform without specific copy number probes. This set is comprised of two high-density arrays, NspI and StyI. All genotyping was performed by the Genetic Analysis Platform at the Broad Institute of Harvard and MIT using standard protocols (Matsuzaki et al. 2004) as previously described (Saxena et al. 2007). Genotypes were called using the Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) (Rabbee & Speed 2006). A panel of 24 markers present on the whole genome product as well as 25 SNPs previously genotyped in the UCL samples were used as genetic fingerprints to detect sample switches.

9.4.3 SZ sample

SZ samples were also genotyped using the Affymetrix Mapping 500K Array and the Genomewide Human SNP Array 5.0 as described elsewhere (Craddock & Sklar 2009; Ferreira et al. 2008; International Schizophrenia Consortium 2008). Genotypes were called from raw intensity data using the Birdseed component of the Birdsuite algorithm (McCarroll et al, 2008; Korn et al, 2008). Controls from UCL were previously genotyped at The Broad Institute as part of a published genome-wide association study of bipolar disorder (Sklar et al 2008).

9.5 DATA ANALYSIS AND QUALITY CONTROL (QC)

BP and SZ: QC was completed using SNPs to remove duplicate samples, poorly genotyped and/or contaminated samples, as part of standard quality control metrics used for whole-genome association studies (Sklar et al 2008). This included individual exclusions (based on call rates <85%;

heterozygosity; outliers who appeared to be close relatives, sample duplications or non-Caucasian) and marker exclusions (call rate <95%; minor allele frequency (MAF) <1%; Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$ in controls and differential rates of missing genotypes between cases and controls, using Fisher's exact test, $p < 1 \times 10^{-3}$). Because genotypes were called on a per-plate basis, we identified "bad plates" where SNPs showed grossly different allele frequencies to all other plates ($p < 1 \times 10^{-10}$) and removed those SNPs. Following the exclusion there remained 372,193 SNPs for further analysis.

9.5.1 Population Stratification and final dataset

To control for possible effects of population stratification within panel, we matched cases and controls based on the proportion of alleles shared identical-by-state (IBS), using complete linkage hierarchical clustering. Within panel, individuals were clustered to ensure a) at least 1 case and 1 control in clusters with 2 or more individuals; b) that no two individuals in the same cluster fail the population pairwise concordance test (PPC) with a $p < 1 \times 10^{-3}$.

IBS clustering was then performed which indicated that the vast majority of individuals belonged to a single cluster; individuals who did not were removed, resulting in a final dataset of:

- i) 143 BP subjects and comorbid AD (BPALC) vs 510 SNC; and BPALC vs 367 BP patients without AD (BPnonALC).
- ii) 77 subjects with schizophrenia and comorbid AD (SZALC) vs 510 SNC; and vs 251 SZ without AD (SZnonALC).
- iii) Meta-analysis of 220 BPALC+SCZALC vs 7510 SNC; and vs 751 BPnonALC+SCZnonALC.

9.6 ASSOCIATION ANALYSES

9.6.1 Primary Whole Genome Association Analysis

The primary analysis was of single SNPs using the Cochran-Mantel-Haenszel (CMH) test to assess allelic association with disease conditional on the strata as defined by the stratification analysis. For each SNP, we also calculated standard, allelic association tests not conditioning on strata, based on a chi-squared test for independence. Because of the quality control steps outlined there was no genomic inflation (our lambdas were effectively 1) and it was not necessary to condition the allelic tests by population strata. In total, we performed 372,193 single SNP tests, 340,925 unique multi-marker tests of association with bipolar disorder, SZ and AD.

9.6.2 Testing of allelic association

I used PLINK (Purcell, 2007) is a free, open-source whole genome association analysis toolset for GWA analyses. Association between a SNP and a phenotype is generally accepted to be genome-wide significant at $\alpha = 0.05$ if the p value is 7.2×10^{-8} or smaller, as this corrects for the total number of independent tests (Dudbridge and Gusnanto, 2008).

Use of control groups

I used two approaches to controls: the first was to use supernormal controls (SNC) – a common approach in many GWAS. The control subjects were selected on the basis of not having a family history of bipolar disorder, schizophrenia or alcoholism. Thus the controls were termed “supernormal controls” (SNC) or “hypernormal controls” due to their low phenotypic and genetic loading for

alcoholism and severe mental illness. It has been argued that use of “hypernormal controls” is more efficient (Morton and Collins 1998). Epidemiologists however would argue that such controls risk introducing new biases (e.g. through exclusion criteria that are different from, and stricter than, those applied to cases), that they do not necessarily allow straightforward epidemiological inferences, and they require more resources in recruitment and screening (Moskvina et al. 2005; Schwartz & Susser 2011).

The second approach was to use “severe mental illness without AD” as the control group. This second control group comprised of BP (vs BPALC), SZ (vs SZALC) or (BP+SZ) vs (BPALC + SZALC). The aim of this approach was to increase the power to detect a difference in the light of the small sample sizes by modern standards; and also to try and minimize the differences between cases and controls in order to elicit the phenotypes of interest – in this case AD (or AD in the context of severe mental illness).

There is scientific precedent for the use of disease controls other than normal controls. Curtis et al’s work in our group in compared SZ with BP as controls (Curtis et al. 2011). Using this approach Curtis et al reported a suggestive marker at the calcium channel gene *CACNG5* ($p = 10^{-6.4}$).

BP and SZ are increasingly understood to share overlapping genetic risk (Purcell et al. 2009; Ruderfer et al. 2014). A recent large analysis of SZ versus BP revealed new loci contributing to our understanding of overlap and uniqueness of the disorders (Ruderfer et al. 2014). Ruderfer commented that combining diseases with similar genetic risk profiles improves power to detect shared risk loci; and that future direct comparisons of BP and SZ are likely to identify loci with significant differential effects. By identifying these loci our understanding of how these diseases differ biologically should improve.

Curtis et al (2011) consider that there are theoretical advantages for this approach including that i) the samples might be better matched epidemiologically leading to patients suffering from either disease having similar genetic and non-genetic backgrounds; ii) recruitment will be from similar settings rather than the usual approach of recruiting unaffected controls from non-clinical settings; iii) having well-matched samples improves the signal-to-noise ratios and should improve both the power and specificity of the association studies; iv) case-case comparisons may “cancel out” common genetic vulnerability and highlight risk alleles unique to each disease, and thus actually reduce shared risk confounders.

To my knowledge this “AD + mental illness versus mental illness controls” comparison combination has not been published before (although Kendler has published on a combined AD + SZ phenotype versus normal controls (Kendler et al. 2011). I thus consider this a novel approach to try and “uncover” any underlying AD genes (or genes which may increase risk of developing comorbid AD) from “pure” BP/SZ/BP+SZ genes.

9.7 GENE-BASED TESTING

Genes are functional groups of nucleotides that code for proteins. Depending on the underlying genetic architecture, the gene-based analyses can be more powerful than traditional individual-SNP-based GWAS (Liu et al. 2010). For example, if a gene contains more than one causative variant, then several SNPs within that gene might show marginal levels of significance that are often indistinguishable from noise in the initial GWAS results. A gene-based test might be better able to detect these effects by combining the effects of all SNPs in a gene into a test-statistic and correcting

for LD. A gene-based test thus detects genes which show a greater signal of association than expected by chance given the LD between the SNPs and the gene's length. The relevance of the gene-based test depends on the underlying genetic architecture of gene which is unknown and which is expected to differ between genes.

In order to determine if there are genes with a surplus of associated variants, we conducted a gene-based test using the Versatile Gene-based Association Study-2 (VEGAS2, <https://vegas2.qimrberghofer.edu.au/>) (Liu et al. 2010). The VEGAS2 test summarises evidence for association on a gene by gene basis by considering the p-value of all SNPs within genes, while accounting for linkage disequilibrium (LD) and number of SNPs per gene. It then compares the resulting gene-based test statistic to a large number of simulated chi-square distributed gene-based test statistics. VEGAS2 uses 1000 Genomes Project (<http://www.1000genomes.org/>) populations to estimate patterns of LD for each gene, and we specified the British and Scottish sub-populations to match our research subjects. The p-value resulting from this gene-based test is thus the proportion of simulated test statistics that exceed the observed test statistic.

9.8 CORRECTION FOR MULTIPLE TESTING

To determine which SNPs are truly associated with a trait, we need some measure of how often we would get this result given how many tests have been performed. One simple way to do this is using Bonferroni correction, which is simply dividing by the number of independent statistical tests performed. The human genome contains roughly 1 million haplotype blocks and therefore, even if one tests 6 million SNPs for association with a trait, that leads to about 1 million independent statistical tests. Therefore it is generally held that the threshold for genome-wide significance is 5×10^{-8} , which is a nominal threshold of $p = 0.05$ Bonferroni-corrected by dividing by 1 million.

For gene based testing, VEGAS2 tests for 15,000 known genes, thus our gene-based significance threshold was set at a Bonferroni-corrected $p = 0.05/15000 = 3.3 \times 10^{-6}$ (Verweij et al. 2010).

9.9 PATHWAY-BASED ANALYSIS WITH VEGAS2 AND PANTHER

GWAS often highlight SNPs with the smallest p-values, paying less attention to lower ranked SNPs. Analysis of those gene pathways may contain valuable information that can enable identification of additional associations (Ngwa et al. 2011). Gene-based results also enable the subsequent use of network or pathway approaches to interpreting the findings from GWAS (Wang et al. 2007; Elbers et al. 2009; Jin et al. 2014). In our study all genes with a p-value ≤ 0.01 from the VEGAS2 gene-based analysis were used as input to perform a pathway analysis by VEGAS2Pathway (<https://vegas2pathway.qimrberghofer.edu.au/>).

The VEGAS2Pathway approach aggregates association strength of individual markers into pre-specified biological pathways. It accounts for gene size and linkage disequilibrium between markers using simulations from the multivariate normal distribution. Pathway size is taken into account via a re-sampling approach. The current implementation uses gene-sets from the gene ontology (GO), curated gene-sets from MSigDB (containing canonical pathways and gene-sets from BIOCARTA, REACTOME, KEGG databases), and PANTHER and pathway commons databases, enabling analysis of a wide range of complex traits. The designers of VEGAS2 applied this method to a colorectal cancer GWAS meta-analysis data set (10934 cases, 12328 controls) from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) which implicated a number of biologically plausible pathways (Mishra & McGregor, paper in press, by personal communication).

All genes with a p-value ≤ 0.01 from the VEGAS analysis were used as input to perform separate pathway analysis by PANTHER. PANTHER is a publically-available database having Gene Ontology (GO) and functional annotation information. GO Pathways are available within the database and these pathways were used in our analysis to estimate pathways and molecular functions of nominally associated genes.

9.10 THE LITERATURE AS CONTROL

There have been no previous GWA studies on BPALC or SZALC as comparators for replication. Because AD was the main phenotype of interest, I carried out a systematic search for published association studies with AD and addiction phenotypes and endophenotypes, and plausibly linked phenotypes, where possible using the standards described by the NCI-NHGRI Working Group on Replication in Association Studies (Chanock et al. 2007). I cross-referenced each of our tentative SNP and gene signals with the published literature (final search 1 November 2015). My database included results from dbGAP (<http://www.ncbi.nlm.nih.gov/gap/>), NHGRI-EBI Catalog (www.ebi.ac.uk/gwas/), OMIM, GADB, KARG, SNPedia, NCBI, EBI, DiscGeNET, GWASDB, HUGE database, and PubMed). I also manually searched the literature for selected genes and SNPs of interest. In total there were 354 unique gene symbols (including variations on gene names for maximum coverage) and nominally associated SNPs from across the literature in the database. This data was combined into a Microsoft Access database, cleaned, and the JOIN command used to search within each GWAS output for matching gene names, identifiers or SNPs.

9.11 GRAPHICAL TECHNIQUES

For Q-Q and Manhattan plots, I used the R package qqman (Turner 2014) available from the Comprehensive R Archive Network (<http://cran.r-project.org/package=qqman>). I used HapMap (CEU) for LD mapping and LocusZoom (Pruim et al. 2010) to plot relevant regional association results. LocusZoom plots show not only the magnitude of association for each SNP, but also the pairwise LD pattern with the most strongly associated SNP or another user-specified SNP. Quick inspection can reveal the extent of the associated region and the location and number of SNPs in strong LD with the index SNP.

I also used Ricopili from the Broad Institute to plot comparison regional loci from the Psychiatric Genetics Consortium. LD (r^2 or D) estimates used by LocusZoom and Ricopili came from the 1000 Genomes Project. The next chapters summarise my results.

CHAPTER 10. RESULTS OF GABRA2 CANDIDATE GENE ANALYSIS IN AD

In the present part I sought to clarify the role of the GABA receptor alpha-2 gene (GABRA2) in alcoholism using seven SNPs used in previous AD/alcoholism association studies.

All markers were in Hardy Weinberg equilibrium (HWE) in the controls. HWE for marker rs279841 in the AD samples was $p=0.0199$ and combined $p=0.0166$. Linkage disequilibrium patterns (Figure 10.1B) appear to be very similar to that observed in the HapMap CEU data (Figure 10.1A.)

None of the selected SNPs showed allelic or haplotypic association with AD in our white UK sample (Table 10.1). This becomes a further negative study for GABRA2 and AD and was published in 2011 (Lydall et al. 2011).

10.1 GABRA2 SNP STUDY DISCUSSION

This case control study of AD cases and ancestrally matched controls found no evidence for allelic or haplotypic association with selected GABRA2 SNPs previously implicated in alcoholism association studies. This would be the sixth negative association study for this gene, while there are also five positively associated studies, all with varying methodological strengths and weaknesses.

In previous studies, GABA system markers have been associated with a variety of phenotypes in a number of different contexts, many of which were not assessed in this study. For example, interaction may occur within and between GABA system genes, particularly among proximal or clustered genes (Uusi-Oukari et al. 2000).

There is evidence that the influence of polymorphisms in the GABA system may vary with age or across developmental stages (Dick et al. 2006; Dick et al. 2009), and be moderated by environmental factors (Enoch et al. 2010; Dick et al. 2009). Markers in GABA system genes have also been associated with less-complex biological markers such as Beta-frequency EEG (Edenberg et al. 2004) and event-related potentials (Winterer et al. 2000) that meet the criteria to be considered endophenotypes more directly reflecting underlying genetic liability than their complex behavioural correlates (Begleiter & Porjesz 2006).

The absence of association in the present study does not invalidate previous positive genetic associations between GABRA2 and alcoholism. This is common for human genetic association studies where extensive locus heterogeneity is present with relatively low frequency susceptibility alleles. Other explanations include unknown bias in sample collection or phenotype heterogeneity. In our study, one weakness is that psychiatric comorbidity was not assessed beyond the four commonest co-occurring disorders (BP, depression, ADHD and antisocial personality disorder). Those with AD + comorbidity were not excluded due to high prevalence and the potential impact on sample size and thus power. The UCL team are about to publish evidence for phenotypic variation contributing towards the non-replication of genetic association studies. This occurred in an alcohol dependence sample using an association study of the ZNF699 gene in a larger well-characterised UCL AD sample with low levels of psychiatric and addictive comorbidity (Ali et al. 2015).

The strengths and weaknesses of case control studies have been discussed in Chapter 6, and all are possibly relevant here. However in our defence, the samples were well characterised, ethnically homogeneous, we chose a severe phenotype (alcohol dependence) and this was a large sample size for a

relatively small number of markers, so it likely not to have been underpowered. Ancestral heterogeneity was reduced by the use of the ancestral questionnaire and QC removing non-Caucasian samples, and confounding further reduced by the use of well characterised AD samples and supernormal controls. I consider it important to publish negative studies to reduce publication bias and contribute to the data for future meta-analyses.

Future work in this direction may wish to ensure larger samples, or meta-analyses, and optimize the quality of AD phenotypes, also excluding as much potentially confounding psychiatric comorbidity as possible by optimising clinical phenotype data.

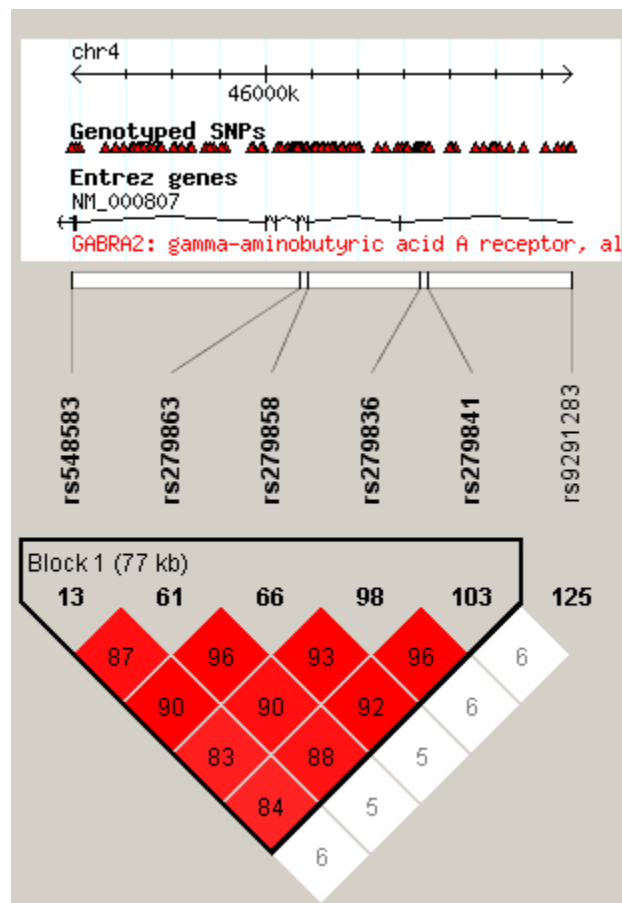
The next chapter reports results of the first GWAS, that of BPALC vs BPhonALC.

Table 10.1. GABRA2 SNP results in the UCL AD sample

#	SNP Name	Batch number	Numbers		Frequencies		Chi square	P value
			Case	Control	Case	Control		
1	rs548583	1	474	549	0.44	0.419	0.910	0.34
2	rs279871	2	524	568	0.442	0.43	0.331	0.57
3	rs279863	2	528	552	0.438	0.432	0.089	0.76
4	rs279858	1	471	582	0.427	0.425	0.005	0.95
5	rs279836	1	496	564	0.572	0.567	0.038	0.85
6	rs279841	1	454	552	0.437	0.425	0.313	0.58
7	rs9291283	2	540	569	0.243	0.227	0.778	0.38

Figure 10-1 GABRA2 comparison of A. HapMap CEU Linkage Disequilibrium (LD) and B. LD from the UCL AD data.

A:



B:

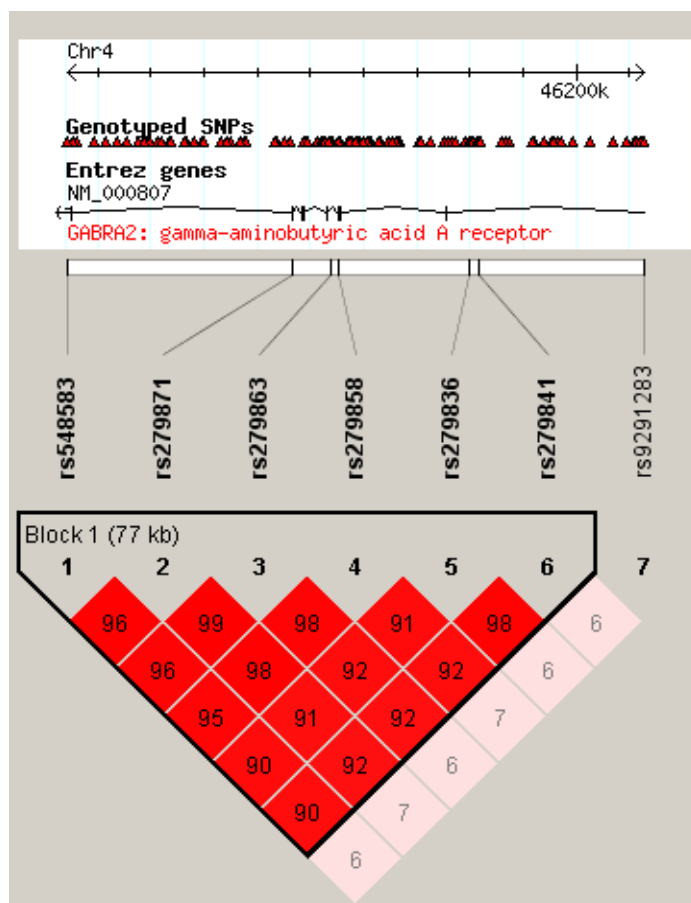


Figure legend: D' is shown by grades of colour, the numbers represent r^2 .

CHAPTER 11. RESULTS OF BPALC vs BPNONALC GWAS

For this analysis of BPALC vs “pure” BP (without AD, i.e. BPnonALC), we examined 143 subjects with bipolar disorder and comorbid AD (BPALC) vs 350 with BPnonALC as controls. We analysed 372,193 SNPs genotyped on the Affymetrix GeneChip Human Mapping Array in 506 BP patients and 510 controls following the quality control and data cleaning procedure described in Methods, above. The overall genotyping call rate for analysed SNPs was 99.4%. Including the multimarker predictors, our data set is estimated to capture 78.7% of common variation (SNPs with minor allele frequency $\geq 5\%$) in the CEPH HapMap (CEU) samples with $r^2 > 0.8$.

No SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p-values for the autosomal associations in relation to the expected p-values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 11.1. A Manhattan plot of our association analyses is shown in Fig.11.2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p-values are presented in Table 11.1.

I then checked whether any of the 40 SNPs with the lowest p-value for each analysis were in or close to a gene of known relevant function. None of the top SNPs were previously related directly to AD or alcoholism-related phenotypes, however a number of genes in or near these SNPs are briefly reported.

Our strongest albeit non-significant SNP signal was near Shugoshin-Like 1 (*SGOL1*) which has been implicated in ethanol-related methylation, and as a risk gene for substance misuse (see Discussion for detail). Figure 11 -3 shows the LocusZoom plot of the strongest SNP and its relationship with the nearest gene SGOL1 - there is no evident LD between the two.

Our second highest-ranked SNP association was near Basic Leucine Zipper Transcription Factor, ATF-Like (*BATF*) which has been associated with multiple sclerosis (Sawcer, 2011) and rheumatoid arthritis (Stahl, 2010). Multiple high-ranked SNPs in gene *GADLI* (6 SNPs in the top 40, Table 11.1) invited further investigation. However gene based analysis was not significant, and a LocusZoom regional plot shows no LD between the highest ranked SNP and the gene (Fig. 11-4).

Of top SNP-implicated genes, only one (*SEMA5A*) was previously reported in linked phenotypes: autism (Weiss et al 2009) and smoking cessation (Uhl et al, 2008; Rose et al 2010). In our gene-based test the *SEMA5A* gene had an uncorrected gene wise p-value of .007 (ranked 6th gene by wise p). I have plotted *SEMA5A* markers from UCL (Fig. 11-5) and PGC Bipolar GWAS (Fig. 11-6) for visual inspection of signal in the region.

In the gene-based tests, we found 255 genes with $p < 0.01$, strongest $p = 5 \times 10^{-4}$. Cross referencing with the literature I noted 8 genes that were nominally associated ($p \leq 0.01$) that were in common with our VEGAS2 gene-based list. The top 40 of these are listed in Table 11.2.

Aside from *SEMA5A*, the most notable gene wise result was the fifth-ranked (by gene based p-value) alcohol metabolism gene *ADH5* $p = 5.07 \times 10^{-3}$ (Fig. 11-7). *ADH5* has been implicated by region in a GWAS of AD in a Korean sample (Park, 2013) and also in AD symptom count in an African American sample (Kendler et al. 2011). Another top 40 SNP signal of interest is the cell-cell contact and scaffolding gene, discs large homolog 5 (*DLG5*) ($p = 4.26 \times 10^{-4}$), a gene that has previously been associated with variation in liver enzymes (Chambers et al. 2011) and childhood obesity (Comuzzie et al. 2012). I have also plotted gene regions around *CDH13* (gene wise non-significant, Fig. 11-8) because of numerous highly ranked SNPs in this gene in our GWA and previous strong literature associations with addictions and AD (Johnson et al. 2010; Uhl & Drgonova 2014; Treutlein et al. 2009).

All gene replications of AD or related phenotypes and the top SNPs in that gene from our data are listed in Table 11.3.

Pathway analysis using VEGAS revealed genetic pathways involving adenylate cyclase activity, tryptophan metabolism, solute carriers, voltage gated ion channels, and amino acid modification (Table 11.4).

PANTHER analysis of our top genes implicate in gene based analysis ($P < 0.001$, $n = 255$) revealed binding, catalytic, nucleic acid binding and receptor activity among the main molecular pathways (Fig. 11-9).

In summary, this was the first small GWAS comparing patients with BP and comorbid AD with patients with BP without AD in an attempt to tease out the AD genetic effects. No marker achieved genome-wide significance. However gene wise and pathway approaches offer a number of biologically plausible genes involving alcohol metabolism, cell signalling, and cell scaffolding.

11.1 DISCUSSION

No SNP achieved conventional genome wide significance levels. Our strongest GWA SNP was not in LD with its nearest gene despite SGOL1 being a risk gene for substance misuse. Our second strongest SNP association was near BATF, but gene based analysis was not even nominally significant. This is probably due to small sample size. Of SNP-implicated genes, only one (SEMA5A) was previously reported in linked phenotypes: autism (Weiss et al 2009) and smoking cessation (Uhl et al, 2008; Rose et al 2010). In our gene-based test the SEMA5A gene had an uncorrected gene wise p-value of .007 (ranked 6th gene by wise p). This makes it a gene of interest but not involvement in AD in this study.

The tentative involvement of Tubulin tyrosine ligase (TTL) on chromosome 2 (uncorrected gene-based $p = 5.83 \times 10^{-3}$) is interesting because TTL has been associated with AD, conduct disorder and suicide attempts (Dick et al, 2010). Arguably, some of the features of bipolar overlap with AD, conduct disorder and self-harming behaviour. Perhaps TTL affects antisocial or risky behaviour in both people with BPALC and those with AD and other externalizing behaviours.

A possible signal near the alcohol metabolism gene ADH5 (uncorrected gene-based $p = 5.07 \times 10^{-3}$) is interesting. ADH5 is not a leading alcohol metabolising gene in the AD literature, and in fact is remarkably ineffective in oxidizing ethanol, but it readily catalyzes the oxidation of long-chain primary alcohols and glutathione. It has however previously been implicated by region in a GWAS of AD in a Korean sample (Park, 2013), and AD symptom count in an African American sample (Kendler, 2011). Unless there is an as yet undiscovered ethanol pathway it is unclear how a gene which metabolizes formaldehyde amongst others could differentiate AD from BP. More encouraging is another top 40 SNP signal of interest is the cell-cell contact and scaffolding gene, discs large homolog 5 (DLG5) (uncorrected gene-based $p = 4.26 \times 10^{-4}$), a gene that has previously been associated with variation in liver enzymes (Chambers, 2011) and obesity (Comuzzie, 2012). One would expect a gene specific to alcohol biology to stand out in any study of AD genetics. It makes intuitive sense for the main difference between people with BP and AD and people with BP alone would be related to alcohol metabolism (including by liver enzymes) amongst others.

Broadly, cadherins are adhesion molecules that mediate Ca^{2+} -dependent intercell adhesion (Ivanov, Philippova, & Tkachuk, 2001). Mapped to 16q24 in the human chromosome, the CDH13 gene is anchored in the membrane via glycosylphosphatidylinositol (GPI). It has been found throughout the central nervous system (including cerebral cortex, midbrain, and medulla), and is believed to play a role in maintaining neural circuitry (Ivanov et al. 2001; Takeuchi et al. 2000). CDH13 as a “cell adhesion

molecule” is well positioned to play roles in regulation of synaptic connectivities that are of interest for brain disorders that include addictions and ADHD (Treutlein et al. 2009; Rivero et al. 2013). GWAS comparing substance-dependent vs control and successful vs unsuccessful abstainers from smoking have each identified SNPs in the 3’ region of the cadherin 13 (CDH13) gene (Uhl et al. 2008). CDH13 variants interacting with environment have been found for age of first drink (rs4389131 x parental monitoring), and another for age of first intercourse (rs12926331 x peer drinking). Thus CDH13 may have a role in initiating risky behaviour (whether in the context of affective disorder or alone) which may be why CDH13 is considered an addictions candidate gene. Early risk behaviour could be a confounder variable and actually represent repeated use of alcohol from a young age which itself is a risk factor for developing AD. The cadherins as a group and CDH13 in particular are worthy of further investigation.

In the next chapter I report on the BPALC vs supernormal controls analysis.

Figure 11-1 Q-Q plot of BPALC vs BP GWAS p-values

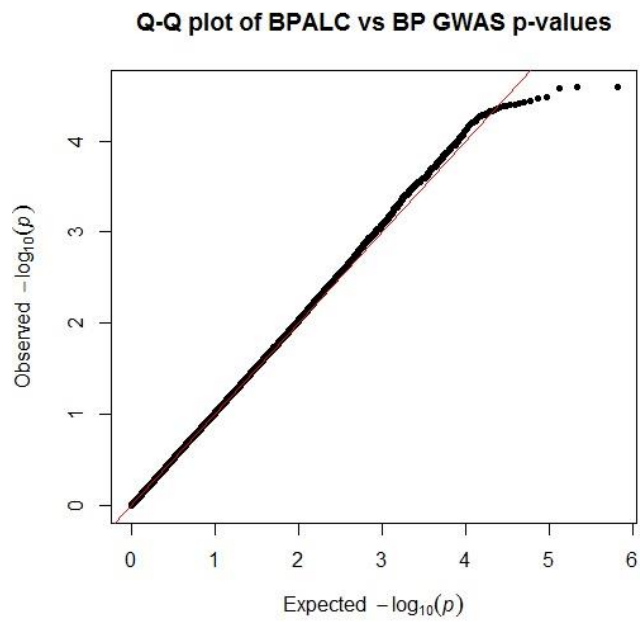


Figure legend: Q-Q plots of observed and expected $-\log_{10}(P)$ of the associations between SNPs

Figure 11-2 *Manhattan Plot of BPALC vs BPnonALC GWAS.*

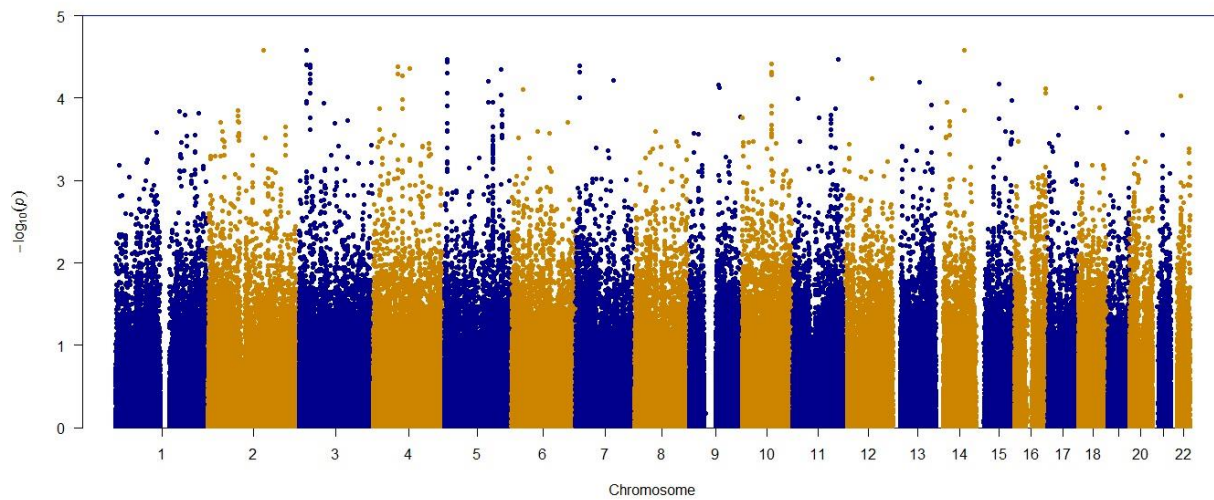


Figure legend: Results of the genome-wide association analyses for BPALC vs BPnonALC. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Table 11.1 BPALC vs BPNALC: Top 40 SNPs, gene wise results and relevant replication studies

CH R	SNP	BP	A 1	F_A	F_U	A 2	CHIS Q	P	OR	Position	Distance	Gene code	Gene based p*
3	rs626491	20514267	T	0.2425	0.3866	G	17.69	2.60E-05	0.508	downstream	34869	SGOL1	---
14	rs10147012	76019185	T	0.1857	0.09	C	17.69	2.60E-05	2.306	downstream	5827	BATF	7.72E-04
2	rs7591997	149589774	C	0.06071	0.01289	T	17.66	2.64E-05	4.948	upstream	6370	---	---
11	rs34143639	122404194	G	0.05036	0.00862	A	17.21	3.35E-05	6.098	downstream	45719	---	---
5	rs2256569	9160380	C	0.1893	0.3214	T	17.19	3.39E-05	0.4929	intron	0	SEMA5A	5.95E-03
5	rs988593	9156684	A	0.1714	0.3	G	17.05	3.65E-05	0.4828	intron	0	SEMA5A	5.95E-03
10	rs1268956	79608277	T	0.3786	0.2471	C	16.96	3.81E-05	1.856	intron	0	DLG5	4.26E-04
3	rs595551	20516530	A	0.25	0.3886	C	16.9	3.93E-05	0.5245	downstream	32606	---	---
3	rs1683768	31074131	C	0.3357	0.4799	T	16.88	3.98E-05	0.5476	upstream	137874	GADL1	1.82E-01
7	rs218971	10986882	C	0.4781	0.3362	G	16.87	4.01E-05	1.809	upstream	26617	PHF14	1.54E-02
4	rs1510738	67500648	G	0.2107	0.1103	A	16.83	4.10E-05	2.153	downstream	203409	---	---
3	rs1683760	31089530	A	0.3321	0.4757	G	16.78	4.21E-05	0.5481	upstream	153273	GADL1	1.82E-01
3	rs294306	31079692	T	0.3321	0.4756	A	16.75	4.27E-05	0.5483	upstream	143435	GADL1	1.82E-01
4	rs17595186	99988713	A	0.3551	0.227	G	16.71	4.37E-05	1.875	exon	0	---	---
5	rs304855	152759013	G	0.277	0.4171	A	16.64	4.52E-05	0.5353	upstream	110162	GRIA1	9.21E-01
7	rs1079478	10989364	A	0.4714	0.3324	G	16.54	4.77E-05	1.791	upstream	24135	PHF14	1.54E-02
10	rs1248631	79582881	T	0.3786	0.2486	C	16.53	4.78E-05	1.842	intron	0	DLG5	4.26E-04
10	rs1248657	79629127	C	0.375	0.2457	T	16.49	4.90E-05	1.842	UTR-3	0	LOC100131132	2.00E-02
5	rs2526122	9158180	C	0.1727	0.2994	T	16.47	4.94E-05	0.4883	intron	0	SEMA5A	5.95E-03
4	rs10517989	67500915	G	0.2107	0.1114	A	16.39	5.14E-05	2.129	downstream	203676	---	---
3	rs182401	31081125	T	0.3321	0.4741	G	16.39	5.14E-05	0.5516	upstream	144868	GADL1	1.82E-01
10	rs1659390	79625407	T	0.3786	0.2493	C	16.35	5.27E-05	1.835	intron	0	DLG5	4.26E-04
4	rs10857059	79854436	T	0.2143	0.1143	C	16.33	5.31E-05	2.114	intron	0	PAQR3	1.51E-03
12	rs1402312	78234525	A	0.02857	0.00143	G	16.19	5.72E-05	20.56	intron	0	NAV3	6.73E-01
3	rs294278	31127911	T	0.3321	0.4728	C	16.1	6.00E-05	0.5546	upstream	191654	GADL1	1.82E-01
7	rs2950183	102020363	G	0.3643	0.2378	T	16.07	6.12E-05	1.836	intron	0	PRKRIP1	4.17E-03
5	rs13154208	119318747	A	0.2679	0.1571	G	16.01	6.31E-05	1.962	downstream	301972	---	---
13	rs9542547	71527068	T	0.3214	0.2014	A	15.99	6.38E-05	1.878	upstream	62205	LINC00348	4.79E-01
3	rs1395162	31092371	C	0.3357	0.4757	A	15.94	6.53E-05	0.557	upstream	156114	GADL1	1.82E-01
15	rs2456470	62777956	G	0.3143	0.1957	C	15.9	6.69E-05	1.884	intron	0	TLN2	7.34E-01

Table 11.1 contd...

9	rs12001902	78715080	T	0.1331	0.05714	G	15.82	6.95E-05	2.533	intron	0	PCSK5	3.25E-02
9	rs2047995	81169950	C	0.3237	0.204	T	15.69	7.46E-05	1.868	downstream	224941	PSAT1	6.39E-01
16	rs4360919	83616310	T	0.1071	0.2157	C	15.66	7.60E-05	0.4363	intron	0	CDH13	1.33E-01
6	rs707937	31731014	G	0.2821	0.17	C	15.6	7.82E-05	1.919	intron	0	SAPCD1	6.69E-03
3	rs172421	31084868	G	0.3393	0.4771	C	15.42	8.62E-05	0.5629	upstream	148611	GADL1	1.82E-01
5	rs1013721	9088165	T	0.2036	0.33	G	15.41	8.67E-05	0.519	intron	0	SEMA5A	5.95E-03
16	rs7499397	83620469	C	0.4607	0.3271	T	15.39	8.75E-05	1.757	intron	0	CDH13	1.33E-01
5	rs304841	152778918	C	0.2464	0.3775	T	15.28	9.26E-05	0.5392	upstream	90257	GRIA1	9.21E-01
22	rs9613519	28110312	C	0.464	0.3299	T	15.26	9.36E-05	1.758	intron	0	---	---
7	rs218967	10988142	T	0.4784	0.3443	A	15.16	9.89E-05	1.747	upstream	25357	PHF14	1.54E-02

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F_A= Frequency of Allele 1; F_U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. *Uncorrected p <0.05 in **bold**. No gene-based tests survived Bonferroni correction.

Table 11.2 BPALC vs BPnonALC: Top 40 gene based analysis genes

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Uncorrected gene based p	Top SNP	Top SNP p
14	BATF	12	1.00E+06	53.12562	7.72E-04	rs10147012	2.60E-05
5	SEMA5A	164	1.00E+05	340.6102	5.95E-03	rs2256569	3.39E-05
10	DLG5	40	1.00E+06	259.6666	4.26E-04	rs1268956	3.81E-05
7	NDUFA4	18	1.00E+05	83.65992	5.19E-03	rs218971	4.01E-05
4	ADH5	20	1.00E+06	104.1915	9.23E-04	rs17595186	4.37E-05
4	PAQR3	14	1.00E+05	74.32575	1.51E-03	rs10857059	5.31E-05
7	SPDYE6	4	1.00E+06	31.91611	4.99E-04	rs2950183	6.12E-05
6	HSPA1L	2	1.00E+06	15.68911	5.25E-04	rs707937	7.82E-05
22	MN1	23	1.00E+05	77.40639	2.09E-03	rs9613519	9.36E-05
10	DLG5-AS1	10	1.00E+06	61.91778	8.26E-04	rs1149725	1.23E-04
17	SLC26A11	19	1.00E+06	82.65686	9.63E-04	rs2269376	1.32E-04
5	HAVCR1	20	1.00E+06	140.5367	2.40E-04	rs7731951	1.41E-04
11	DYNC2H1	79	1.00E+05	271.3367	4.93E-03	rs2671349	1.58E-04
5	TIMD4	15	1.00E+05	73.62017	2.26E-03	rs9790967	1.59E-04
3	SNORA58	2	1.00E+06	15.15453	5.48E-04	rs9843898	1.88E-04
14	RALGAPA1	16	1.00E+06	131.1212	6.86E-04	rs10147261	1.92E-04
2	TTC27	31	1.00E+05	97.41738	5.67E-03	rs17012300	1.97E-04
2	FASTKD2	17	1.00E+05	71.196	3.22E-03	rs17280449	2.26E-04
15	LOC101929478	11	1.00E+05	54.68458	2.25E-03	rs2292416	2.57E-04
19	ZNF816-ZNF321P	17	1.00E+05	53.29152	6.15E-03	rs17770717	2.57E-04
1	CHIA	56	1.00E+05	157.7963	6.39E-03	rs9645351	2.62E-04
17	TP53I13	6	1.00E+05	21.68943	8.96E-03	rs8081085	2.82E-04
14	INSM2	5	1.00E+05	35.13945	3.11E-03	rs4982292	2.83E-04
4	PI4K2B	17	1.00E+05	81.24482	3.06E-03	rs7661189	3.12E-04
2	HAAO	24	1.00E+05	129.8898	1.76E-03	rs2119014	3.33E-04
10	KIAA1462	33	1.00E+04	102.9859	9.20E-03	rs11007874	3.33E-04
15	ST8SIA2	33	1.00E+05	81.05471	7.13E-03	rs7170557	3.38E-04
10	PITRM1	47	1.00E+05	149.6708	2.73E-03	rs12678	3.43E-04
1	RFWD2	23	1.00E+05	96.15749	4.06E-03	rs791743	3.47E-04
17	RAP1GAP2	33	1.00E+05	75.5678	4.51E-03	rs4790372	3.58E-04
13	SPATA13_1	113	1.00E+05	216.2478	7.62E-03	rs4770587	3.86E-04
7	EGFR-AS1	22	1.00E+05	77.93638	2.67E-03	rs2075110	3.99E-04
8	IDO2	33	1.00E+05	112.979	4.52E-03	rs10504014	4.60E-04
2	DBI	13	1.00E+06	87.71992	6.42E-04	rs1374313	4.92E-04
5	GPR150	4	1.00E+05	14.9653	7.81E-03	rs6556877	5.26E-04
4	ZCCHC4	14	1.00E+05	67.89398	4.97E-03	rs7684466	5.38E-04
9	TAL2	13	1.00E+05	54.60498	4.13E-03	rs7039618	5.94E-04

Table 11.3 BPALC vs BPnonALC gene based replications of AD or related phenotypes from the literature

Chr	Gene symbol	Gene name	n SNPs	n Sims	VEGAS Test result	Uncorrected gene based p	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
10	DLG5	Discs, large homolog 5	40	1000000	259.6	4.26E-04	rs1268956	3.81E-05	Liver enzymes (Chambers, 2011) and obesity (Comuzzie, 2012)
4	ADH5	Alcohol dehydrogenase 5	22	100000	117.9	5.07E-03	rs17595186	4.37E-05	AD (Park, 2013); AD symptom count (Kendler, 2011)
2	HAAO	3-Hydroxy-anthranilate 3,4-dioxygenase	24	100000	129.9	1.76E-03	rs2119014	3.33E-04	AD with conduct disorder and suicide attempts (Dick et al, 2010)
2	DBI	Diazepam binding inhibitor	13	1000000	87.7	6.42E-04	rs1374313	4.92E-04	AD (Waga, 2007)
1	MTHFR	Methylenetetrahydrofolate reductase	20	10000	73.6	9.60E-03	rs2236797	6.46E-04	AD severity, withdrawal symptoms and hyperhomocysteinemia (Saffroy, 2008; Benyamina, 2009; Lutz, 2006, 2007)
3	GHSR	Growth hormone secretagogue Receptor	14	1000000	68.8	6.96E-04	rs509035	2.46E-03	Alcohol intake and body mass (Haplotypic association) (Landgren, 2008)
2	TTL	Tubulin tyrosine ligase	10	100000	45.9	5.83E-03	rs4849074	2.96E-03	AD with conduct disorder and suicide attempts (Dick et al, 2010)

Figure 11-3 BPALC vs BPnonALC: LocusZoom regional plot of top SNP and nearest gene SGOL1

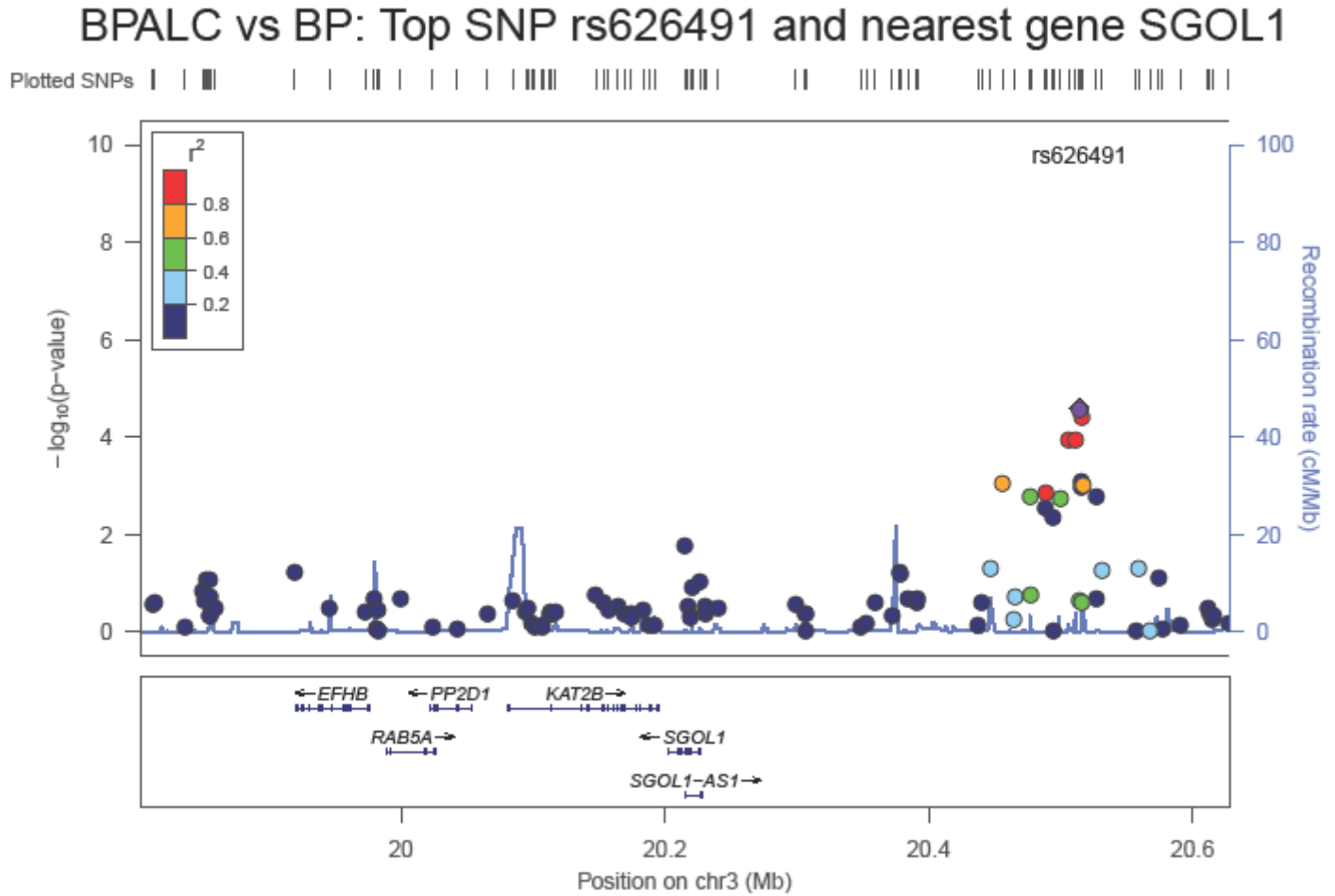


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Table 11.4: Top 10 BPALC vs BP VEGAS2 pathways

GO ID and pathway	Initial pathway length	Final pathway length	Pathway P	Empirical P	Pathway gene list
BIOCARTA P53 HYPOXIA	19	18	1.33E-03	1.14E-03	GADD45A, FHL2, CSNK1A1, CDKN1A, IGFBP3, ABCB1, MAPK8, ATM, MDM2, HIF1A, HSP90AA1, AKT1, NQO1, HIC1, TP53, NFKBIB, BAX, EP300
GO:0031280 negative regulation of adenylate cyclase/lyase activity	49	48	1.57E-03	1.84E-03	CORT, OPRD1, GNAI3, ADCY10, RGS1, ADORA1, ADCY3, GRM7, CCR2, GNAI2, GRM2, DRD3, ADCY5, NPY2R, NPY1R, ADCY2, NPR3, HTR1A, GABBR1, GRM4, OPRM1, ADCY1, GNAI1, GRM3, GRM8, OPRK1, ADCY8, S1PR3, GABBR2, ADRA2A, DRD4, ADCY6, RXFP2, EDNRB, ADCY4, CHRM5, SSTR5, ADCY9, ADCY7, SSTR2, GNAL, GALR1, PALM, HRH3, OPRL1, GNAZ, GALR3, MCHR1
PC Tryptophan catabolism	17	16	3.33E-03	3.40E-03	NADK, NMNAT1, CCBL2, NMNAT2, KMO, HAAO, ACMSD, KYNU, NMNAT3, TDO2, AADAT, IDO2, CCBL1, NADSYN1, QPRT, AFMID
BIOCARTA Tumor Suppressor Arf Inhibits Ribosomal Biogenesis	17	17	3.54E-03	4.00E-03	POLR1A, POLR1B, PIK3CA, PIK3R1, POLR1C, RAC1, TWIST1, PIK3CG, MYC, CDKN2A, ABL1, MDM2, POLR1D, RB1, TP53, TBX2, E2F1
GO:0031365 N-terminal protein amino acid modification	12	12	3.61E-03	3.90E-03	KAT2B, NAA50, MAP6D1, METAP1, NAA15, NMT2, METAP2, NAA16, CREBBP, PDF, NMT1, EP300
REACTOME TRYPTOPHAN CATABOLISM	11	10	3.67E-03	2.50E-03	CCBL2, KMO, HAAO, ACMSD, KYNU, TDO2, AADAT, IDO2, CCBL1, AFMID

Table 11.4 contd...

GO ID and pathway	Initial pathway length	Final pathway length	Pathway P	Empirical P	Pathway gene list
GO:0030817 regulation of cAMP biosynthetic process	94	91	4.12E-03	3.70E-03	CORT, GPR3, OPRD1, CAP1, GNAI3, ADORA3, NTRK1, ADCY10, RGS1, ADORA1, ADCY3, OXER1, FSHR, CALCRL, GRM7, CCR2, GNAI2, GRM2, DRD3, ADCY5, DRD5, EDNRA, NPY2R, NPY1R, ADCY2, NPR3, HTR1A, ADRB2, DRD1, EDN1, CAP2, GABBR1, GRM4, GLP1R, HTR1B, OPRM1, ADCYAP1R1, ADCY1, GNAI1, GRM3, CALCR, GRM8, OPRK1, ADCY8, GNAQ, NTRK2, S1PR3, GABBR2, ABCA1, UCN3, HTR7, ADRA2A, ADRB1, DRD4, PTH, CALCA, PTHLH, ADCY6, RXFP2, EDNRB, ADCY4, TSHR, GPR65, CHRM5, SSTR5, ADCY9, ADCY7, MC1R, ADORA2B, NF1, CRHR1, SSTR2, GALR2, ADCYAP1, GNAL, MC4R, GALR1, PALM, S1PR4, P2RY11, AVP, GHRH, MC3R, GNAS, HRH3, OPRL1, GNAZ, ADORA2A, GALR3, MCHR1, ACR
GO:0015491 cation:cation antiporter activity	14	13	4.27E-03	3.80E-03	SLC9A1, SLC8A1, SLC9A2, SLC11A1, SLC9A9, SLC9A3, SLC22A4, SLC3A2, SLC8A3, SLC24A1, SLC8A2, SLC24A3, SLC9A8
GO:0008484 sulfuric ester hydrolase activity	11	11	4.38E-03	3.80E-03	ARSJ, ARSB, ARSK, ARSI, SULF1, GNS, GALNS, ARSG, SGSH, SULF2, ARSA
GO:0005247 voltage-gated chloride channel activity	12	11	4.43E-03	3.50E-03	CLCN6, CLCNKB, CLIC4, CLCN2, PKD2, CLCN3, CLIC1, CLIC5, CLCN1, CLCN7, CLIC6

Fig. 11.4 BPALC vs BPnonALC: Region plot of SNP nearest GADL1

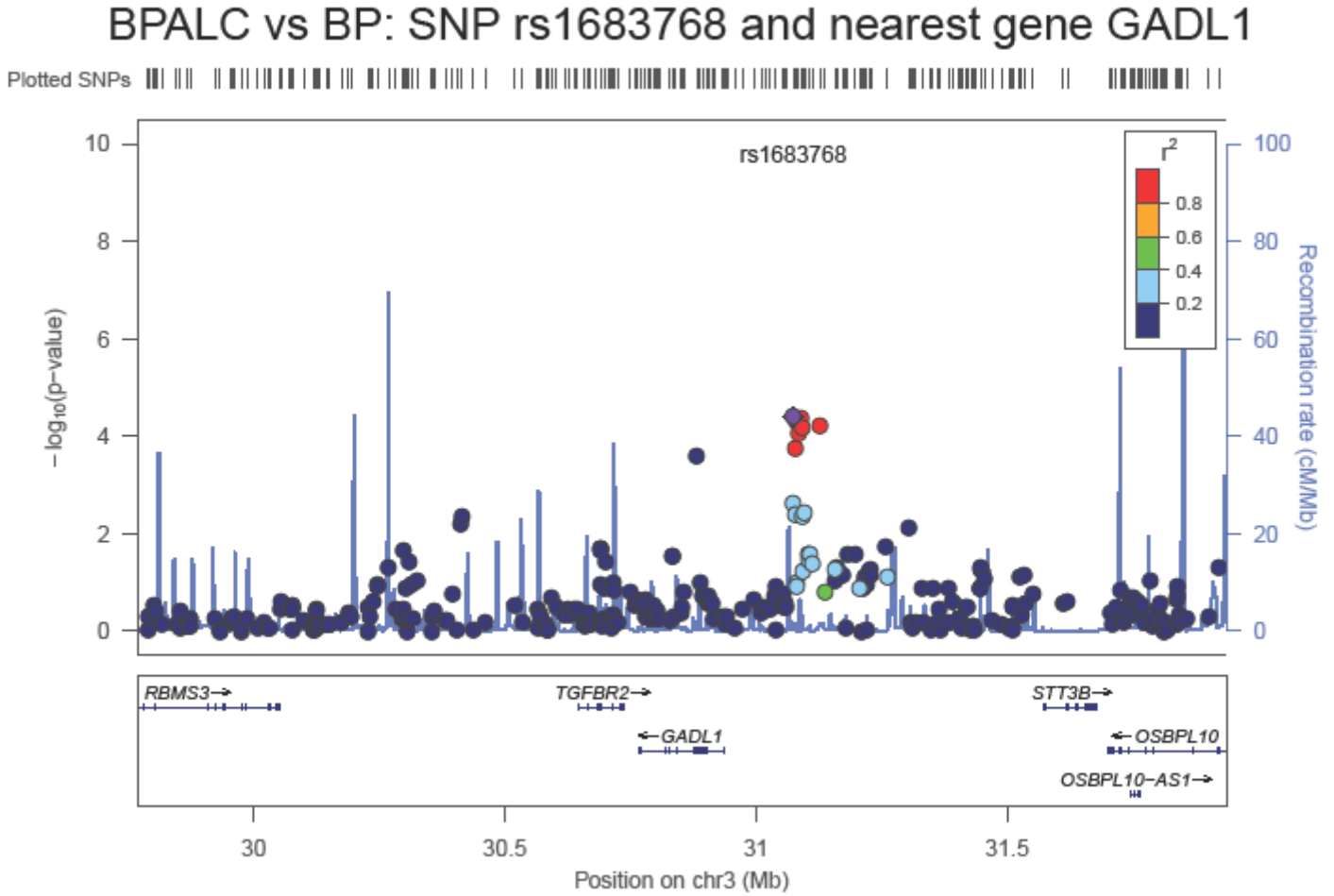


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 11-5 UCL BPALC vs BPnonALC plot of SEMA5A and strongest SNP.

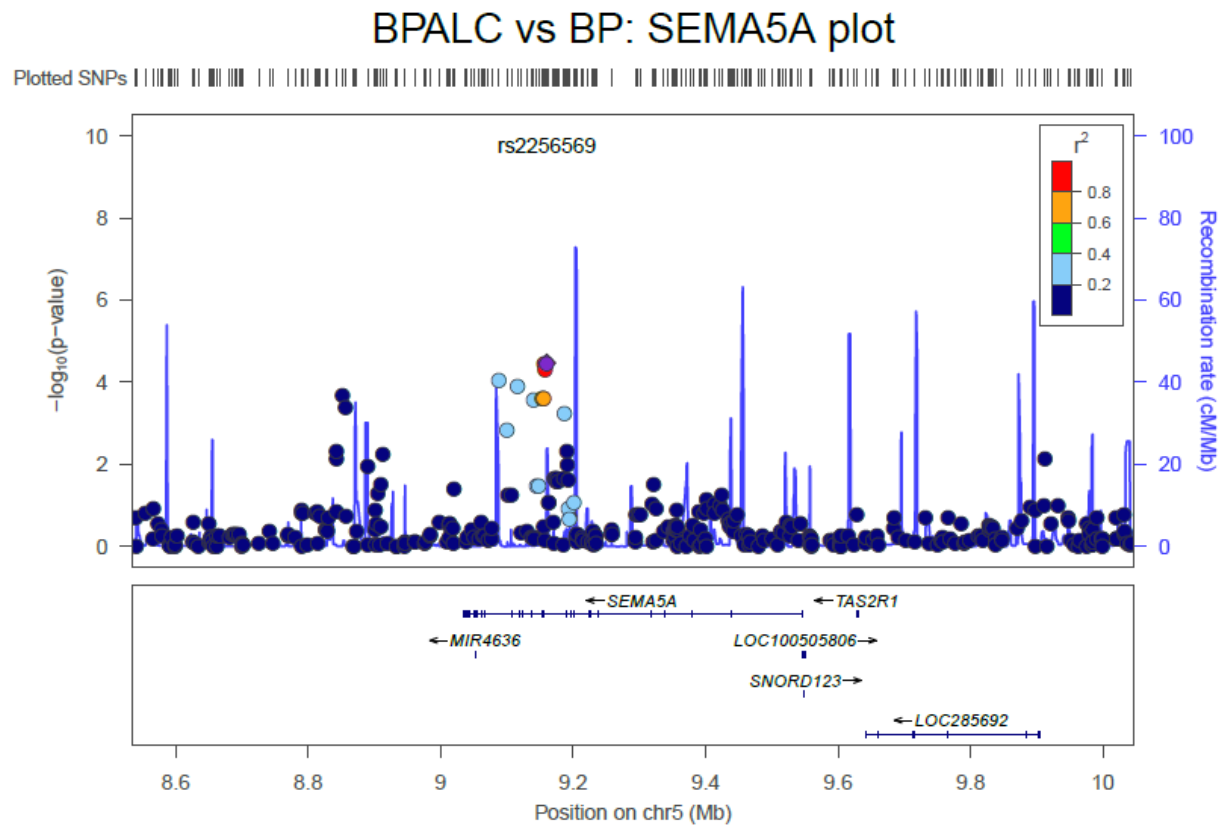


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 11-6 PGC Bipolar GWAS plot of same region as Fig. 11-5 (near SEMA5A).

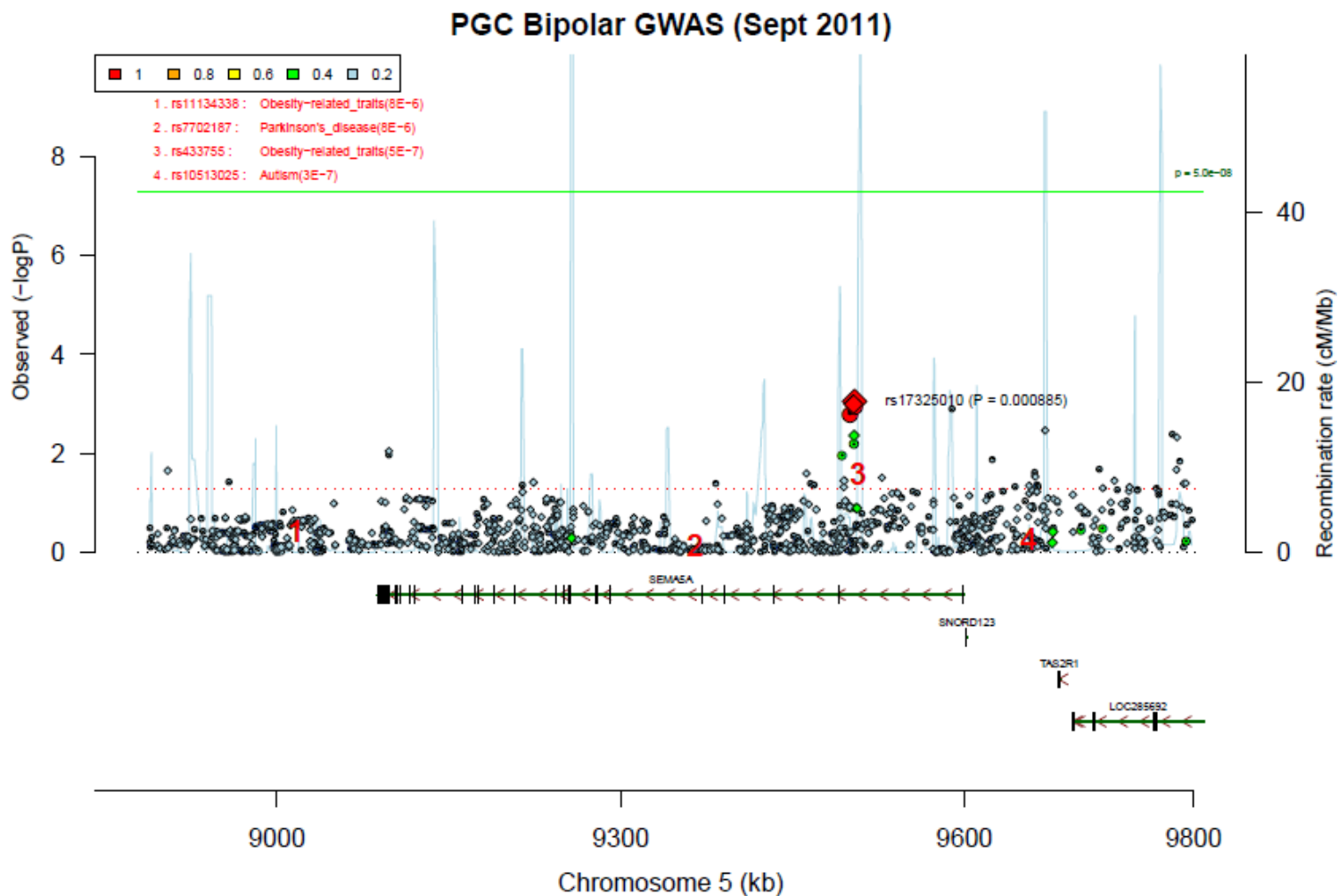


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position
Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 11-7: BPALC vs BPnonALC: ADH5 region plot

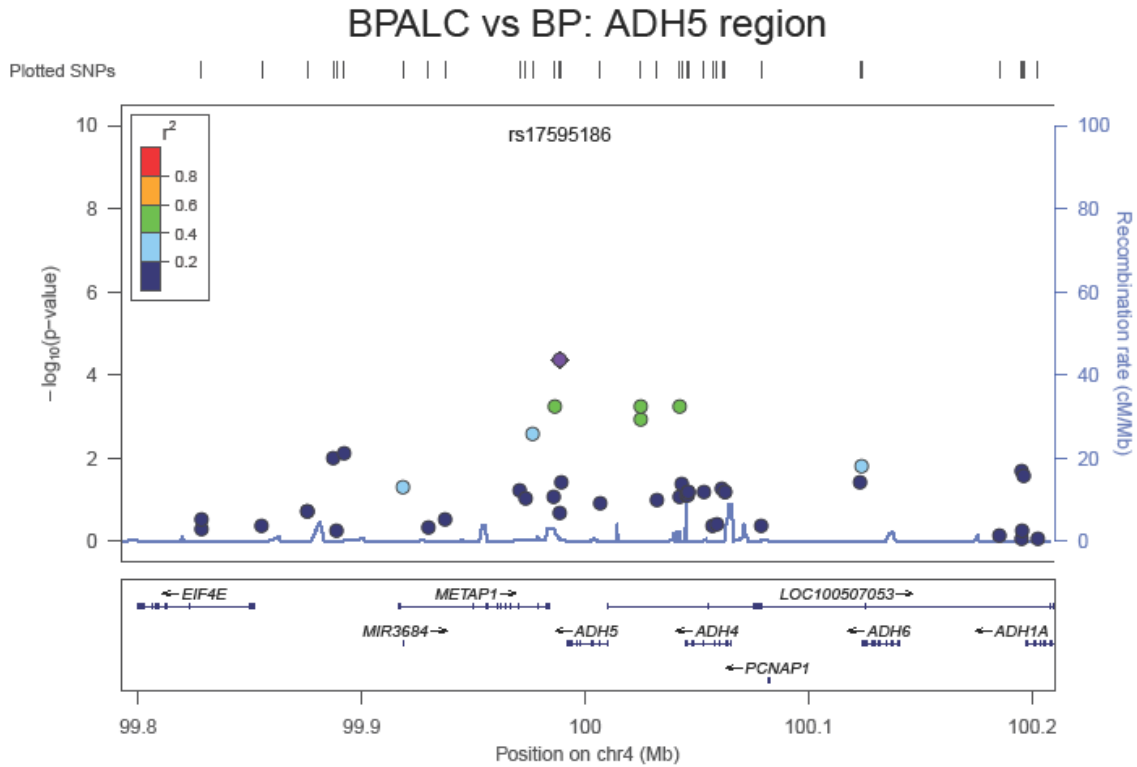


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 11-8 BPALC vs BP: CDH13 region plot

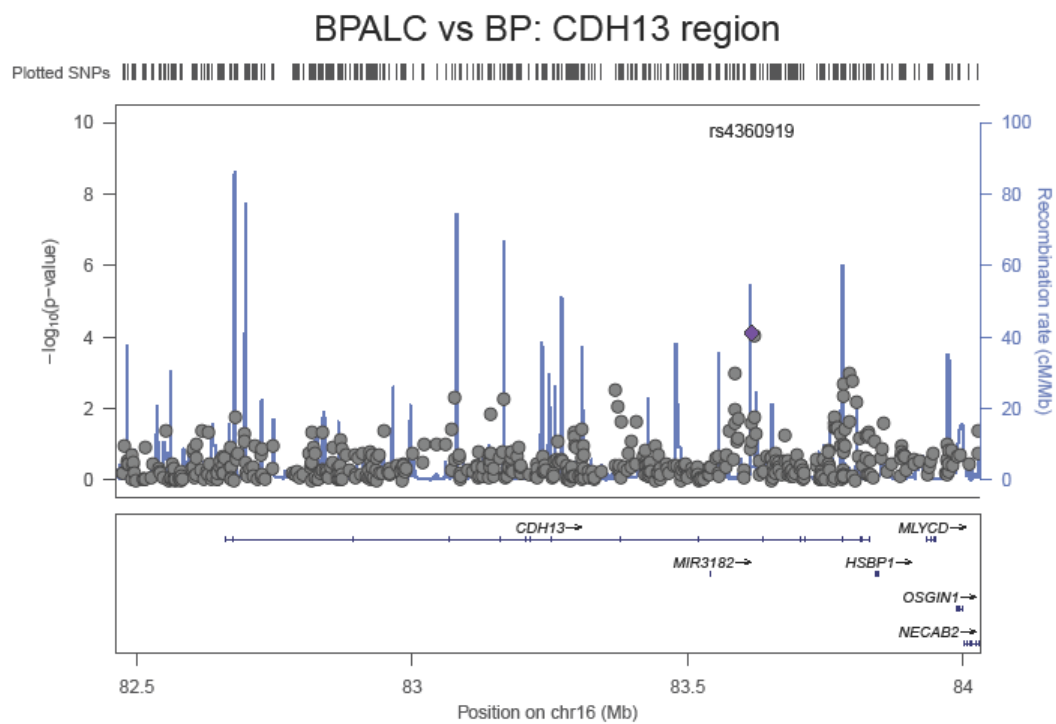


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 11-9 BPALC vs BP gene pathway analysis

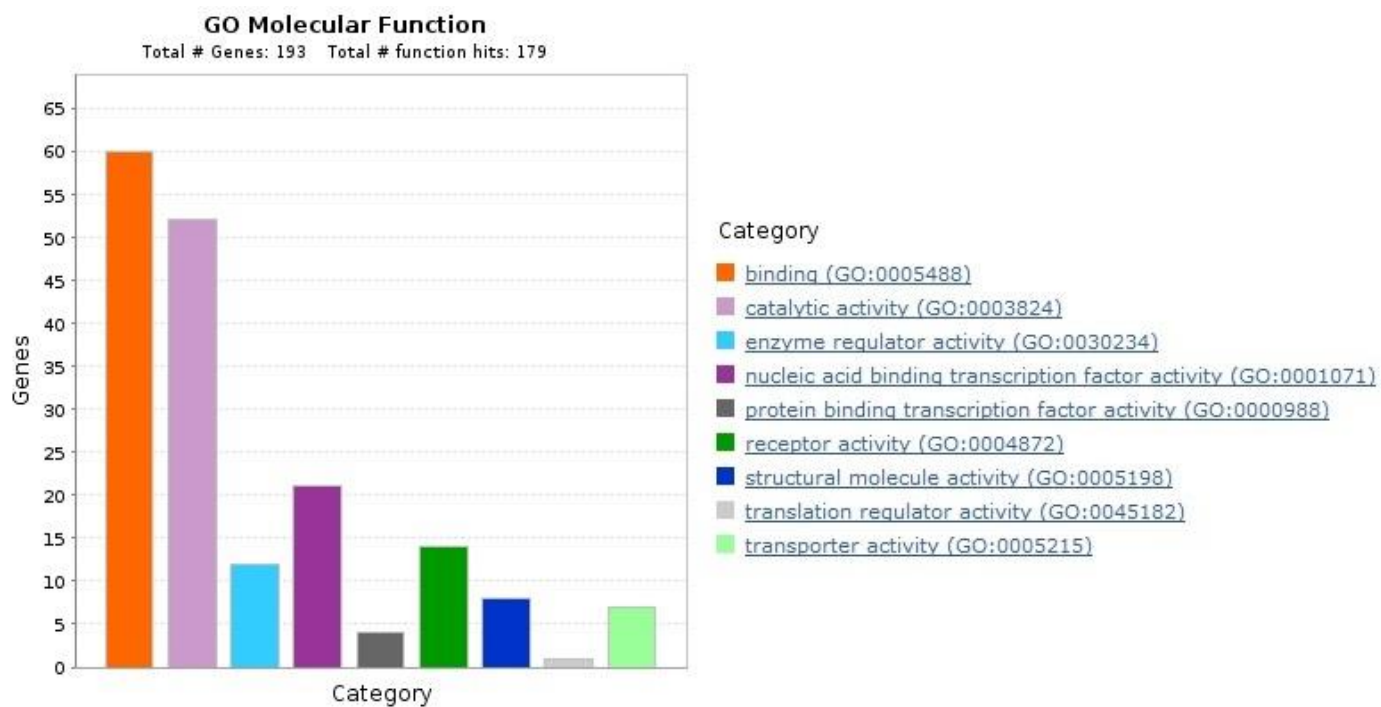


Figure legend: x axis = number of genes; x axis = GO molecular category

CHAPTER 12. RESULTS OF BPALC vs SNC GWAS

This analysis was to try and differentiate BPALC genes from supernormal controls. Unlike the previous chapter which sought AD genes as separate from BP, this approach looks at the combined phenotype – a commonly occurring dual diagnosis picture and thus clinically relevant. For this analysis of BPALC vs normal ancestrally matched controls, we examined 143 subjects with bipolar disorder and comorbid AD (BPALC) vs 510 normal controls.

We analysed 372,193 SNPs genotyped on the Affymetrix GeneChip Human Mapping Array in 506 BP patients and 510 controls following the quality control and data cleaning procedure described in Methods, above. The overall genotyping call rate for analysed SNPs was 99.4%. Including the multimarker predictors, our data set is estimated to capture 78.7% of common variation (SNPs with minor allele frequency $\geq 5\%$) in the CEPH HapMap (CEU) samples with $r^2 > 0.8$.

No SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p-values for the autosomal associations in relation to the expected p-values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 12-1. A Manhattan plot of our association analyses is shown in Fig.12-2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p-values are presented in Table 12.1.

I then checked whether any of the 40 SNPs with the lowest p-value for each analysis were in or close to a gene of known relevant function. None of the top SNPs themselves were previously related directly to AD or alcoholism-related phenotypes, however genes in or near these SNPs are briefly reported.

Our strongest SNP rs17498319 ($p = 1.0 \times 10^{-5}$) was distant from any gene. The second ranked SNP by p value was rs1024820 which is intronic to Musashi RNA-Binding Protein 2 (*MSI2*), a gene which codes for RNA binding protein that regulates the expression of target mRNAs at the translation level. It may play a role in the proliferation and maintenance of stem cells in the central nervous system, and may contribute to disease progression in chronic myeloid leukemia. Third ranked came one of two SNPs intronic to the collagen, type I, alpha 2 (*COL1A2*, highest ranked SNP rs369982, $p = 1.49 \times 10^{-5}$) gene which was also implicated by gene wise analysis. Ranked fourth was SNP rs8062326 ($p = 1.62 \times 10^{-5}$), a single SNP placed midway between the synaptotagmin 17 (*SYT17*) and Inositol 1,4,5-triphosphate receptor-interacting protein-like 2 (*ITPR1L2*) genes.

We found six nominally BPALC-associated SNPs near or within the GABA-A beta-3 receptor (*GABRB3*, highest ranked SNP $p = 2.6 \times 10^{-4}$, uncorrected gene-wise significance $p = 2 \times 10^{-3}$). The tachykinin receptor 1 (*TACR1*, also known as neurokinin-1 receptor, *NK1R*) was nominally associated with BPALC in three SNPs. There were eight SNPs nominally associated with BPALC near the gene for insulin-like growth factor 1 (*IGF1*, highest ranked SNP: rs12426318, 154 kb downstream, $p = 1.8 \times 10^{-3}$). Neither *GABRB3*, *TACR1* nor *IGF1* survived gene wise testing for either phenotype.

In our publication on this data we were interested in commonalities with the “connectivity cluster” genes. We noted forty five SNPs nominally associated with BPALC were located in and around cadherin 11 (*CDH11*, highest ranked SNP rs429065, $p = 1.0 \times 10^{-4}$, gene-wise $p = 6 \times 10^{-4}$). SNP rs429065 is located 22 Kb downstream of *CDH11* (Figure 12-4). Two of the associated SNPs from a genome-wide study of pooled DNA from the COGA sample (Johnson et al. 2006) were also nominally associated in our sample. These were intronic SNP rs35200 ($p = 1.9 \times 10^{-2}$) and rs35164 ($p = 1.4 \times 10^{-4}$) which is 5,484 base pairs downstream from *CDH11*. rs35164 was also implicated with early age of onset alcoholism in the GWAS by Edenberg and colleagues (Edenberg et al. 2010). Excluding the

CDH11 SNPs replicated above, no strongly associated SNPs from previous GWAS replicated in this study.

I then performed VEGAS2 gene permutation tests on all SNPs within and next to a previously implicated gene and corrected for multiple tests. In the gene-based tests, we found 236 genes with $p < 0.01$, strongest gene Hydroxyanthranilate 3,4-Dioxygenase (*HAAO*) on Chromosome 24 $p = 9.13 \times 10^{-4}$ (Fig. 12-5). The top 40 of these are listed in Table 12.2. Cross referencing with the literature database of gene and SNP associations, I found 8 genes that were nominally associated ($p \leq 0.05$) that were in common with our VEGAS2 gene wise list. All gene replications of AD or related phenotypes and the top SNPs in that gene where $p < 0.05$ from our data are listed in Table 12.3. Most notable of these are *HAAO* and *TTL* (Tubulin tyrosine ligase) both of which have been linked to conduct disorder, depression and suicide (Dick et al. 2010), and *TGOLN2* (Trans-Golgi Network Protein 2) which has been implicated in suicide attempts in alcoholism (Mahon, 2010).

Pathway analysis using VEGAS revealed genetic pathways involving adenylate cyclase activity, tryptophan metabolism, solute carriers, voltage gated ion channels, and amino acid modification (Table 12.4).

PANTHER analysis of our top genes implicate in gene based analysis ($P < 0.001$, $n=236$) revealed binding, catalytic, nucleic acid binding and receptor activity among the main molecular pathways (Fig. 12-7).

In summary, this was the first small GWAS comparing patients with BP and comorbid AD with normal controls in an attempt to tease out the BPALC phenotype (as opposed to AD) genetic effects. No marker reached genome-wide significance. However gene wise, replication, and pathway approaches offer a number of biologically plausible genes worthy of further investigation.

12.1 DISCUSSION

This genome-wide study comparing individuals with BP with and without comorbid AD diagnoses has not achieved genome wide significance, but has replicated several previous associations with AD and addiction phenotypes by gene based analysis.

Our most interesting tentative SNP associations with BPALC were in or near genes involved in cell adhesion, neurotransmitter pathways, enzymatic activity, cellular messengers, connective tissue, and cell regulation but none of these reached genome wide significance ($p < 10^{-8}$) and can therefore the associations can only be described as suggestive.

First we discuss SNP signals of association, then gene based findings. Our top ranked genome wide SNP suggestively associated with BPALC is located equidistant from Synaptotagmin 17 (*SYT17*) and Inositol 1,4,5-triphosphate receptor-interacting protein-like 2 (*ITPRIPL2*). *SYT17* is expressed abundantly in the frontal and temporal lobes, hippocampus, hypothalamus, amygdala, substantia nigra, and pituitary. Synaptotagmins are considered to be important in the docking and fusion of synaptic vesicles and plasma membrane in neurotransmitter release and is a worthy candidate gene (Fukuda et al. 1994). *ITPRIPL2* is also expressed in the brain but little is known about its function.

Genes suggestively implicated by gene based significance in the present study include those that are functionally involved in second-messenger systems, ion and neurotransmitter function, neuronal adhesion, differentiation and architecture. These are considered part of the “connectivity constellation” of genes (Uhl et al. 2008b). The cadherin (CDH) gene family are a useful example of connectivity genes. These genes encode a large group of transmembrane proteins that mediate calcium-dependent cell-cell adhesion and the generation of synaptic complexity in the developing brain (Benson & Tanaka 1998). Cadherins have been implicated in mnemonic processes, addictions and bipolar disorder

(Johnson et al. 2009, Uhl et al. 2008b). CDH11 (Fig 12-4) was implicated in a genome-wide study of pooled COGA DNA (Johnson et al. 2006), a recent GWA of alcoholism from COGA families (Edenberg et al. 2010), and was implicated in our study with uncorrected gene based significance tests and high (but not significant) genome-wide ranking. Twenty two CDH11 SNPs in the UCL sample were nominally implicated in BPALC. Our strongest CDH11 SNP, rs429065 was not associated with bipolar disorder in the Wellcome Trust Case Control Consortium ($p=1.2 \times 10^{-2}$) and STEP-BD samples ($p = 2.7 \times 10^{-2}$) (WTCCC, 2007; Sklar et al. 2008).

At gene based level, this study revealed a number of further genes of suggestive interest. Taste receptor 2, 28 (TAS2R38, Fig. 12-3) on 7q36 is of great interest because taste affects alcohol consumption and has known associations with AD. Duffy et al reported a study on 6-n-propylthiouracil (PROP) which is a probe for genetic variation in bitter taste. They found that homozygotes of the alanine-valine-isoleucine (AVI) /AVI haplotype in gene TAS2R38 reported higher alcohol use than heterozygotes (Duffy, 2004). Using COGA data, Wang and colleagues noted a positive correlation between TAS2R38 haplotypes and Maxdrinks in high-risk women of African-American origin. The common taster haplotype was significantly associated with a lower mean Maxdrinks compared with the other haplotypes, but not alcohol dependence (Wang, 2007).

TAS2R38 was found at nominal gene based level only in the BPALC vs SNC comparison, and not at meta-analysis level. This is explainable by the possibility that taste receptors play a more prominent role in alcohol dependence in the context of bipolar disorder, then in AD alone, at least in this population. The AD association has been reported in African Americans (in pure AD not BPALC or similar). Perhaps this is evidence of pleiotropy in that the gene has an effect on expression of different phenotypes in different populations of people.

The tentative associations in our study should be replicated in other BPALC samples including non-Caucasians to fully evaluate the gene. There are numerous taste receptor genes, and taste is a key factor in the tolerability of a drink, so it would be worth considering all as candidate genes for AD. Sequencing of individuals with risk alleles and possible cross-matching with laboratory based taste testing, or surveys of subjective taste (e.g. of popular alcoholic drinks) may allow for further clarification.

HAAO (Fig 12-5) was nominally associated in our BPALC vs SNC sample (and, also in the BPALC vs BPnonALC, and meta-analysis studies). The phenotypes of AD, conduct disorder, and suicide attempts, which show linkage on chromosome 2, are all characterized by elements of impulsivity and poor behavioural control. Conduct disorder is a robust predictor of both concurrent and future alcohol problems (Crowley et al. 1998; Moss and Lynch, 2001; White et al. 2001) Furthermore, numerous twin studies indicate that the overlap between childhood conduct disorder and adult alcohol dependence is largely due to shared genetic factors (Slutske et al. 1998; Krueger, 1999; Young et al. 2000; Kendler et al. 2003). This common genetic liability is believed to be a predisposition toward behavioural undercontrol/disinhibition, which can manifest as conduct disorder in childhood and alcohol dependence later in life (Slutske et al. 2002).

It is not unreasonable to theorize that the characteristics which predispose to substance use and externalizing problems represent an expression of an underlying tendency toward behavioural disinhibition. Longitudinal studies have found that behavioural disinhibition measured in childhood/adolescence predicts both the development of substance use disorders and a propensity toward suicidal behaviour in young adulthood (Tarter et al. 2004). Behavioural disinhibition is a key diagnostic criteria for BP and is arguably also evident in SZ. It is possible that HAAO in our sample is involved in this impulsive propensity across the disorders.

Of further note is that Tachykinin receptor 1 (TACR1) also known as Neurokinin receptor 1 (NKR1) had a gene based p value of 0.016. TACR1 has been previously implicated in BP, AD, and Attention Deficit Hyperactivity Disorder (ADHD) (Sharp, 2014). It should be noted that the Sharp study was from unlinked UCL AD and ADHD samples, but the BP sample was the same as used in this GWAS (ie PBALC vs SNC). The association with TACR1 and BPAD, ADS, and ADHD suggests a shared molecular pathophysiology between these affective disorders.

The GABA system has a wealth of overlapping evidence for involvement in BP, SZ and AD. Important negative findings include the genes of the Gamma-aminobutyric acid (GABA) system which encode receptors for the major inhibitory transmitter of the nervous system. GABA receptors are of strong interest in addictions, alcoholism, and affective disorders (Agrawal et al. 2008, Bierut et al. 2010). None of the GABA receptor genes including GABRA2 were implicated at corrected gene wise level in the present study.

In our sample, GABRB3 (Fig 12-6) showed tentative gene based associations in several groups. Noble found a significant decrease in the prevalence ($P = 4.5 \times 10^{-3}$) and frequency ($P = 2.7 \times 10^{-2}$) of the GABRB3 major (G1) allele in “severe alcoholics”, compared to non-alcoholics. Furthermore, a significant progressive decrease was noted in G1 allelic prevalence ($P = 2.4 \times 10^{-3}$) and frequency ($P = 1.9 \times 10^{-2}$) in non-alcoholics, less severe and severe alcoholics, respectively (Noble, 1998). Sequencing GABRB3 in subjects with the G1 allele may reveal novel mutations.

The SEMA5A gene belongs to the semaphorin gene family that encodes membrane proteins containing a semaphorin domain and several thrombospondin type-1 repeats. Members of this family are involved

in axonal guidance during neural development. This gene has been implicated as an autism susceptibility gene and cri-du-chat syndrome. SEMA5A is of interest because it has been replicated in two studies on autistic spectrum disorder, another neuropsychiatric condition with a high heritability. Weiss and colleagues (Weiss et al, 2009) published a linkage and association mapping study using half a million SNPs for 1553 individuals with autism and their families. Initial analysis did not yield genome-wide significant associations; however, genotyping of top hits in additional families revealed a SNP on chromosome 5p15 (between SEMA5A and TAS2R1) that was significantly associated with autism ($P = 2 \times 10^{-7}$). They also demonstrated that expression of SEMA5A is reduced in brains from autistic patients, further implicating SEMA5A as an autism susceptibility gene. Cheng and colleagues' GWAS of autism spectrum disorders showed significant association with an intergenic SNP near Semaphorin 5A (SEMA5A) and provided evidence for reduced expression of the same gene. In a novel GWAS follow-up approach, they mapped an expression regulatory pathway SEMA5A in silico by using population expression and genotype data sets. They found that the SEMA5A regulatory network significantly overlapped with rare autism-specific CNVs (Cheng et al. 2013).

How might autism and alcoholism be linked? They have no phenotypic commonalities to suggest underlying aetiological or biological connections. Clinically, one would expect the social interaction difficulties and reduced novelty-seeking of autism would protect against alcohol use and alcohol-using peers. Small studies of individuals with autism have reported very low rates of drinking and substance use in people with autism spectrum disorders (Hofvander et al. 2009; Santosh & Mijovic 2006).

However De Alwis and colleagues found an increased risk of developing alcohol and other substance misuse with increasing numbers of autism spectrum traits. The researchers studied autistic traits in 3080 Australian twins and found just under 20 percent of those with no autistic traits met the criteria for

alcoholism. In contrast thirty five percent of people with six or more traits were alcohol dependent (De Alwis et al. 2014).

There is considerable genetic overlap between ADHD and autism; and alcoholism and ADHD. There is also evidence of common vulnerability genes to both autism and alcoholism. Certainly families with autistic children have higher rates of alcoholism. Miles et al analysed 167 pedigrees ascertained through an autistic child; 39% had alcoholism in patterns consistent with transmission of a genetic trait (Miles et al. 2003). Schumann et al's GWAS meta-analysis of more than 40,000 people revealed an association between SNP rs6943555 in the Autism Susceptibility Candidate 2 gene (AUTS2) and alcohol consumption at genome-wide significance ($P = 4 \times 10^{-9}$) (Schumann et al. 2011). In our data neither this SNP nor its related gene on gene based analysis were significant ($p=0.6$) (Fig. 15-4). However there are likely, as with all complex diseases, to be many common genetic pathways crossing disorders, further complicated by the possibility of pleiotropy (Consortium 2015).

Genes of interest with suggestive involvement that did not stand up to correction included ADH5, CHRNA6, and KCNMA1. ADH5 (uncorrected gene based $p = 0.03$) was implicated in our BPALC vs BPnonALC data is a potential candidate gene for AD because of its role in alcohol metabolism. It has previously been associated with AD (Park, 2013) and AD symptom count in African Americans (Kendler et al. 2011). This lack of association may reflect ancestral differences because our sample was of White British ethnicity. CHRNA6 on chromosome 22 (uncorrected gene based $p = 0.029$; strongest SNP rs10427722, $p = 0.01752$) is also intriguing from a substance misuse perspective because it has been associated with subjective response to tobacco, but not alcohol (Zeiger et al. 2008). Lastly, Potassium Channel, Calcium Activated Large Conductance Subfamily M Alpha, Member 1 (KCNMA1), on Chromosome 10, uncorrected gene based $p = 0.04$; top SNP rs2673464, $p = 0.006$) was

associated with AD in people with SZ (Kendler et al. 2011). KCNMA1 is the human homologue of the slo-1 gene in *C. Elegans*, a gene involved in ethanol resistance in this species (Wang et al. 2001).

12.2 BPALC vs SNC GWAS CONCLUSIONS

The relationship between bipolar and unipolar affective disorder and alcohol dependence is bidirectional (Sonne and Brady, 2002) and there is consensus that the two disorders have reciprocal effects on each other. AD may also change expression and function of the same neurotransmitters as those involved in bipolar disorder, thereby “prompting” the symptoms of unipolar and bipolar affective disorder (Tohen et al. 1998).

The present study has benefitted from a well-characterized, ancestrally homogenous and ancestrally-matched control sample. The main drawback was the modest sample size of bipolar patients with AD for a genome-wide study.

Although low significance values are needed for convincing GWAS results there are genes previously implicated in the genetics of alcoholism which were suggestively associated with bipolar alcoholism after gene based association tests. This suggests that our results are worthy of further investigation. The “connectivity cluster” genes are strongly represented in this study. We also report tentative new alcoholism gene associations in a bipolar-alcoholic sample which are in need of further investigation in additional samples.

In the next chapter I report on the SZALC vs SZnonALC analysis.

Fig. 12-1 Q-Q Plot of BPALC vs SNC

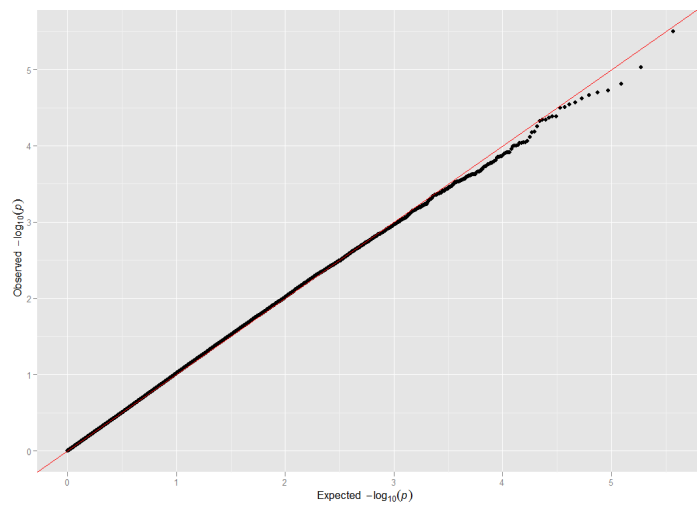


Figure legend: $Q-Q$ plots of observed and expected $-\log_{10}(P)$ of the associations between SNPs

Fig. 12-2 Manhattan plot of BPALC vs SNC

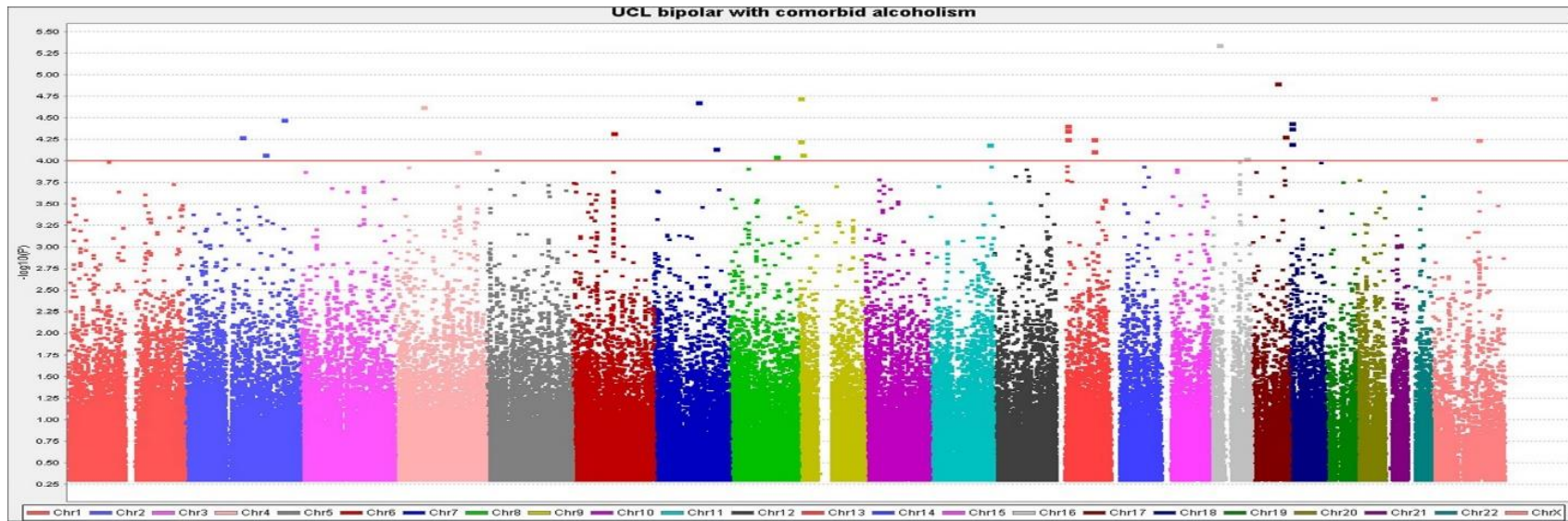


Figure legend: Results of the genome-wide association analyses for BPALC vs SNC. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Table 12.1 BPALC vs SNC: Top 40 SNPs and uncorrected gene based results

CHR	SNP	Position (base pairs)	A1	AF Cases	AF Control	A2	CHISQ	P	OR	Position	Distance	Nearest gene code	Gene based p*
4	rs17498319	58147714	C	0.25	0.1394	T	19.5	1.01E-05	2.058	downstream	71451	---	---
17	rs1024820	55607591	T	0.4317	0.2926	C	19.11	1.24E-05	1.836	intron	0	MSI2	3.28E-01
7	rs369982	94048482	A	0.5357	0.3909	G	18.75	1.49E-05	1.798	intron	0	COL1A2	1.74E-04
16	rs8062326	19156016	A	0.075	0.02222	G	18.59	1.62E-05	3.568	intron	0	SYT17/ ITPRIPL2	---
2	rs10205049	119899933	G	0.5321	0.3927	A	17.37	3.07E-05	1.759	downstream	13886	C1QL2	5.93E-03
17	rs17836920	72078926	A	0.275	0.4119	G	17.28	3.23E-05	0.5416	intron	0	C17orf54	---
9	rs10967953	2756638	C	0.2607	0.1535	T	17.2	3.37E-05	1.945	intron	0	KIAA0020	1.79E-02
17	rs12051795	45413384	A	0.0219	0.001027	G	16.7	4.38E-05	21.78	intron	0	EFCAB13	1.29E-01
13	rs4770393	23692789	T	0.06429	0.01829	C	16.66	4.46E-05	3.687	upstream	140653	---	---
13	rs1330910	78582044	T	0.1143	0.04757	C	16.62	4.56E-05	2.583	intron	0	RNF219-AS1	8.59E-02
5	rs12516376	129720170	C	0.1107	0.2202	T	16.6	4.62E-05	0.4409	downstream	2144	---	---
7	rs10488600	136605458	T	0.2122	0.1173	C	16.34	5.29E-05	2.028	intron	0	CHRM2	5.40E-02
13	rs7318572	23678935	A	0.04745	0.01029	T	16.29	5.44E-05	4.792	upstream	126799	---	---
13	rs17078254	23678405	T	0.04643	0.0101	G	16.2	5.70E-05	4.772	upstream	126269	---	---
18	rs17288584	69134662	A	0.1036	0.04141	G	16.19	5.72E-05	2.674	downstream	52536	LOC100505776	3.33E-01
14	rs7158201	71010546	T	0.1475	0.07041	C	16.12	5.94E-05	2.284	intron	0	RP11-486O13.4	---
5	rs2189381	129792625	C	0.1964	0.3198	T	16.05	6.16E-05	0.5198	upstream	72487	---	---
13	rs7322449	23693091	T	0.06115	0.01721	C	16.03	6.23E-05	3.72	upstream	140955	---	---
13	rs17068683	78568871	A	0.1143	0.04848	G	16.01	6.30E-05	2.532	intron	0	RNF219-AS1	8.59E-02
9	rs16925783	7288704	A	0.1214	0.05354	G	15.77	7.17E-05	2.443	downstream	113056	KDM4C	6.70E-01
11	rs7104304	125283262	A	0.3489	0.4827	G	15.63	7.70E-05	0.5744	intron	0	PKNOX2	1.35E-01
9	rs7857841	2756756	A	0.1464	0.07085	C	15.6	7.82E-05	2.25	intron	0	KIAA0020	1.79E-02
6	rs1028714	1833041	G	0.3429	0.2273	C	15.44	8.52E-05	1.774	intron	0	GMDS	8.08E-01
4	rs17661170	171175320	G	0.1964	0.3172	A	15.44	8.53E-05	0.5263	upstream	19750	LOC101928223	3.28E-01

Table 12.1 contd...

17	rs7222497	72087797	G	0.3536	0.4859	C	15.39	8.73E-05	0.5788	intron	0	RP11-101O21.1	---
3	rs918244	10480628	C	0.06786	0.02134	G	15.39	8.74E-05	3.338	intron	0	ATP2B2	2.19E-01
2	rs3815475	160114315	A	0.03571	0.006085	G	15.34	8.97E-05	6.049	intron	0	WDSUB1	5.11E-03
7	rs6960667	130923445	G	0.1	0.0404	A	15.3	9.18E-05	2.639	intron	0	MKLN1	4.06E-01
18	rs280983	4451489	C	0.275	0.403	T	15.27	9.33E-05	0.5618	intron	0	DLGAP1	3.04E-01
5	rs6879343	23056410	T	0.06835	0.1602	C	15.21	9.63E-05	0.3844	downstream	41774	---	---
17	rs8080915	72083792	C	0.3714	0.503	T	15.16	9.87E-05	0.5838	intron	0	RP11-101O21.1	---
14	rs10137987	28371914	A	0.5821	0.4505	G	15.15	9.91E-05	1.699	upstream	87297	---	---
14	rs11158863	70981166	C	0.1393	0.0668	A	15.09	1.02E-04	2.261	intron	0	---	---
17	rs10512591	70879560	G	0.02143	0.09091	A	15.08	1.03E-04	0.219	intron	0	SLC39A11	6.31E-02
20	rs6061245	61044227	T	0.6	0.4687	C	15.05	1.04E-04	1.7	intron	0	GATA5	5.46E-04
18	rs280984	4451130	T	0.2714	0.398	C	14.99	1.08E-04	0.5636	intron	0	DLGAP1	3.04E-01
22	rs2267123	28284061	T	0.2679	0.1657	A	14.92	1.12E-04	1.843	intron	0	PITPNB	3.83E-03
12	rs17835106	76405801	C	0.25	0.1515	T	14.78	1.21E-04	1.867	upstream	28006	---	---
5	rs2256569	9160380	C	0.1893	0.3057	T	14.66	1.29E-04	0.5304	intron	0	SEMA5A	2.30E-02
20	rs6587239	61040951	T	0.3857	0.5152	C	14.64	1.30E-04	0.5909	synon	0	GATA5	5.46E-04

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F_A= Frequency of Allele 1; F_U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. *Uncorrected p <0.05 in **bold**.

Table 12.2 BPALC vs SNC: Top 40 VEGAS2 Gene Based Permutation Testing

Chr	Gene	n SNPs	n VEGAS2 Sims	VEGAS2 Test result	Gene based P value	Top SNP	Top SNP P
7	COL1A2	16	1.00E+06	77.79	1.74E-04	rs369982	1.49E-05
14	ADAM20	3	1.00E+06	39.02	2.07E-04	rs7158201	5.94E-05
1	SLC9A1	8	1.00E+06	65.13	3.96E-04	rs11247612	3.08E-04
1	CEP170	5	1.00E+06	46.91	4.84E-04	rs12091558	1.57E-04
14	ADAM20P1	2	1.00E+06	22.90	4.91E-04	rs11158863	1.02E-04
20	GATA5	9	1.00E+06	46.89	5.46E-04	rs6061245	1.04E-04
17	EFNB3	7	1.00E+06	38.26	5.48E-04	rs2909430	2.11E-04
2	OXER1	23	1.00E+06	143.77	8.15E-04	rs7583085	5.75E-04
2	HAAO	24	1.00E+06	144.03	9.13E-04	rs7583085	5.75E-04
1	FAM131C	5	1.00E+06	25.25	9.35E-04	rs6696416	5.54E-03
2	RETSAT	8	1.00E+05	45.79	9.80E-04	rs1053561	9.94E-04
3	SNORA58	2	1.00E+05	14.07	1.00E-03	rs9843898	4.14E-04
7	LOC100131257	17	1.00E+05	71.43	1.09E-03	rs7793374	4.33E-04
16	CDH11	64	1.00E+05	264.32	1.11E-03	rs429065	1.64E-04
17	WRAP53	9	1.00E+05	38.89	1.11E-03	rs2909430	2.11E-04
2	POLR1B	13	1.00E+05	67.72	1.11E-03	rs4849074	5.07E-04
17	TP53	10	1.00E+05	41.34	1.13E-03	rs2909430	2.11E-04
10	ZNF33A	16	1.00E+05	89.63	1.14E-03	rs645064	1.50E-04
18	GREB1L	15	1.00E+05	91.36	1.18E-03	rs1318841	2.43E-03
17	DNAH2	10	1.00E+05	45.62	1.21E-03	rs2909430	2.11E-04
12	MSI1	7	1.00E+05	35.72	1.22E-03	rs2701632	1.11E-03
20	RBP8NL	8	1.00E+05	39.44	1.23E-03	rs6061245	1.04E-04
14	MED6	5	1.00E+05	32.66	1.24E-03	rs7158201	5.94E-05
2	CHCHD5	12	1.00E+05	63.77	1.30E-03	rs4849074	5.07E-04
20	LOC284798	10	1.00E+05	54.07	1.37E-03	rs4815389	3.11E-04
19	BRD4	5	1.00E+05	27.58	1.45E-03	rs11665899	4.76E-03

Table 12.2contd...

2	TTL	10	1.00E+05	59.73	1.47E-03	rs4849074	5.07E-04
12	AQP6	10	1.00E+05	42.54	1.61E-03	rs7309966	4.22E-03
10	ZNF37A	21	1.00E+05	130.71	1.64E-03	rs645064	1.50E-04
6	MIR1275	11	1.00E+05	43.09	1.64E-03	rs4713731	2.71E-04
9	KCNV2	21	1.00E+05	70.17	1.67E-03	rs10967953	3.37E-05
17	SPNS3	14	1.00E+05	45.96	1.70E-03	rs4239046	1.44E-03
1	CLCNKA	7	1.00E+05	31.69	1.85E-03	rs6696416	5.54E-03
18	ABHD3	4	1.00E+05	25.83	1.87E-03	rs2850565	2.77E-03
2	MTA3	37	1.00E+05	192.29	1.92E-03	rs7583085	5.75E-04
4	HGFAC	9	1.00E+05	39.59	1.98E-03	rs13128489	1.64E-03
22	MN1	23	1.00E+05	76.38	2.15E-03	rs5752613	1.90E-03
6	DAAM2	34	1.00E+05	124.46	2.16E-03	rs2504090	1.62E-04
2	TGOLN2	14	1.00E+05	59.33	2.18E-03	rs1053561	9.94E-04
22	MIR3199-2	6	1.00E+05	28.57	2.26E-03	rs2267123	1.12E-04
18	MIR320C1	3	1.00E+05	22.88	2.31E-03	rs2850565	2.77E-03

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations

Table 12.3 BPALC vs normal controls: Genes Yielding Evidence for Association at $P < 0.05$ After Permutation Testing and Replication

Chr	Gene symbol	Gene name	n SNPs	n VEGAS Sims	VEGAS Test result	P	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
2	HAAO	3-Hydroxyanthranilate 3,4-dioxygenase	24	1000000	144.0322	9.13E-04	rs7583085	5.75E-04	AD with conduct disorder, depression and suicide (P=0.0327) (Dick, 2010)
2	TTL	Tubulin tyrosine ligase	10	100000	59.73078	1.47E-03	rs4849074	5.07E-04	AD with conduct disorder, depression and suicide (P=0.0302) (Dick, 2010)
2	TGOLN2	Trans-Golgi Network Protein 2	14	100000	59.32617	2.18E-03	rs1053561	9.94E-04	Suicide attempts in alcoholism (p = 0.03-0.008) (Mahon, 2010)
7	TAS2R38	Taste receptor 2 38	11	100000	50.00965	2.31E-03	rs2204607	3.17E-03	Bitterness perception and alcohol intake (Duffy, 2004)
1	MTHFR	methylenetetrahydrofolate reductase	20	100000	85.99447	3.42E-03	rs2076001	2.65E-04	Serum homocysteine levels in alcoholics (de La Vega 2001)
6	HEY2 (HERP1)	Hes-Related Family BHLH Transcription Factor With YRPW Motif 2	13	100000	58.73213	6.55E-03	rs9321054	6.54E-03	Reduced HERP mRNA expression in AD (Bleich, 2006)
15	GABRB3	GABA(A) receptor beta3 subunit	37	100000	85.24025	7.49E-03	rs2114485	3.54E-04	Severity of alcoholism (Noble, 1998)
11	GSTP1	Glutathione S-Transferase Pi 1	3	10000	13.34865	9.50E-03	rs6591256	1.90E-03	Val/val polymorphism and cirrhosis and pancreatitis in AD (Burim 2004)

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations

Table 12.4 VEGAS2 pathway analysis results

GO ID and pathway	Initial pathway length	Final pathway length	Pathway P	Empirical P	Pathway gene list
GO:0001756somitogenesis	33	32	8.33E-05	5.00E-05	PSEN2, WNT3A, MSGN1, ZEB2, RBPJ, LEF1, SFRP2, RIPPLY2, TBX18, T, DLL1, MEOX2, NKX3-1, SFRP1, PRKDC, ROR2, ATM, MYF5, PCDH8, PSEN1, MESP2, HES7, TCAP, KAT2A, MEOX1, AXIN2, MIB1, DLL3, TCF15, PAX1, POFUT1, EP300
BIOCARTA ARF PATHWAY	17	17	1.79E-04	1.02E-04	POLR1A, POLR1B, PIK3CA, PIK3R1, POLR1C, RAC1, TWIST1, PIK3CG, MYC, CDKN2A, ABL1, MDM2, POLR1D, RB1, TP53, TBX2, E2F1
REACTOMETRYPTOPHANCATABOLISM	11	10	2.74E-04	1.48E-04	CCBL2, KMO, HAAO, ACMSD, KYNU, TDO2, AADAT, IDO2, CCBL1, AFMID
GO:0035282segmentation	46	45	2.21E-04	1.70E-04	PSEN2, WNT3A, MSGN1, TCF7L1, ZEB2, HOXD8, TDGF1, RBPJ, LEF1, SFRP2, RIPPLY2, TBX18, T, DLL1, MEOX2, HOXA2, NKX3-1, SFRP1, PRKDC, MLLT3, ROR2, BMI1, EGR2, ATM, FRS2, MYF5, TBX3, PCDH8, PSEN1, MESP2, RPGRIP1L, DVL2, HES7, TCAP, KAT2A, MEOX1, HOXB6, AXIN2, MIB1, DLL3, TCF15, PAX1, POFUT1, MAFB, EP300
GO:0005247voltage-gated chloride channel activity	12	11	3.92E-04	2.82E-04	CLCN6, CLCNKB, CLIC4, CLCN2, PKD2, CLCN3, CLIC1, CLIC5, CLCN1, CLCN7, CLIC6
PANTHER BIOLOGICAL PROCESS Chromatin packaging and remodeling	162	113	7.51E-04	7.80E-04	KDM1A, RBBP4, NASP, BRDT, SYCP1, HIST2H2AB, SETDB1, GATAD2B, ASH1L, H3F3A, HIST3H2A, CENPA, MTA3, TSPYL6, HAT1, KAT2B, SS18L2, SETD2, PBRM1, L3MBTL3, ARID1B, HDAC9, CBX3, H2AFV, BAZ1B, TRRAP, SUV39H2, TRDMT1, SIRT1, H2AFY2, SMC3, SIRT3, NAP1L4, SPTY2D1, TAF6L, UBTF1, H2AFX, NCAPD3, NCAPD2, PHC1, H2AFJ, H3F3C, SMARCD1, CBX5, BAZ2A, NAP1L1, SIRT4, SETD1B, SETDB2, TOX4, BAZ1A, MTA1, INO80, TNP2, SETD1A, TOX3, CBFA2T3, SUPT6H, SUZ12, TADA2A, SMARCE1, EZH1, UBTF, CBX1, HILS1, SUPT4H1, SMARCD2, BPTF, H3F3B, CBX4, L3MBTL4, SS18, SETBP1, CHAF1A, SAFB, BRD4, GATAD2A, SUPT5H, TRIM28, CBFA2T2, SS18L1, CHAF1B, CECR2, SMARCB1, H1FO, HDAC10

Table 12.4 contd...

GO:0007389 pattern specification process	240	196	6.54E- 04	5.60E- 04	HES3, CTNNBIP1, PAX7, GRHL3, CYR61, CELSR2, ALX3, HIPK1, NOTCH2, PBX1, RNF2, BTG2, DISP1, LEFTY2, PSEN2, WNT3A, MSGN1, IFT172, SIX3, VAX2, CYP26B1, NOTO, TCF7L1, EN1, GLI2, ZEB2, ACVR2A, ACVR1, DLX2, HOXD3, NEUROD1, SF3B1, SATB2, ACVR2B, CTNNB1, TDGF1, ARL6, PLXND1, ZIC1, CHR1, TP63, HES1, HTT, NKX3-2, RBPJ, GSX2, SHROOM3, NKX6-1, PKD2, BMPR1B, CXXC4, LEF1, PITX2, BBS7, HHIP, SMAD1, EDNRA, SFRP2, SEMA5A, FGF10, FST, APC, SMAD5, CDX1, PROP1, EDN1, RING1, MDFI, VEGFA, BMP5, RAB23, RIPPLY2, TBX18, CITED2, T, DLL1, UNCX, LFNG, MEOX2, SP8, EVX1, TBX20, GLI3, RELN, SMO, HIPK2, SHH, MNX1, GATA4, NKX3-1, SFRP1, PRKDC, ASPH, EYA1, ARC, FOXH1, RFX3, CER1, MLLT3, PCSK5, GAS1, ROR2, BARX1, PTCH1, TGFBR1, PBX3, LMX1B, ENG, NOTCH1, BMI1, ZEB1, NRP1, EGR2, NODAL, BMPR1A, CYP26A1, FRAT1, LDB1, EMX2, ALX4, LRP5, C2CD3, ATM, ZBTB16, DSCAML1, TULP3, LRP6, WNT1, HOXC4, GDF11, GLI1, FRS2, MYF5, ALX1, ASCL1, TBX3, IFT88, FLT1, ALG5, PCDH8, FOXG1, NKX2-1, BMP4, SIX1, PSEN1, GSC, YY1, GREM1, ALDH1A2, SMAD6, MESP2, NR2F2, AXIN1, FTO, FOXC2, DVL2, HES7, LHX1, HNF1B, PCGF2, TCAP, KAT2A, MEOX1, WNT3, HOXB9, NOG, AXIN2, FOXJ1, RFNG, MIB1, SMAD2, SMAD4, RAX, RTTN, TSHZ1, DAND5, DLL3, NTF4, TCF15, MKKS, OVOL2, PAX1, FOXA2, KIF3B, MAFB, IFT52, BMP7, SIM2, CRKL, MFNG, EP300
REACTOME DAG AND IP3 SIGNALING	29	29	1.06E- 03	9.20E- 04	PRKACB, ADCY3, PRKCE, CALM2, PDE1A, CREB1, PRKAR2A, PRKCD, ADCY5, ADCY2, CAMK4, ITPR3, ADCY1, PRKAR2B, ADCY8, PRKACG, ADRBK1, ITPR2, ADCY6, PDE1B, ADCY4, CALM1, ADCY9, ADCY7, PRKCA, PRKAR1A, CALM3, PRKCG, PLCG1
GO:0034199 activation of protein kinase A activity	15	15	1.39E- 03	1.04E- 03	PRKACB, ADCY10, ADCY3, PRKAR2A, ADCY5, ADCY2, ADCY1, PRKAR2B, ADCY8, PRKACG, ADCY6, ADCY4, ADCY9, ADCY7, PRKAR1A
PC Tryptophan catabolism	17	16	1.41E- 03	1.22E- 03	NADK, NMNAT1, CCBL2, NMNAT2, KMO, HAAO, ACMSD, KYNU, NMNAT3, TDO2, AADAT, IDO2, CCBL1, NADSYN1, QPRT, AFMID

Fig. 12-5 BPALC vs SNC: TAS2R38 region plot

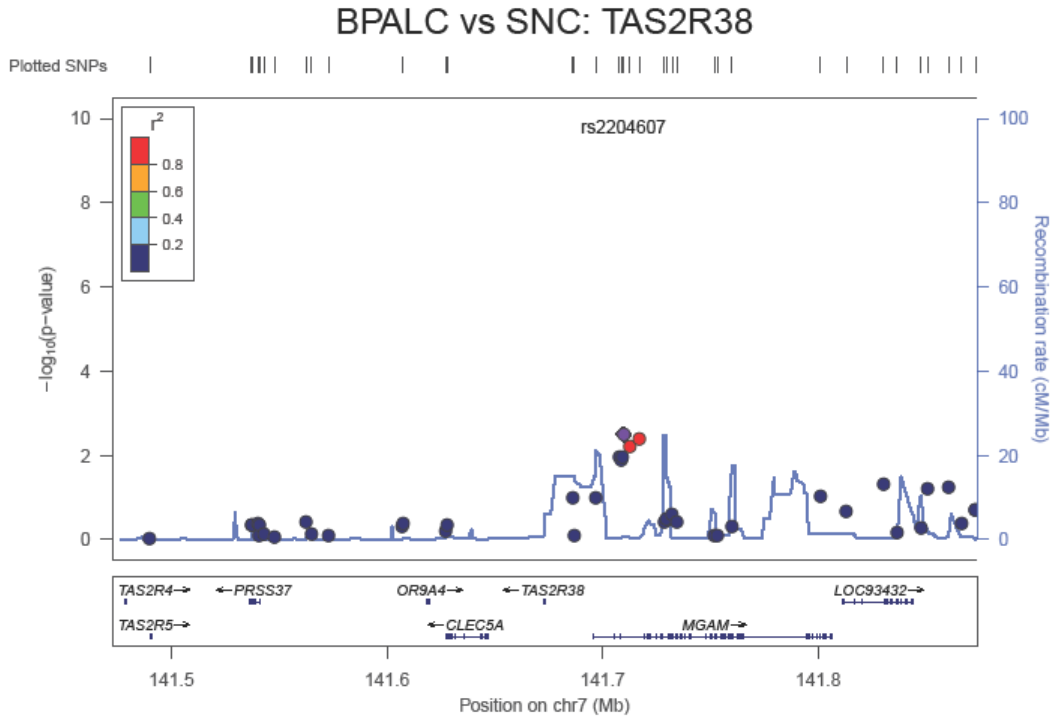


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 12-6 BPALC vs SNC: CDH11 region plot

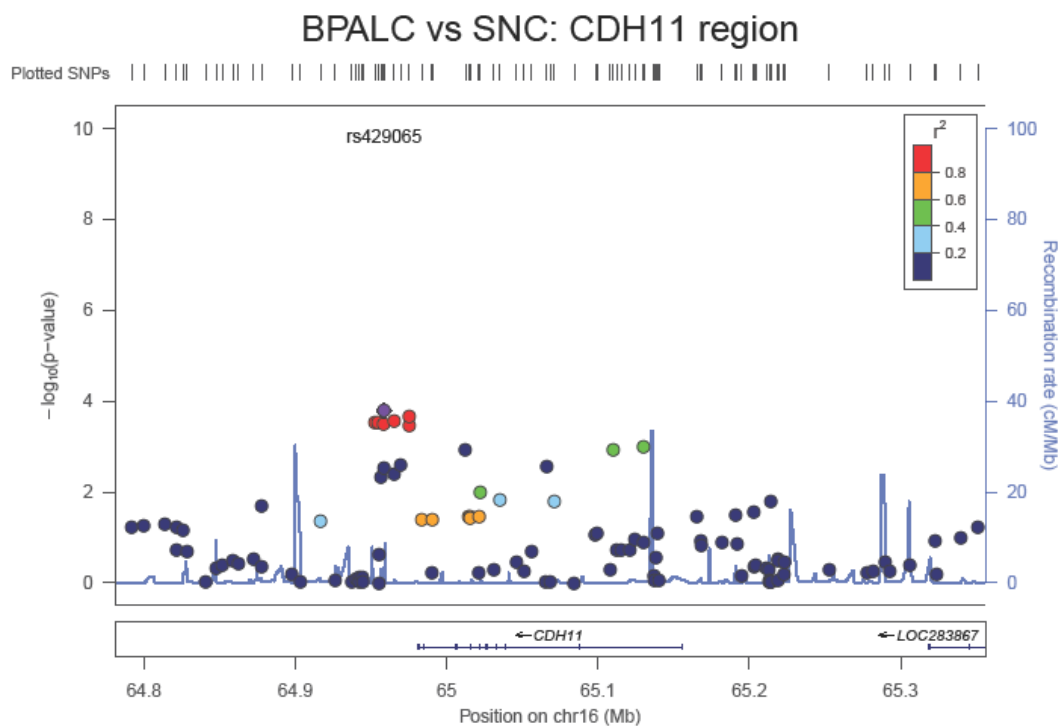


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 12-7 BPALC vs SNC: HAAO region plot

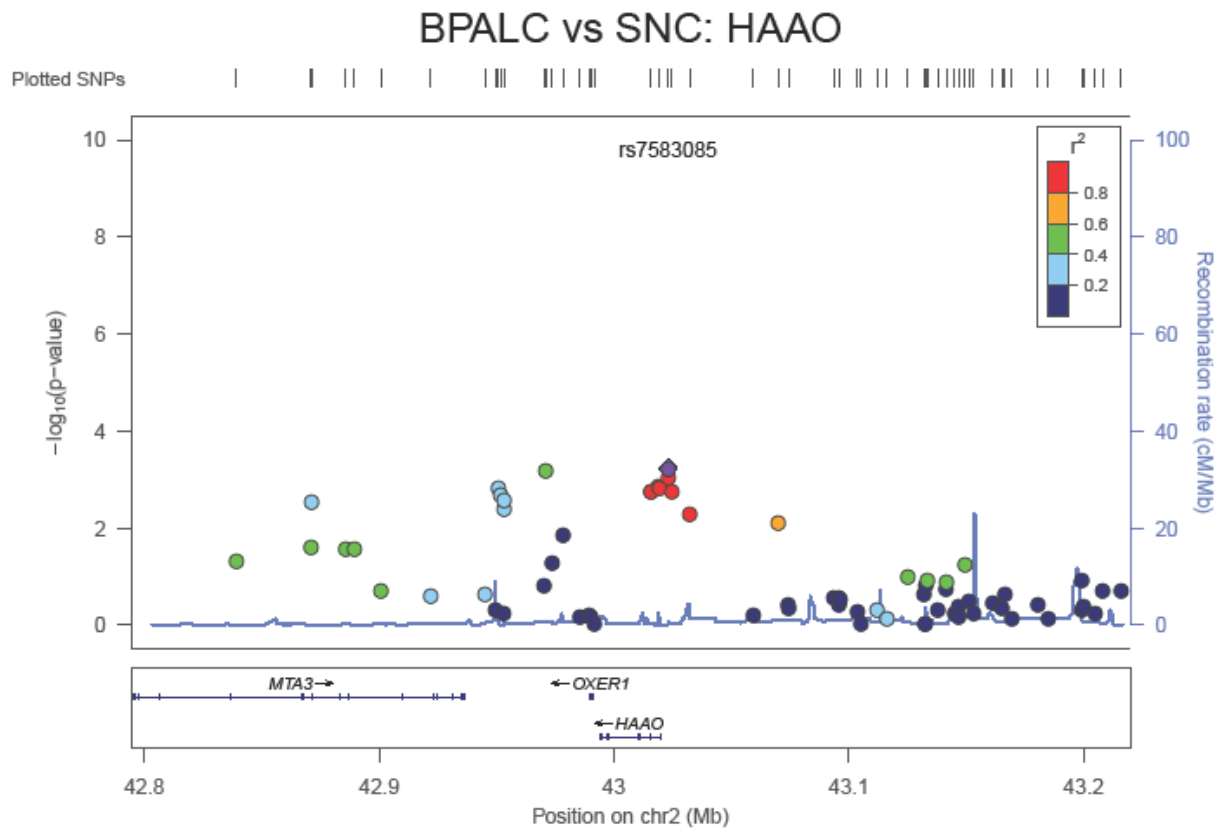


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 12-8 BPALC vs SNC: GABRB3 region plot

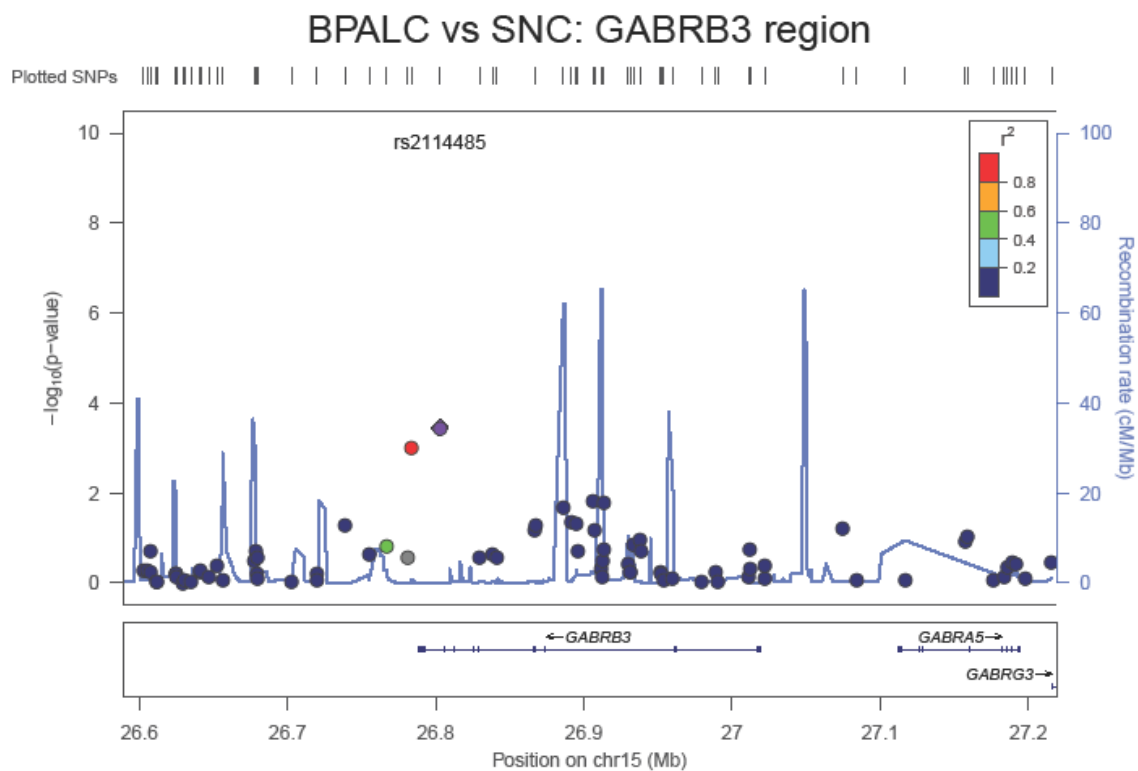


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 12-9 BPALC vs SNC: PANTHER histogram of numbers of genes in each GO pathway

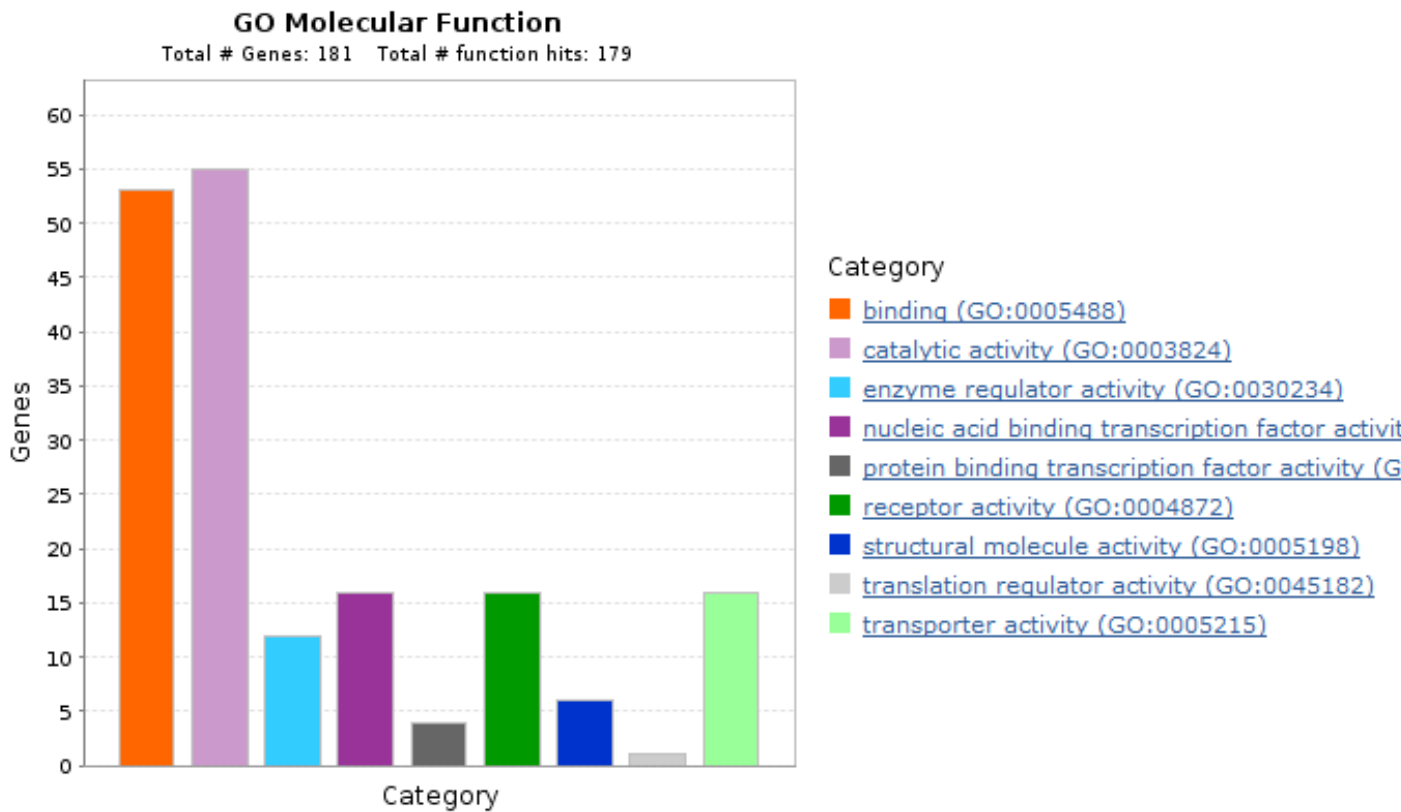


Figure legend: x axis = number of genes; x axis = GO molecular category

CHAPTER 13: RESULTS OF SZALC vs SZNONALC

This analysis was to try and differentiate SZALC genes from SZnonALC controls in a similar way to the BPALC vs BPnonALC analysis. For this analysis of SZALC vs “pure” SZ (without AD, SZnonALC), we examined 77 subjects with SZ and comorbid AD (SZALC) vs 384 SZ without AD (SZnonALC).

We analysed 372,193 SNPs genotyped on the Affymetrix GeneChip Human Mapping Array following the quality control and data cleaning procedure described in Methods, above. The overall genotyping call rate for analysed SNPs was 99.4%. Including the multimarker predictors, our data set is estimated to capture 78.7% of common variation (SNPs with minor allele frequency $\geq 5\%$) in the CEPH HapMap (CEU) samples with $r^2 > 0.8$.

No SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p-values for the autosomal associations in relation to the expected p-values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 13-1. A Manhattan plot of our association analyses is shown in Fig. 13-2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p-values are presented in Table 13.1.

I then checked whether any of the 40 SNPs with the lowest p-value for each analysis were in or close to a gene of known relevant function. None of the top SNPs themselves were previously related directly to AD or alcoholism-related phenotypes, however genes in or near these SNPs are briefly reported.

Our strongest SNP rs4727695 ($p = 1.97 \times 10^{-6}$) was intronic to Laminin, Beta 1 (LAMB1, Fig. 13-3), one of a group of extracellular matrix glycoproteins and the major noncollagenous constituent of basement membranes. Laminins have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signalling, neurite outgrowth and metastasis, and may be involved in the organization of the laminar architecture of cerebral cortex.

2nd SNP by GWA p-value was s7621179 which is intronic to Cell adhesion molecule 2 (CADM2, Fig. 13-4) a gene that codes for a member of the synaptic cell adhesion molecule 1 (SynCAM) family (part of the immunoglobulin superfamily). A lower-ranked SNP rs10191156 ($p = 8.4 \times 10^{-5}$) was intronic to Senataxin (SETX, gene based $p = 5.57 \times 10^{-3}$) which may be involved in both DNA and RNA processing. Mutations in this gene have been associated with ataxia-ocular apraxia-2 (AOA2) and an autosomal dominant form of juvenile amyotrophic lateral sclerosis (ALS4).

I then performed VEGAS2 gene based permutation tests on all SNPs within and next to a previously implicated gene and corrected for multiple tests. In the gene-based tests, I found 45 genes with a p-value < 0.01 , strongest gene long intergenic non-protein coding RNA 1366 (LOC257358, gene based $p = 5.5 \times 10^{-4}$) on Chromosome 5 (Fig. 13-5). The top 40 of these are listed in Table 13.2.

Cross referencing with the literature database of gene and SNP associations, I found 9 genes that were nominally associated ($p \leq 0.05$) that were in common with our VEGAS2 gene wise list. No genes survived Bonferroni correction for multiple tests. All gene replications of AD or related phenotypes and the top SNPs in that gene where $p < 0.05$ from our data are listed in Table 13.4.

Most interesting of this group is CAMK2A (Fig. 13-6) on chromosome 5, gene based $p = 3.13 \times 10^{-2}$ which has a role in learning and memory, and has been associated with both AD (Easton, 2013), BP

(Ament et al. 2015) and transition to cocaine dependence (Easton, 2014). Also of interest is CHRM2, uncorrected gene based $p = 1.86 \times 10^{-2}$, which has been associated with AD, drug dependence and affective disorders (Luo et al. 2005), and AD with depression (Wang et al. 2004). The finding of NPY2R with an uncorrected gene based $p = 5.27 \times 10^{-3}$ is intriguing because of its previous association with AD, alcohol withdrawal symptoms, comorbid alcohol and cocaine dependence, and cocaine dependence (Wetherill et al. 2008). Genes BUB1, CHN1, COBLL1 and TRIB2 have all been associated with AD with conduct disorder, depression and suicide (Dick, 2010). Our finding of SYNGR1 with nominal gene based $p = 0.02$ is of interest because it Chr 22q11-13, a region with consistent linkage regions for SZ and BP but not AD or SUD (Han et al. 2013).

Pathway analysis using VEGAS did not show significant results and will not be reported here.

PANTHER analysis of our top genes implicate in gene based analysis ($P < 0.001$, $n = 236$) revealed adenylate cyclase activity, tryptophan metabolism, solute carriers, voltage gated ion channels, and amino acid modification (Table 13.4) among the main molecular pathways.

In summary, this was the first GWAS comparing patients with SZ and comorbid AD versus “pure” SZ without AD in an attempt to tease out the genetic factors contributing to AD in people with SZ. No marker met genome-wise or corrected gene based significance. However gene based approaches hint at a number of biologically plausible genes involving learning and memory, and synaptic biology.

13.1 DISCUSSION

In this analysis of SZ with and without AD, the hope was to find genes implicated in AD. A number of gene based tests were nominally associated and found firm published evidence to support their involvement in AD.

Perhaps most interesting is the tentative signal of CAMK2A on chromosome 5 (uncorrected gene based $p = 3.13 \times 10^{-2}$) which has a role in learning and memory, and has been associated with both AD (Easton, 2013), BP (Ament et al. 2015) and transition to cocaine dependence in humans and mice (Easton, 2014). Also of interest is CHRM2, gene based $p = 1.86 \times 10^{-2}$, which has been associated with AD, drug dependence and affective disorders (Luo et al. 2005), and AD with depression (Wang et al. 2004). The suggestive finding of NPY2R (uncorrected $p = 5.27 \times 10^{-3}$) is intriguing because of its previous association with AD, alcohol withdrawal symptoms, comorbid alcohol and cocaine dependence, and cocaine dependence (Wetherill et al. 2008).

Our finding of SYNGR1 with an uncorrected gene based $p = 0.02$ is of interest because it is located in Chr 22q11-13, a region with consistent linkage regions for SZ and BP but not AD or SUD (Han et al. 2013). The SYNGR1 gene in this region is associated with presynaptic vesicles in neuronal cells and association studies have confirmed the involvement of SYNGR1 in SZ and BP (Verma et al. 2005; Iatropoulos et al. 2009; Kato 2007) but not in Han Chinese (Wang et al. 2009). SYNGR1 is considered a probable susceptibility gene in both disorders.

This is an unexpected finding because we were explicitly trying to separate AD genes from SZ genes. If this is a true association (noting that it would not survive Bonferroni correction) there are several explanations: 1) It is a SZALC gene which has not previously been seen because of phenotypic heterogeneity. It is possible that previous SZ GWA studies did not exclude SZALC comorbidity leading to a broader phenotype. Indeed a detailed literature search found no evidence that the main studies excluded patients with AD or alcohol misuse (i.e. O'Donovan et al. 2008; Shi et al. 2009; Moskvina et al. 2009; Potkin et al. 2009; Ripke et al. 2011; Walters et al. 2013; Fanous et al. 2012; Ripke et al. 2013); 2) It is an AD gene commonly inherited with SZ (however it did not feature in the SZALC vs normal controls analysis). In the next chapter I report on the SZALC vs SNC analysis.

Table 13.1 SZALC vs SZnonALC: Top 40 SNPs and gene based results

CHR	SNP	Position (base pairs)	A1	AF Cases	AF Control	A2	CHISQ	P	OR	Position	Distance	Nearest gene code	Gene based p*
7	rs4727695	107614003	G	0.2078	0.08073	A	22.62	1.97E-06	2.987	intron	0	LAMB1	6.31E-02
3	rs7621179	85820416	G	0.07143	0.01172	A	21.55	3.45E-06	6.487	intron	0	CADM2	4.31E-01
11	rs7106819	12570446	G	0.5519	0.3581	C	20.26	6.76E-06	2.208	upstream	125523	TEAD1	6.97E-01
20	rs17800346	11552387	C	0.2273	0.09922	A	19.82	8.52E-06	2.67	upstream	225884	---	
7	rs3893568	107629659	T	0.2597	0.1224	C	19.48	1.02E-05	2.516	intron	0	LAMB1	6.31E-02
12	rs2010060	53070502	T	0.5068	0.3184	A	19.14	1.21E-05	2.2	intron	0	KRT1	1.05E-03
8	rs13281810	89780482	G	0.6623	0.4701	A	18.97	1.33E-05	2.211	downstream	31119	---	
9	rs2900171	102251637	A	0.09091	0.02214	T	18.67	1.55E-05	4.418	upstream	42341	LOC100507438	
9	rs16918616	102272438	A	0.09091	0.02214	G	18.67	1.55E-05	4.418	upstream	21540	LOC100507438	
6	rs12213573	4210858	T	0.5649	0.3776	C	18.62	1.59E-05	2.14	downstream	20818	---	
1	rs17427125	5706597	A	0.08442	0.01953	C	18.34	1.85E-05	4.628	intron	0	---	
1	rs1499396	5707224	C	0.08442	0.01953	T	18.34	1.85E-05	4.628	intron	0	---	
18	rs524273	24390432	C	0.2987	0.4857	A	18.09	2.11E-05	0.451	intron	0	AQP4-AS1	7.19E-02
4	rs9654198	114710727	C	0.3571	0.2005	T	17.95	2.27E-05	2.215	upstream	27644	CAMK2D	1.79E-02
4	rs4642242	19555435	A	0.5526	0.3708	C	17.47	2.92E-05	2.097	intron	0	---	
3	rs2042686	107066743	G	0.5	0.3255	A	17.06	3.62E-05	2.072	intron	0	LOC101929579	
1	rs17467353	5706560	G	0.08442	0.02083	A	17.02	3.69E-05	4.333	intron	0	---	
4	rs4417971	114753435	A	0.3506	0.1992	G	16.9	3.95E-05	2.171	upstream	70352	CAMK2D	1.79E-02
5	rs2731694	171878751	A	0.5065	0.3329	G	16.73	4.31E-05	2.057	intron	0	SH3PXD2B	1.51E-02
1	rs10875053	97636966	G	0.2922	0.1549	C	16.53	4.80E-05	2.252	intron	0	DPYD	1.86E-02
4	rs6827998	114710196	T	0.3506	0.201	G	16.4	5.12E-05	2.146	upstream	27113	CAMK2D	1.79E-02
16	rs16950218	79352274	G	0.06494	0.01302	A	16.29	5.43E-05	5.264	downstream	53804	---	
16	rs16950213	79351872	G	0.06494	0.01302	A	16.29	5.43E-05	5.264	downstream	53402	---	
12	rs1561560	8762963	C	0.1494	0.05729	T	16.13	5.91E-05	2.889	intron	0	AICDA	5.99E-02
10	rs2812495	71418185	C	0.6234	0.4466	A	16.07	6.11E-05	2.051	downstream	24833	C10orf35	2.66E-04

Table 13.1 contd...

10	rs2616076	71418213	G	0.6234	0.4466	T	16.07	6.11E-05	2.051	downstream	24861	C10orf35	2.66E-04
4	rs7667580	19558358	T	0.5455	0.3724	C	15.99	6.37E-05	2.022	intron	0	---	
1	rs1474797	186967193	G	0.3636	0.2135	A	15.91	6.65E-05	2.105	downstream	9080	PLA2G4A	1.69E-02
1	rs11165805	97652000	G	0.2857	0.1523	C	15.83	6.92E-05	2.226	intron	0	DPYD	1.86E-02
16	rs11150443	82171367	C	0.2013	0.09162	T	15.74	7.25E-05	2.499	intron	0	---	
12	rs1464042	53091326	T	0.5584	0.3867	C	15.58	7.92E-05	2.006	intron	0	KRT77	6.66E-04
3	rs3774081	10961939	A	0.1234	0.04308	T	15.5	8.24E-05	3.126	intron	0	SLC6A11	7.49E-02
2	rs1192340	84948021	A	0.1234	0.04308	G	15.5	8.24E-05	3.126	intron	0	DNAH6	2.22E-03
9	rs10901156	135184237	G	0.1494	0.05859	A	15.47	8.40E-05	2.821	intron	0	SETX	5.57E-03
17	rs1532457	31607870	A	0.2468	0.4154	G	15.37	8.84E-05	0.4611	intron	0	ASIC2	9.79E-02
16	rs2081257	82170725	C	0.2039	0.09424	G	15.29	9.23E-05	2.462	intron	0	---	
16	rs8047857	82165577	C	0.2532	0.1302	A	15.17	9.84E-05	2.265	intron	0	---	
7	rs6973199	138411712	C	0.2078	0.09766	T	15.17	9.84E-05	2.424	intron	0	ATP6V0A4	4.12E-02
4	rs4128132	19562384	T	0.5455	0.3766	C	15.15	9.93E-05	1.986	intron	0	---	
11	rs16916055	91160029	C	0.04545	0.00651	T	15.15	9.96E-05	7.267	downstream	259567	---	

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F_A= Frequency of Allele 1; F_U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. *Uncorrected p <0.05 in **bold**.

Table 13.2 SZALC vs SZ Top 40 Genes After VEGAS2 Permutation Testing

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Uncorrected gene based p	Top SNP	Top SNP P
5	LOC257358	24	1.00E+06	116.76	6.50E-05	rs17072564	5.55E-04
2	MEIS1-AS3	5	1.00E+06	41.86	1.02E-04	rs10496135	2.99E-04
2	C2orf82	16	1.00E+06	118.17	2.09E-04	rs1284	1.36E-04
10	C10orf35	22	1.00E+06	102.11	2.66E-04	rs2616076	6.11E-05
3	CCDC54	14	1.00E+06	66.73	2.68E-04	rs2042686	3.62E-05
14	PPP2R5C	11	1.00E+06	50.45	2.95E-04	rs8016207	6.33E-04
2	GIGYF2	28	1.00E+06	144.27	3.07E-04	rs1284	1.36E-04
16	HSD17B2	30	1.00E+06	144.50	5.06E-04	rs11150443	7.25E-05
4	PDGFC	17	1.00E+06	112.07	5.31E-04	rs17811863	1.35E-04
16	MPHOSPH6	35	1.00E+06	228.09	5.62E-04	rs11150443	7.25E-05
11	POLD3	12	1.00E+06	75.03	6.30E-04	rs10793095	3.60E-04
18	ACAA2	12	1.00E+06	64.29	6.38E-04	rs7244563	6.04E-04
17	AXIN2	10	1.00E+06	54.15	6.54E-04	rs7216210	1.51E-04
12	KRT77	32	1.00E+06	151.45	6.66E-04	rs2010060	1.21E-05
11	LOC100507384	17	1.00E+06	85.85	6.85E-04	rs6485638	3.00E-03
11	MIR7154	19	1.00E+06	100.04	6.95E-04	rs6485638	3.00E-03
19	KLK7	13	1.00E+06	68.22	7.10E-04	rs1654559	5.36E-04
2	NGEF	42	1.00E+06	150.15	7.39E-04	rs1284	1.36E-04
18	AQP4	13	1.00E+06	51.86	7.58E-04	rs524273	2.11E-05
19	KLK6	16	1.00E+06	70.52	8.43E-04	rs1654559	5.36E-04
19	KLK5	14	1.00E+06	63.88	8.56E-04	rs1654559	5.36E-04
8	CHMP7	8	1.00E+06	34.87	8.92E-04	rs12544414	6.12E-04
12	KRT2	26	1.00E+06	127.18	9.06E-04	rs2010060	1.21E-05
12	KRT1	27	1.00E+06	132.32	1.05E-03	rs2010060	1.21E-05
19	KLK9	10	1.00E+05	55.45	1.08E-03	rs1654559	5.36E-04
21	MIR6508	7	1.00E+05	38.90	1.16E-03	rs2837006	2.28E-03

Table 13.2 contd...

14	RBM25	12	1.00E+05	58.57	1.20E-03	rs362350	1.40E-03
3	LINC00883	14	1.00E+05	52.91	1.20E-03	rs2042686	3.62E-05
19	KLK8	10	1.00E+05	55.45	1.21E-03	rs1654559	5.36E-04
1	DPYD-AS1	22	1.00E+05	79.73	1.21E-03	rs10875053	4.80E-05
14	PSEN1	12	1.00E+05	53.59	1.25E-03	rs2293796	8.98E-04
2	COL6A3	27	1.00E+05	93.22	1.40E-03	rs4433949	8.43E-04
9	TTF1	8	1.00E+05	41.23	1.41E-03	rs4637976	2.24E-03
4	PIGG	8	1.00E+05	52.46	1.42E-03	rs4499656	1.71E-04
6	RPS10	7	1.00E+05	32.26	1.48E-03	rs6912971	4.04E-03
19	KLK11	12	1.00E+05	55.56	1.50E-03	rs1654559	5.36E-04
19	KLK10	12	1.00E+05	55.56	1.71E-03	rs1654559	5.36E-04
8	R3HCC1	16	1.00E+05	73.91	1.72E-03	rs12544414	6.12E-04
19	LILRA2	9	1.00E+05	36.52	1.77E-03	rs10402506	1.97E-03
5	KCNMB1	16	1.00E+05	68.55	1.88E-03	rs17072564	5.55E-04

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations

Table 13.3 SZALC vs SZ Genes Yielding Evidence for Association at P <0.05 After Permutation Testing and Literature Replication

Chr	Gene symbol	n SNPs	n Sims	VEGAS Test result	Uncorrected gene based p-value	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
7	CHRM2	34	10000	97.98	1.86E-02	rs6943675	2.38E-03	AD/drug dependence/affective disorders (Luo et al. 2005), and AD with depression (Wang et al. 2004).
11	BDNF	14	10000	33.98	3.85E-02	rs11602246	5.72E-03	BP (Serretti and Mandelli 2008; Craddock and Sklar 2009).
7	GRM8	101	10000	252.06	1.02E-02	rs11763603	5.15E-03	Theta EEG and alcoholism (Chen, 2009)
3	GHSR	9	10000	27.59	1.82E-02	rs16845462	2.21E-02	Alcohol consumption and body weight (Landgren, et al. 2008)
4	NPY2R	10	100000	40.43	5.72E-03	rs2342653	8.59E-03	AD, alcohol withdrawal symptoms, comorbid alcohol and cocaine dependence, and cocaine dependence (Wetherill et al. 2008).
2	BUB1	3	10000	9.43	3.22E-02	rs4443010	1.45E-02	AD with conduct disorder, depression and suicide (Dick, 2010)
2	CHN1	19	100000	81.96	8.22E-03	rs17222746	3.21E-03	
2	COBLL1	33	10000	77.81	2.48E-02	rs16849792	7.46E-03	
2	TRIB2	16	10000	42.59	4.93E-02	rs477599	1.57E-02	
5	CAMK2A	22	10000	44.06	3.13E-02	rs9688182	6.23E-03	
22	SYNGR1	10	10000	41.02	2.11E-02	rs1010170	1.03E-02	Nonsyndromic intellectual disability (Berrier, 2012, Hamdan, 2011)

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations

Fig. 13-1 Q-Q plot of SZALC vs SZ

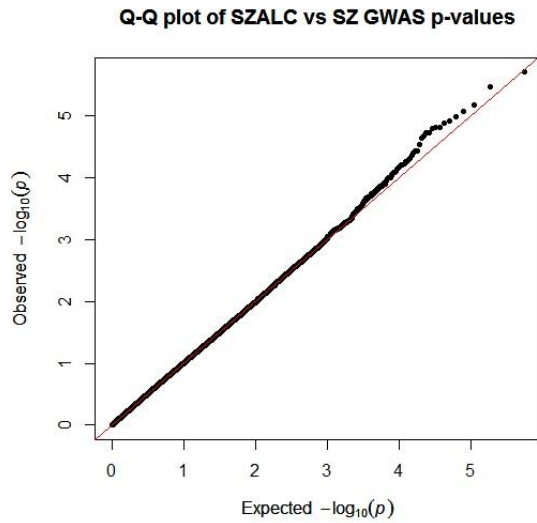


Figure legend: Q-Q plots of observed and expected $-\log_{10}(P)$ of the associations between SNPs

Fig. 13-2 Manhattan plot of SZALC vs SZ

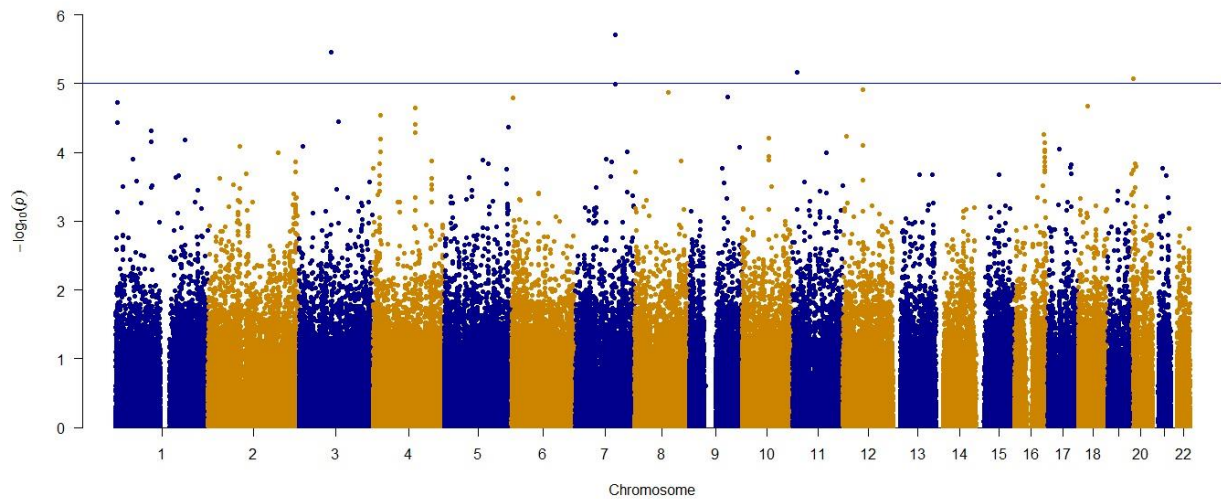


Figure legend: Results of the genome-wide association analyses for SZALC v SZ. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Fig. 13-3 SZALC vs SZnonALC: Top SNP in LAMB1

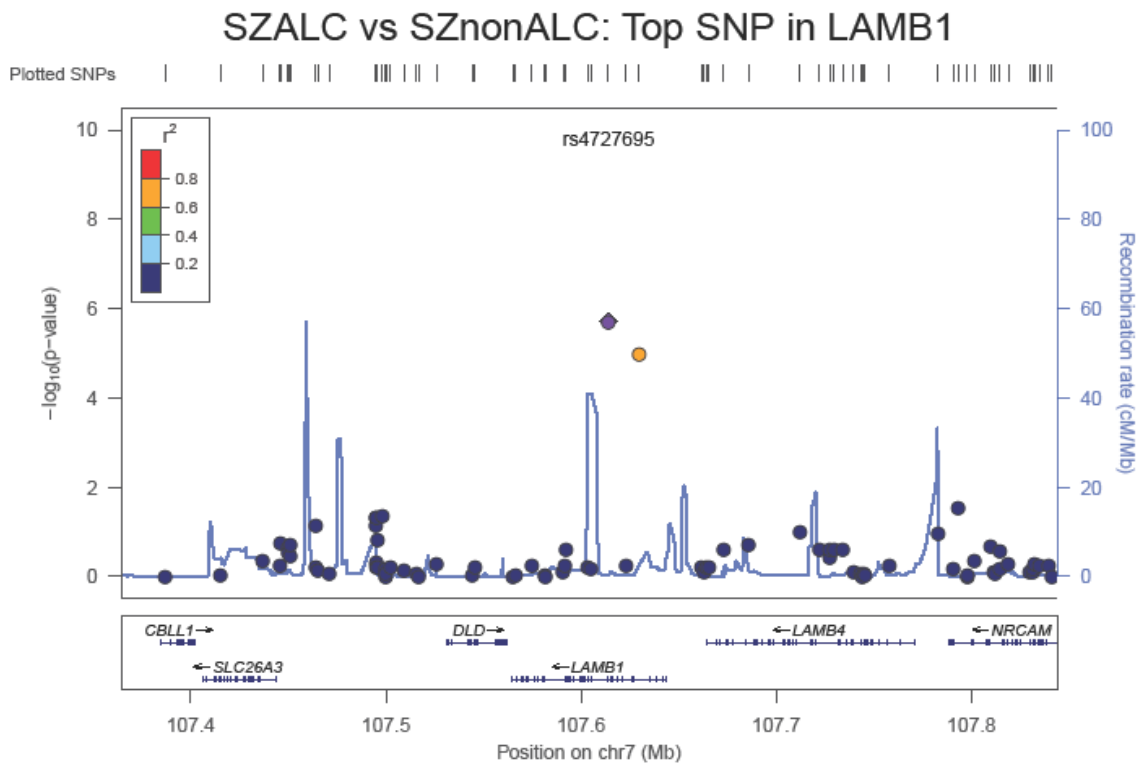


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 13-4 SZALC vs SZnonALC: CADM2 region

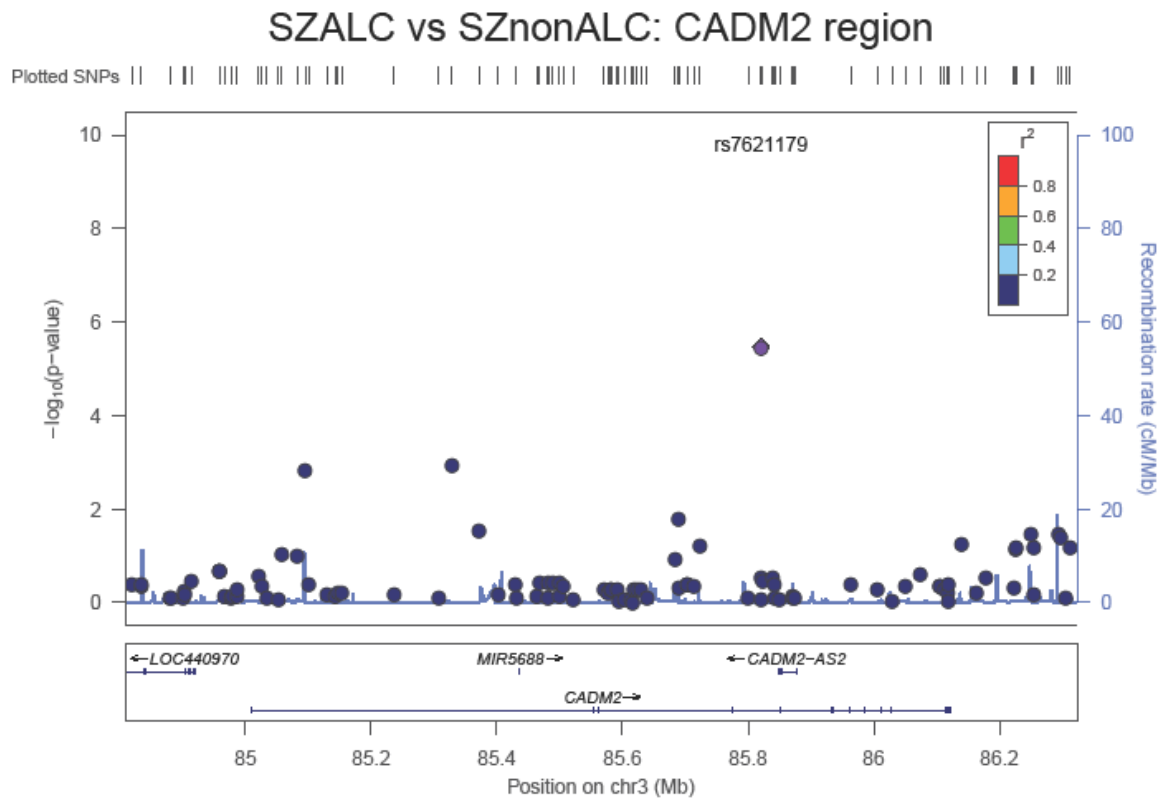


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

CHAPTER 14: RESULTS OF SZALC VS NORMAL CONTROLS

The current analysis aimed to try and differentiate SZALC genes from supernormal controls. Unlike the previous chapter which sought AD genes as separate from SZ, this approach looks at the combined SZALC phenotype – a common and clinically important dual diagnosis combination, similar to the BPALC approach in Chapter 12.

We analysed 372,193 SNPs genotyped on the Affymetrix GeneChip Human Mapping Array following the quality control and data cleaning procedure described in Methods, above. The overall genotyping call rate for analysed SNPs was 99.4%. Including the multimarker predictors, our data set is estimated to capture 78.7% of common variation (SNPs with minor allele frequency $\geq 5\%$) in the CEPH HapMap (CEU) samples with $r^2 > 0.8$. For this analysis of SZALC vs normal ancestrally matched controls, we examined 77 subjects with SZ and comorbid AD (SZALC) vs 510 normal controls.

No SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p-values for the autosomal associations in relation to the expected p-values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 14.1. A Manhattan plot of our association analyses is shown in Fig. 14.2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p-values are presented in Table 14.1.

I then checked whether any of the 40 SNPs with the lowest p-value for each analysis were in or close to a gene of known relevant function. None of the top SNPs themselves were previously related directly to AD or alcoholism-related phenotypes, however genes in or near these SNPs are briefly reported.

Our strongest SNP rs17030670 on Chromosome 2, intronic to uncharacterised gene LOC101927533 ($p = 5.89 \times 10^{-6}$, Fig. 14-3).

I then performed VEGAS2 based permutation tests on all SNPs within and next to a previously implicated gene and corrected for multiple tests. In the gene-based tests, we found 43 genes with $p < 0.01$, strongest gene LOC101927166 on Chromosome 17, gene based $p = 4.96 \times 10^{-5}$ (Fig. 14-4). The top 40 of these are gene based results listed in Table 14.2.

Cross referencing with the literature database of gene and SNP associations, I found 29 genes that were nominally associated ($p \leq 0.05$) that were in common with our VEGAS2 gene based list. No genes survived Bonferroni correction for multiple tests.

Most notable of these is the opioid precursor prodynorphin (PDYN, uncorrected gene based $p = 2.03 \times 10^{-2}$, Fig. 14-5) which is a strong addictions candidate gene from the literature. It has been implicated in disinhibited behaviour (Flory et al. 2011); opioid dependence in African Americans (Ray et al. 2005); cocaine dependence (Dahl et al. 2005); cocaine and alcohol dependence in African Americans (Williams et al. 2007); methamphetamine dependence (Nomura et al. 2006); and altered brain methylation in alcoholics (Taqi et al. 2011). PDYN has also been implicated in SZ (Zhang et al. 2004) and less strongly in SZ by epistatic interaction with the DRD3 Gly allele (Ventriglia et al. 2002).

Again we see gene based suggestive results from CHRM2, GRM8, BDNF, GPD1L and TLR1.

Fascinatingly we found a gene based signal from both neuropeptide Y (NPY) and the neuropeptide Y 2 receptor (NPY2R, Fig. 14-6) which are both associated with AD and alcohol withdrawal symptoms (Bhaskar et al, 2013; Wetherill et al 2008). All gene replications of AD or related phenotypes and the top SNPs in that gene where $p < 0.01$ from our data are listed in Table 14.3.

VEGAS2 Pathway analysis did not show significant results and are not presented here.

In summary, this was the first small GWAS comparing patients with SZ and comorbid AD with normal controls in an attempt to tease out the SZALC phenotype (as opposed to AD) genetic effects. No marker was genome-wide significant. However gene wise and pathway approaches offer a number of biologically plausible genes involving the opioid, cholinergic and glutamatergic systems, and neurotrophins.

14.1 DISCUSSION

In this analysis of SZALC vs SNC, the hope was to find genes implicated in the SZALC phenotype. Again, despite a larger sample size, no marker met genome wide significance thresholds. A number of gene based tests were nominally associated and found firm published evidence to support their involvement in AD. I will mention the most significant in my opinion.

PDYN1 is of significant interest because of the key role of the opioid system in reward, pleasure and the strong evidence for involvement in addictive behaviours (see Section 5.2.5). PDYN1 has much circumstantial evidence supporting its role in addiction: Methylation of PDYN CpG-SNPs in post-mortem brains of alcoholics was found to be different to control subjects (Taqi et al. 2011). PDYN1 has been implicated (weakly) in opiate dependence in African Americans (Ray et al. 2005); cocaine dependence (Dahl et al. 2005); cocaine and alcohol dependence in African Americans (Williams et al. 2007) ; methamphetamine dependence (Nomura et al. 2006); and disinhibited behaviour (Flory et al. 2011).

Flory's analysis of AD and disinhibited behaviour did not find an association between the PDYN polymorphism and the diagnosis of AD, however they did observe an association between the "low"

expressing L allele of the PDYN gene and disinhibited behaviour (Flory et al. 2011). The authors suggest that variation in the PDYN gene is associated with a dimensional trait or intermediate phenotype that reflects a preference for heavy drinking and engaging in related risky behaviours. Unsafe and impulsive behaviours are often evident during episodes of SZ and BP, with risk increased when the patient has been drinking. It is an interesting theory that this effect may be mediated through PDYN's effects on risky behaviours – which may be a useful phenotype for future work.

Cholinergic muscarinic 2 receptor (CHRM2) is implicated in memory and cognition, which are both key functions impaired in many neuropsychiatric disorders. Association has been found between CHRM2 and AD and major depressive syndrome (Wang et al. 2004); and AD/drug dependence/affective disorders (Luo et al. 2005).

Neuropeptide (NPY) is a neurotransmitter involved in cortical excitability, stress response, food intake, circadian rhythms, and cardiovascular function. It has been associated with AD (Bhaskar et al. 2013; Mottagui-Tabar et al. 2005); and alcohol withdrawal seizures (Okubo et al. 2003). NPY2R has been associated with AD, alcohol withdrawal symptoms, comorbid alcohol and cocaine dependence, and cocaine dependence (all $p < 0.03$) (Wetherill et al. 2008). A German study found no association with AD (Zill et al. 2008).

14.2 SZALC GWAS CONCLUSIONS

This larger GWAS of SZALC vs normal controls suggested some interesting gene signals implicating the opioid, cholinergic and glutamatergic systems in SZALC.

In the next chapter I report on the first meta-analysis of combined psychosis and AD phenotype versus psychosis without AD.

Fig. 14-1: Q-Q plot of SZALC vs SNC

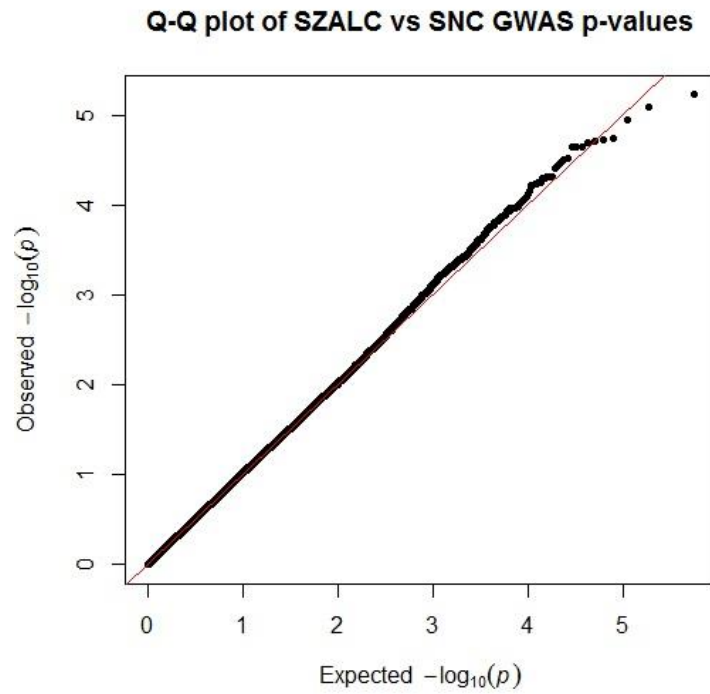


Figure legend: Q-Q plots of observed and expected $-\log_{10}(P)$ of the associations between SNPs

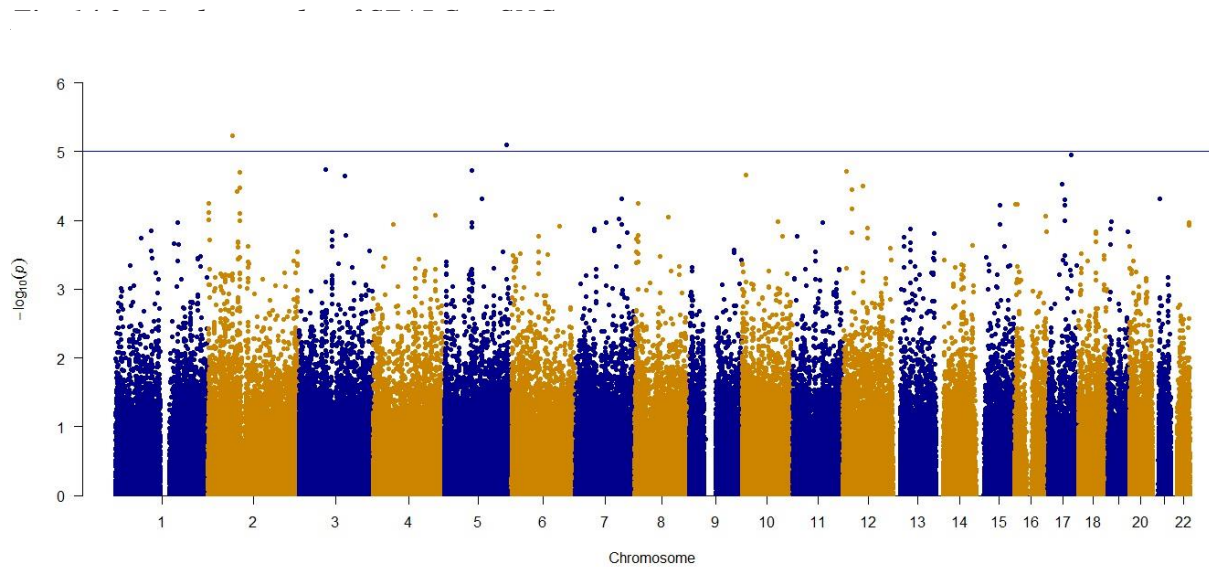


Figure legend: Results of the genome-wide association analyses for SZALC vs SNC. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Table 14.1 SZALC vs SNC: Top 40 SNPs and gene based significance values

CHR	SNP	Position (base pairs)	A1	AF Cases	AF Control	A2	CHISQ	P	Position	Distance from gene	Nearest gene code	Gene based p
2	rs17030670	65935157	T	0.1883	0.066	C	20.53	5.89E-06	intron	0	LOC101927533	---
5	rs17498388	166775014	G	0.539	0.3386	A	19.96	7.91E-06	intron	0	TENM2	4.00E-02
17	rs4791140	63660230	T	0.2403	0.102	C	19.33	1.1E-05	intron	0	CEP112	1.06E-01
3	rs11709093	71128054	A	0.3701	0.2012	G	18.39	1.8E-05	intron	0	FOXP1	3.33E-04
5	rs9942407	74540187	A	0.5325	0.3406	G	18.3	1.88E-05	downstream	75793	---	---
12	rs1561560	8762963	C	0.1494	0.04781	T	18.27	1.92E-05	intron	0	AICDA	2.67E-02
2	rs1192340	84948021	A	0.1234	0.03386	G	18.21	1.98E-05	intron	0	DNAH6	1.37E-03
10	rs1750738	11156472	C	0.1948	0.0757	T	17.99	2.22E-05	intron	0	CELF2	1.35E-02
3	rs7652858	121998105	T	0.05844	0.006	C	17.98	2.24E-05	intron	0	CASR	3.39E-02
3	rs2134225	122002253	A	0.05844	0.006	C	17.98	2.24E-05	intron	0	CASR	3.39E-02
17	rs2315504	39046881	C	0.3377	0.5301	A	17.44	2.96E-05	upstream	5402	KRT20	4.60E-04
12	rs2010060	53070502	T	0.5068	0.3187	A	17.35	3.11E-05	intron	0	KRT1	1.29E-03
2	rs1947465	84618105	C	0.1169	0.032	T	17.21	3.34E-05	downstream	32542	SUCLG1	1.91E-04
12	rs1487654	24820731	G	0.4026	0.2331	C	17.07	3.6E-05	downstream	24631	---	---
2	rs6708183	78384746	T	0.2105	0.3924	C	16.97	3.81E-05	downstream	255173	---	---
21	rs719500	19036144	A	0.2792	0.4641	G	16.54	4.78E-05	upstream	3442	---	---
5	rs4515270	101180049	A	0.4805	0.3024	T	16.5	4.86E-05	upstream	27565	---	---
7	rs1526408	125045427	T	0.2013	0.08367	A	16.49	4.89E-05	downstream	26052	LOC101928283	7.79E-02
17	rs1027894	46527718	A	0.1104	0.02988	G	16.46	4.96E-05	upstream	10014	---	---
17	rs8081430	46541483	C	0.1104	0.02988	G	16.46	4.96E-05	intron	0	---	---
8	rs2945908	8192229	C	0.2468	0.1155	A	16.24	5.58E-05	intron	0	SGK223	8.56E-04
2	rs7355580	970417	T	0.2013	0.3765	C	16.23	5.62E-05	intron	0	SNTG2	6.56E-03
16	rs8046065	3838298	T	0.2013	0.08434	C	16.18	5.75E-05	intron	0	CREBBP	7.93E-03
16	rs9930513	7886652	A	0.1623	0.05976	G	16.14	5.87E-05	upstream	109699	---	---

Table 14.1 contd...

17	rs11655120	46542355	T	0.1104	0.03024	A	16.13	5.92E-05	intron	0	---	---
15	rs10519180	63126806	T	0.08442	0.01793	G	16.07	6.1E-05	intron	0	TLN2	3.80E-02
12	rs1976130	24818342	C	0.4221	0.255	T	15.85	6.87E-05	downstream	22242	---	---
2	rs6761698	973109	C	0.2078	0.3805	T	15.66	7.57E-05	intron	0	SNTG2	6.56E-03
2	rs13000280	84696996	A	0.1118	0.03187	G	15.57	7.94E-05	upstream	46583	DNAH6	1.37E-03
4	rs922418	168525403	A	0.1429	0.3028	G	15.47	8.4E-05	upstream	203498	---	---
16	rs4782854	83911556	T	0.1776	0.3454	G	15.43	8.58E-05	upstream	21175	MLYCD	2.80E-01
8	rs13281810	89780482	G	0.6623	0.482	A	15.35	8.92E-05	downstream	31119	---	---
7	rs10953842	117044682	T	0.1908	0.07968	G	15.26	9.36E-05	intron	0	ASZ1	1.09E-02
2	rs6710556	969866	C	0.2727	0.4494	T	15.18	9.79E-05	intron	0	SNTG2	6.56E-03
2	rs4832087	84724673	C	0.1104	0.032	G	15.1	0.000102	upstream	18906	DNAH6	1.37E-03
17	rs12449855	46485603	C	0.1104	0.032	T	15.1	0.000102	intron	0	SKAP1	1.33E-02
10	rs10882205	95032661	C	0.3896	0.2311	T	15.09	0.000103	downstream	11100	---	---
19	rs10403787	10166375	A	0.2792	0.1434	G	15.03	0.000106	exon	0	C3P1	1.38E-03
7	rs2367910	81786043	A	0.2078	0.092	C	15.03	0.000106	intron	0	CACNA2D1	1.07E-01
1	rs12061855	167426797	C	0.04545	0.003984	G	14.98	0.000109	intron	0	CD247	4.01E-01

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F_A= Frequency of Allele 1; F_U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. Uncorrected p <0.05 in **bold**.

Table 14.2 SZALC vs SNC: SZ Top 40 Genes After VEGAS2 Permutation Testing

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Uncorrected gene based P value	Top SNP	Top SNP p value
17	LOC101927166	9	1.00E+06	82.2973952	1.46E-04	rs1027894	4.96E-05
3	EPHA3	31	1.00E+06	211.784169	1.48E-04	rs907713	1.45E-04
2	SUCLG1	11	1.00E+06	97.249129	1.91E-04	rs1947465	3.34E-05
18	ACAA2	12	1.00E+06	71.6553817	2.70E-04	rs669350	1.48E-04
10	UNC5B	24	1.00E+06	76.8357146	2.89E-04	rs6480502	2.59E-03
3	FOXP1	67	1.00E+06	196.287595	3.33E-04	rs11709093	1.80E-05
13	MIR8073	25	1.00E+06	90.5143254	3.68E-04	rs9521717	1.55E-04
2	LINC01249	15	1.00E+06	82.1009024	3.97E-04	rs10187813	1.90E-04
17	KRT20	10	1.00E+06	47.8296839	4.60E-04	rs2315504	2.96E-05
19	RDH8	6	1.00E+06	28.8757304	4.62E-04	rs10403787	1.06E-04
21	LCA5L	13	1.00E+06	91.5925885	5.15E-04	rs2837006	6.64E-04
21	MIR6508	7	1.00E+06	43.2985084	5.30E-04	rs2837006	6.64E-04
6	SMAP1	14	1.00E+06	83.9298365	5.69E-04	rs781977	1.70E-04
12	MFAP5	5	1.00E+06	29.828979	5.82E-04	rs1561560	1.92E-05
5	ANKRD31	16	1.00E+06	105.768886	6.42E-04	rs9942407	1.88E-05
18	SCARNA17	7	1.00E+06	48.8804324	6.62E-04	rs669350	1.48E-04
10	UNC5B-AS1	11	1.00E+06	47.1054825	8.56E-04	rs6480502	2.59E-03
8	SGK223	25	1.00E+06	114.394018	8.56E-04	rs2945908	5.58E-05
11	LINC00301	9	1.00E+06	74.2288671	9.62E-04	rs11825362	2.86E-04
21	WRB	14	1.00E+06	84.6687584	9.98E-04	rs2837006	6.64E-04
18	SNHG22	23	1.00E+05	170.219148	1.19E-03	rs669350	1.48E-04
3	PIK3CA	10	1.00E+05	54.2942636	1.28E-03	rs3975506	8.64E-04
17	KRT12	10	1.00E+05	41.2119099	1.29E-03	rs2315504	2.96E-05
2	DNAH6	34	1.00E+05	180.985815	1.37E-03	rs1192340	1.98E-05
19	C3P1	8	1.00E+05	28.924666	1.38E-03	rs10403787	1.06E-04
17	EPN2	2	1.00E+05	13.5660971	1.40E-03	rs3862148	1.62E-03
3	MIR1284	6	1.00E+05	37.0379343	1.40E-03	rs6780776	2.08E-03
5	HMGCR	18	1.00E+05	91.1304538	1.43E-03	rs2335418	1.09E-04
11	LINC00678	10	1.00E+05	56.3450902	1.45E-03	rs4923460	2.42E-03
4	NPY2R	10	1.00E+05	51.5342839	1.45E-03	rs2342653	1.41E-03
9	TPD52L3	16	1.00E+05	98.9996404	1.47E-03	rs2381438	4.82E-04
9	NACC2	14	1.00E+05	41.7128461	1.50E-03	rs4842092	3.82E-04
20	SIRPA	19	1.00E+05	64.1855772	1.62E-03	rs6035140	3.00E-03
15	PEX11A	3	1.00E+05	17.615686	1.63E-03	rs8179043	4.58E-04
5	ZCCHC10	9	1.00E+05	77.9505451	1.64E-03	rs7721953	4.96E-04
2	MIR548AD	13	1.00E+05	100.688985	1.65E-03	rs1011798	6.21E-04
19	COL5A3	10	1.00E+05	32.4166066	1.73E-03	rs10403787	1.06E-04
8	C8orf46	15	1.00E+05	57.6899107	1.79E-03	rs1030420	1.87E-03
9	TTF1	8	1.00E+05	39.6917086	1.79E-03	rs7037570	8.36E-03

Table 14.3 SZALC vs SNC Genes Yielding Evidence for Association at P <0.05 After Permutation Testing and Literature Replication

Chr	Gene symbol	n SNPs	n Sims	VEGAS Test result	Uncorrected gene based p-value	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
20	PDYN	19	10000	46.11	2.03E-02	rs6035140	3.00E-03	Disinhibited behaviour (Flory et al. 2011); opioid dependence in African Americans (Ray et al. 2005); cocaine dependence (Dahl et al. 2005); cocaine and alcohol dependence (Williams et al. 2007); methamphetamine dependence (Nomura et al. 2006); altered brain methylation AD (Taqi et al. 2011). SZ (Zhang et al. 2004). SZ by epistatic interaction with DRD3 Gly allele (Ventriglia et al. 2002).
8	NAT1	29	10000	55.19	4.50E-02	rs13268900	1.93E-02	NAT1: severity of substance abuse (Comings, Muhleman, Wu, & MacMurray, 2000)
7	NPY	28	10000	72.15	3.66E-02	rs10487603	1.05E-02	NPY: AD (Bhaskar et al. 2013; Mottagui-Tabar et al. 2005); alcohol withdrawal seizures (Okubo, Harada, Higuchi, & Matsushita, 2003)
7	CHRM2	34	10000	83.48	3.34E-02	rs6943675	3.77E-03	AD/drug dependence/affective disorders (Luo et al. 2005), and AD with depression (Wang et al. 2004).
11	BDNF	14	10000	46.06	9.70E-03	rs4923460	2.42E-03	BP (Serretti and Mandelli 2008; Craddock and Sklar 2009).
1	TBX19	12	10000	27.38	4.50E-02	rs2075976	1.78E-02	AD (Yu et al. 2008)
7	GRM8	101	10000	243.76	1.31E-02	rs2299555	3.38E-03	Theta EEG and alcoholism (Chen, 2009); SZ (Takaki, 2004; Zhang, 2014; Li 2015), depression (Lee, 2012)
4	NPY2R	10	100000	51.53	1.45E-03	rs2342653	1.41E-03	AD, alcohol withdrawal symptoms, alcohol and cocaine dependence (Wetherill et al. 2008)
2	NTSR2	15	10000	30.01	4.85E-02	rs6432233	2.08E-02	AD with conduct disorder and suicide attempts (Dick et al, 2010)
3	GPD1L	18	100000	73.02	5.90E-03	rs1562878	1.07E-03	AD (suggestive) (Frank, 2012)
4	TLR1	9	10000	22.84	3.25E-02	rs7696175	2.15E-02	AD, maxdrinks (Kapoor et al. 2013)

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations. BP = bipolar disorder. SZ = schizophrenia. AD = alcohol dependence

Fig. 14-3 SZALC vs SNC: uncharacterised gene LOC101927533 region plot

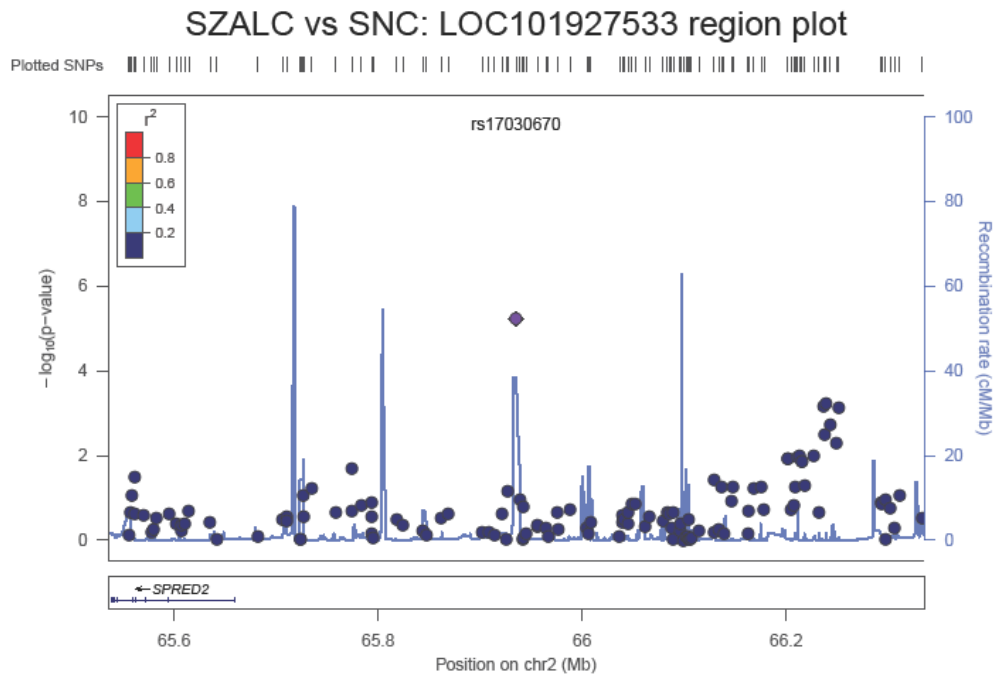


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 14-4 SZALC vs SNC: uncharacterised gene LOC101927166 region plot

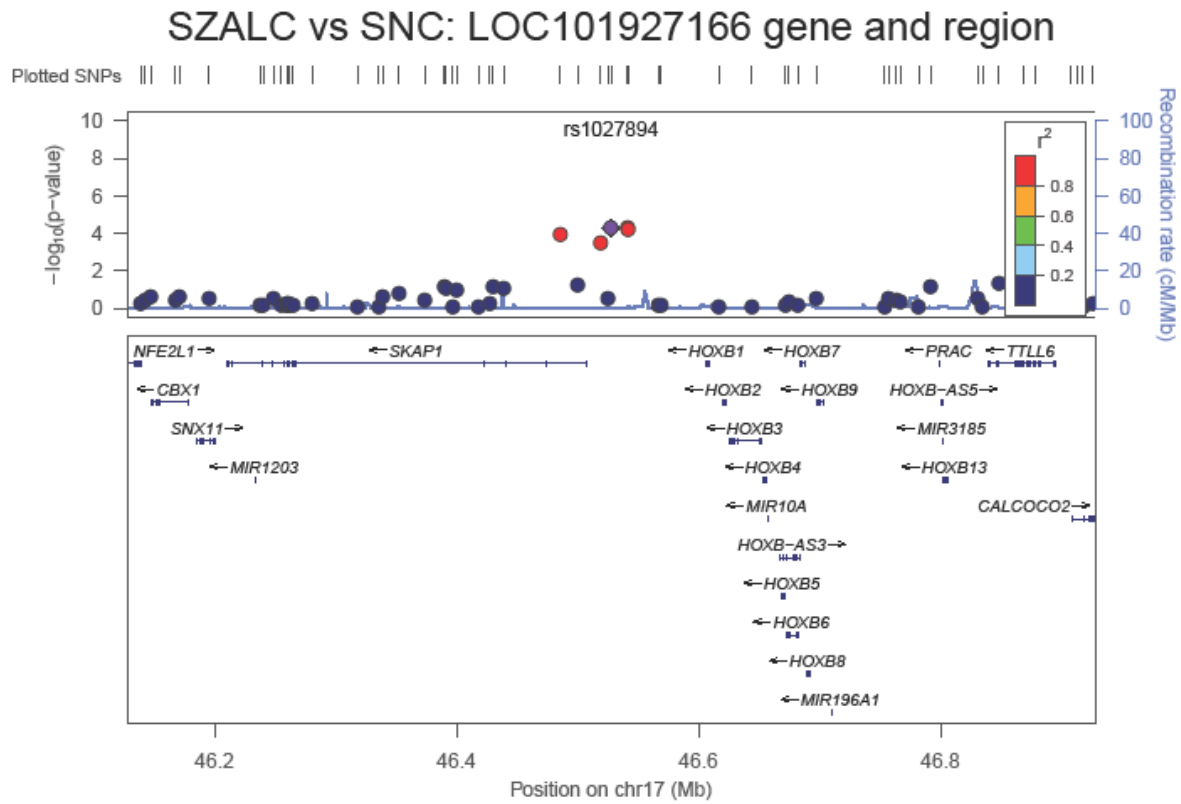


Figure legend: 1 Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position. Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

CHAPTER 15: RESULTS OF META-ANALYSIS: BPALC + SZALC vs BPNONALC + SZNONALC

For this meta-analysis we examined 220 BPALC + SZALC vs “pure” BP + SZ (ie without AD) as controls. This approach was to increase the power to detect genes which affect risk of developing AD in people with severe mental illness. We used PLINK’s meta-analysis function.

We analysed 372,193 SNPs genotyped on the Affymetrix GeneChip Human Mapping Array following the quality control and data cleaning procedure described in Methods, above. The overall genotyping call rate for analysed SNPs was 99.4%. Including the multimarker predictors, our data set is estimated to capture 78.7% of common variation (SNPs with minor allele frequency $\geq 5\%$) in the CEPH HapMap (CEU) samples with $r^2 > 0.8$.

While the larger sample size increased the significance values by a factor of 10, no SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p -values for the autosomal associations in relation to the expected p -values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 15-1. A Manhattan plot of our association analyses is shown in Fig. 15-2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p -values are presented in Table 15.1.

I then checked whether any of the 40 SNPs with the lowest p -value for each analysis were in or close to a gene of known relevant function. None of the top SNPs were previously related directly to AD or alcoholism-related phenotypes, however genes in or near these SNPs are briefly reported.

Our strongest SNP was intronic to *TLN2* (talin 2, Fig. 15-3) which gene encodes a protein related to talin 1, a cytoskeletal protein that plays a significant role in the assembly of actin filaments and in spreading and migration of various cell types, including fibroblasts and osteoclasts. Next was Senataxin (SETX) a gene encodes for a protein which contains a DNA/RNA helicase domain at its C-terminal end which suggests that it may be involved in both DNA and RNA processing. Mutations in this gene have been associated with ataxia-ocular apraxia-2 (AOA2) and an autosomal dominant form of juvenile amyotrophic lateral sclerosis (ALS4). T-Cell Immunoglobulin And Mucin Domain Containing 4 (TIMD4) was next by SNP implication. This is implicated in sarcoma and malaria. Fifth ranked SNP was neuron navigator 2 (NAV2), a gene associated with neuroblastoma. GO annotations related to this gene include *heparin binding* and *helicase activity*. The seventh strongest SNP was glutamate decarboxylase-like protein 1 (GADL1) a variant of which is linked to response to lithium therapy in bipolar disorder (Chen et al. 2014). SNPs in GADL1 were highly representative in the BPALC vs BPnonALC analysis, too.

In the gene-based tests, we found 238 genes with $p < 0.01$, strongest $p = 3.2 \times 10^{-5}$. The top 40 of these are listed in Table 15.2. Cross referencing with the literature AD and related phenotypes database I found 22 genes that were nominally associated ($p \leq 0.05$) that were in common with our VEGAS2 gene wise list. No genes survived Bonferroni correction for multiple tests.

The most notable suggestive gene based results were GABRB3 which is associated with severity of alcoholism (Noble et al. 1998); MTHFR is associated with milder (Babor type A) alcohol dependence (Benyamina et al. 2009); AD (Saffroy et al. 2008); and serum homocysteine levels in alcoholics (de la Vega et al. 2001). GPRIN1's allele A-containing genotype is over-represented in alcohol withdrawal seizures (Rujescu et al. 2005). Also interesting are OSBPL5 associated with AD (Edenberg, 2010) and PKNOX2 which has been repeatedly associated with AD in numerous studies AD (Chen et al. 2009);

Wang et al. 2011; Zuo et al.; Mulligan et al. 2006), and with substance dependence (Chen et al. 2011). TTL, HAAO, COBLL1 are all associated with AD with conduct disorder, depression and suicide (Dick et al. 2010).

An important negative finding is the absence of taste receptors, alcohol metabolising genes, and AUTS2 (autism susceptibility gene 2), recently associated with AD (Schumann et al. 2011). A regional plot of our AUTS2 data confirms little signal in this region (Fig. 15-4).

Pathway analysis using VEGAS revealed pathways involving generation of neurons (GO:0048699), inflammation (GO:0050727), response to inorganic substance (GO:0010035), transportation (GO:00151-03), and tyrosine metabolism (GO:004251) (Table 15.4).

In summary, this was the first meta-analysis of BPALC+SZALC vs “pure” BP+SZ (i.e. severe mental illness [SEMI] without AD) in an attempt to tease out the AD genetic effects. No marker was genome-wide significant. However gene based, replication and pathway approaches offer a number of biologically plausible candidate genes worthy of further investigation which we now discuss.

15.1 DISCUSSION

To increase power we combined the BPALC and SZALC groups in a meta-analysis and compared them with either “pure” psychosis or normal controls. Again, genome wide significance was not achieved, probably due to insufficient sample sizes.

Combining the BPALC and SZALC samples increases sample size (for “psychosis + AD”) but there are drawbacks: if BP and SZ are indeed biologically distinct disorders (and Kraepelin was correct in predicting this in 1921 (Kraepelin, 1921)) then power to detect a difference will be diluted by phenotypic heterogeneity. If, as we suspect, there is sufficient genetic overlaps between SZ and BP, this dilution may be reduced. However we have little data on any overlaps between SZALC and BPALC as phenotypes.

A number of genes tentatively implicated by gene based analysis may offer some clues as to pathways differentiating addictions from psychotic disorders. Below I discuss the most pertinent.

The glutamate system has long been involved in SSZ, BP and addictions. GRM8 has been implicated in event-related theta oscillations during information processing and also in vulnerability to alcoholism using a family-based association test (FBAT) in the COGA sample (Chen, 2009). GRM8 was chosen by Chen et al because it is located at 7q31.3-q32.1 under the peak region of previously identified significant linkage (peak LOD = 3.5) using a genome-wide linkage scan of the same phenotype (event-related theta band for the target visual stimuli).

Ghrelin, an orexigenic peptide, acts on growth hormone secretagogue receptors (GHS-R1A), and expressed in the hypothalamus as well as in important reward nodes such as the ventral tegmental area. In a haplotype analysis of Spanish individuals ranging from abstinent to heavy drinkers, SNP rs2232165 of the GHS-R1A gene was associated with heavy alcohol consumption and SNP rs2948694 of the same gene as well as haplotypes of both the pro-ghrelin and the GHS-R1A genes were associated with body mass in heavy alcohol consuming individuals (Landgren, 2008).

Glycerol-3-phosphate dehydrogenase 1-like (GPD1L) gene has been suggested in a GWAS via a SNP located at a distance of 30kb (Frank, 2012). In humans, mutations in the GPD1L gene have been reported to cause Brugada syndrome 2, which is a progressive disorder of myocardial conduction (OMIM MIM ID #611777). Interestingly, this gene has received support for the phenotype “alcohol stimulated activity” in the mouse (Palmer et al. 2006).

In the next chapter I report on the final meta-analysis, using comorbid BPALC+SZALC vs normal controls analysis.

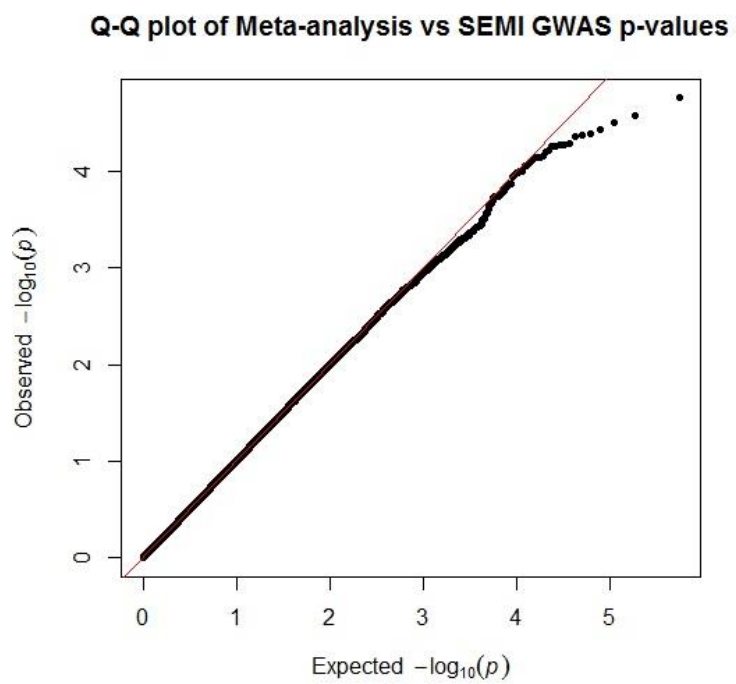


Figure legend: $Q-Q$ plot of observed and expected $-\log_{10}(P)$ of the associations between SNPs

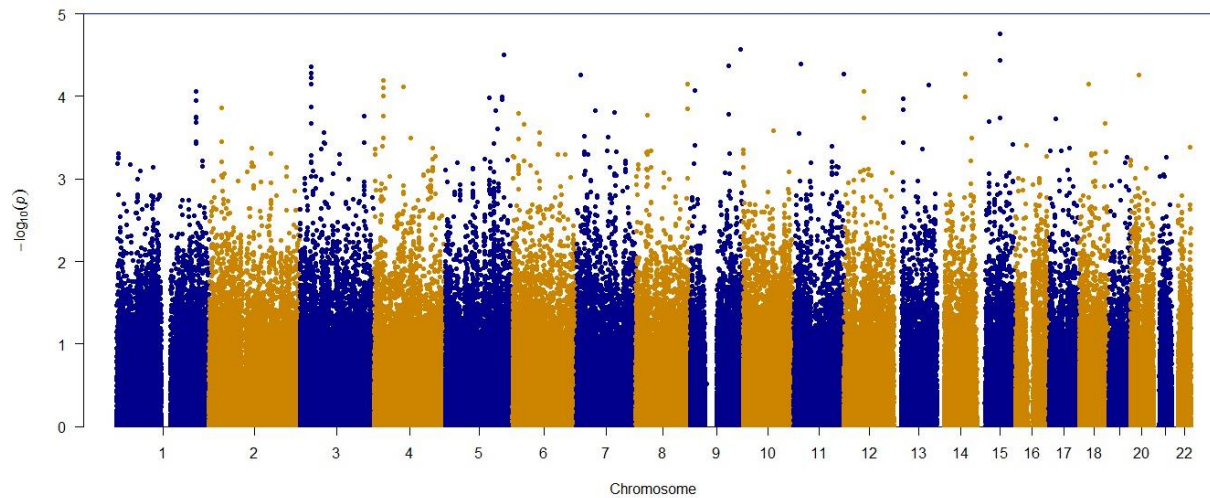


Figure legend: Results of the genome-wide association analyses for Meta-analysis vs BP+SZ. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Table 15.1 Meta vs SEMI Top 40 SNPs and gene based results

CHR	SNP	Position (base pairs)	A1	A2	Q	P	OR	Position	Distance	Nearest gene code	Gene based p*
15	rs2456470	62777956	G	C	1.7037	1.74E-05	1.74E-05	intron	0	TLN2	2.19E-01
9	rs10901156	135184237	G	A	2.1542	2.67E-05	0.002131	intron	0	SETX	3.21E-03
5	rs9790967	156434655	G	T	1.5905	3.13E-05	3.13E-05	upstream	44389	TIMD4	1.85E-03
15	rs8039781	62836025	G	A	1.6775	3.64E-05	3.64E-05	intron	0	TLN2	2.19E-01
11	rs10766606	19960636	C	T	0.63	4.08E-05	4.08E-05	intron	0	NAV2	4.43E-02
9	rs16918616	102272438	A	G	3.0302	4.26E-05	0.00499	upstream	21540	LOC100507438	---
3	rs1395162	31092371	C	A	0.6336	4.41E-05	0.001856	upstream	156114	GADL1	2.21E-01
3	rs294306	31079692	T	A	0.641	5.20E-05	0.00947	upstream	143435	GADL1	2.21E-01
11	rs7929735	133864938	T	C	1.6573	5.32E-05	0.001251	upstream	38058	IGSF9B	1.79E-01
14	rs10147012	76019185	T	C	1.8391	5.38E-05	0.01495	downstream	5827	BATF	2.15E-04
20	rs6048438	22901657	A	G	0.5564	5.51E-05	5.51E-05	upstream	34951	---	---
7	rs12699238	11610592	C	G	1.6644	5.55E-05	5.55E-05	intron	0	THSD7A	3.03E-02
3	rs1683760	31089530	A	G	0.6446	5.97E-05	0.01305	upstream	153273	GADL1	2.21E-01
4	rs7661189	25250165	T	C	0.633	6.33E-05	6.33E-05	intron	0	PI4K2B	6.07E-04
8	rs12674962	139708408	C	T	0.6357	7.03E-05	7.03E-05	intron	0	COL22A1	2.48E-02
18	rs524273	24390432	C	A	0.5976	7.09E-05	0.05368	intron	0	AQP4-AS1	3.65E-02
3	rs182401	31081125	T	G	0.6467	7.09E-05	0.01226	upstream	144868	GADL1	2.21E-01
13	rs4324007	93242263	A	T	0.589	7.18E-05	0.001001	intron	0	GPC5	9.99E-02
4	rs7688071	79126653	T	C	1.5816	7.75E-05	7.75E-05	intron	0	FRAS1	3.98E-02
4	rs7684466	25386468	C	T	0.6389	7.81E-05	7.81E-05	intron	0	ANAPC4	1.47E-03
9	rs10491739	12847826	C	T	1.6824	8.49E-05	8.49E-05	downstream	25696	LURAP1L	1.34E-02
1	rs1078308	213705895	A	C	1.6009	8.60E-05	8.60E-05	intron	0	---	---
12	rs1464042	53091326	T	C	1.6015	8.68E-05	0.02597	intron	0	KRT77	2.50E-03
4	rs10027814	25386559	G	A	0.6428	9.94E-05	9.94E-05	intron	0	ANAPC4	1.47E-03
4	rs7654358	25227917	G	A	0.6308	9.94E-05	9.94E-05	intron	0	PI4K2B	6.07E-04
14	rs957370	76008487	C	T	1.8938	1.00E-04	4.32E-04	intron	0	BATF	2.15E-04

Table 15.1 contd...

5	rs304855	152759013	G	A	0.6436	1.02E-04	0.02738	upstream	110162	GRIA1	7.38E-01
5	rs13154208	119318747	A	G	1.6438	1.03E-04	0.01299	downstream	301972	---	---
13	rs4770587	24692835	C	T	1.5435	1.07E-04	1.07E-04	intron	0	---	---
5	rs304841	152778918	C	T	0.6283	1.10E-04	0.007309	upstream	90257	GRIA1	7.38E-01
1	rs6680294	213715274	G	A	1.5906	1.13E-04	1.13E-04	intron	0	---	---
3	rs294278	31127911	T	C	0.6617	1.35E-04	0.02984	upstream	191654	GADL1	2.21E-01
2	rs743236	33346798	T	A	1.5298	1.37E-04	1.37E-04	intron	0	LTBP1	3.07E-01
8	rs10104531	139702693	T	C	0.6349	1.40E-04	1.40E-04	intron	0	COL22A1	2.48E-02
13	rs2146687	24700616	T	C	1.5301	1.43E-04	1.43E-04	intron	0	---	---
5	rs1148364	134924962	G	C	1.5268	1.48E-04	1.48E-04	intron	0	---	---
7	rs17554438	51180346	T	C	1.8833	1.49E-04	0.001809	intron	0	COBL	1.04E-02
7	rs2950183	102020363	G	T	1.5298	1.58E-04	0.03439	intron	0	PRKRIP1	1.05E-02
6	rs1267719	16060085	G	A	1.5197	1.61E-04	1.61E-04	upstream	47768	---	---
9	rs2900171	102251637	A	T	2.7206	1.64E-04	0.04234	upstream	42341	LOC100507438	---

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F_A= Frequency of Allele 1; F_U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. *Uncorrected p <0.05 in **bold**.

Table 15.2 Top 40 Genes Yielding Evidence for Association After VEGAS2 Permutation Testing

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Uncorrected gene based p	Top SNP	Top SNP P
3	GHSR	9	1.00E+06	74.86	3.10E-05	rs509035	1.74E-04
8	PPP2CB	5	1.00E+06	45.33	5.50E-05	rs2111862	4.69E-04
14	BATF	10	1.00E+06	53.66	2.15E-04	rs10147012	5.38E-05
2	CHCHD5	8	1.00E+06	44.99	3.67E-04	rs4849073	4.17E-04
16	MT2A	14	1.00E+06	55.92	3.80E-04	rs6499850	2.09E-03
4	ZCCHC4	12	1.00E+06	89.99	4.88E-04	rs7684466	7.81E-05
16	MT1E	13	1.00E+06	53.06	5.09E-04	rs6499850	2.09E-03
14	LOC102724153	12	1.00E+06	54.60	5.10E-04	rs10147012	5.38E-05
6	SMAP1	14	1.00E+06	89.54	5.20E-04	rs781977	2.73E-04
4	PI4K2B	13	1.00E+06	96.78	6.07E-04	rs7661189	6.33E-05
16	MT1JP	11	1.00E+06	48.80	6.45E-04	rs6499850	2.09E-03
16	MT1M	11	1.00E+06	48.80	6.91E-04	rs6499850	2.09E-03
2	POLR1B	9	1.00E+06	45.55	8.00E-04	rs4849073	4.17E-04
16	MT1L	13	1.00E+06	49.61	8.01E-04	rs6499850	2.09E-03
16	MT1DP	12	1.00E+06	48.81	8.65E-04	rs6499850	2.09E-03
16	MT1A	12	1.00E+06	48.81	9.38E-04	rs6499850	2.09E-03
13	LINC00346	5	1.00E+05	30.90	9.60E-04	rs9521967	3.15E-03
4	DOK7	14	1.00E+06	66.38	9.91E-04	rs13128489	4.29E-04
4	LRPAP1	14	1.00E+05	65.80	1.04E-03	rs13128489	4.29E-04
19	BRD4	5	1.00E+05	28.14	1.07E-03	rs2079014	2.79E-03
3	MFSD1	9	1.00E+05	48.01	1.08E-03	rs340271	1.87E-03
12	KRT2	25	1.00E+05	117.30	1.16E-03	rs1464042	8.68E-05
4	HGFAC	9	1.00E+05	41.61	1.29E-03	rs13128489	4.29E-04
12	MIR4303	7	1.00E+05	37.84	1.45E-03	rs11109314	2.34E-03

Table 15.2 contd...

4	ANAPC4	11	1.00E+05	67.60	1.47E-03	rs7684466	7.81E-05
7	SPDYE6	4	1.00E+05	26.73	1.50E-03	rs2950183	1.58E-04
7	LOC100289561	5	1.00E+05	27.91	1.53E-03	rs2950183	1.58E-04
16	MT1B	12	1.00E+05	46.78	1.54E-03	rs6499850	2.09E-03
4	FABP2	14	1.00E+05	86.40	1.64E-03	rs11722940	4.18E-03
10	LOC102031319	13	1.00E+05	57.90	1.71E-03	rs2505349	2.08E-03
2	COBLL1	33	1.00E+05	122.52	1.73E-03	rs355839	1.11E-03
13	SPATA13_1	93	1.00E+05	202.08	1.77E-03	rs4770587	1.07E-04
3	TNFSF10	19	1.00E+05	82.31	1.80E-03	rs9881097	3.66E-04
5	TIMD4	13	1.00E+05	65.56	1.85E-03	rs9790967	3.13E-05
6	C6orf203	9	1.00E+05	33.80	1.98E-03	rs1714632	2.15E-03
2	DBI	11	1.00E+05	64.88	2.09E-03	rs1374313	7.06E-04
8	TEX15	8	1.00E+05	49.57	2.09E-03	rs2111862	4.69E-04
20	PAX1	3	1.00E+05	28.43	2.16E-03	rs6035924	1.31E-03
10	EPC1	17	1.00E+05	72.32	2.20E-03	rs2505349	2.08E-03
1	C1orf122	2	1.00E+05	12.94	2.22E-03	rs12563037	6.74E-04
1	MTF1	2	1.00E+05	12.94	2.22E-03	rs12563037	6.74E-04

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations

Table 15.3 Meta-analysis vs BP+SZ Genes Yielding Evidence for Association at P <0.05 After Permutation Testing and replication

Chr	Gene symbol	n SNPs	n Sims	VEGAS Test result	Uncorrected P	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
15	GABRB3	29	10000	61.61864	1.55E-02	rs2114485	2.88E-03	Severity of alcoholism (Noble, 1998)
1	MTHFR	17	100000	69.76328	4.29E-03	rs2075539	1.30E-03	Serum homocysteine levels in alcoholics (de La Vega 2001)
5	GPRIN1	6	10000	19.46595	1.40E-02	rs10476198	3.17E-03	2108A allele/A-containing genotype over-represented in alcohol withdrawal seizures (Rejescu 2005)
7	GRM8	100	100000	262.9882	7.49E-03	rs1211384	4.03E-03	Theta EEG and alcoholism (Chen, 2009); SZ (Takaki, 2004; Zhang, 2014; Li 2015), depression (Lee, 2012)
3	GHSR	9	10000	32.70371	9.00E-03	rs9881097	3.48E-03	Alcohol consumption and body weight (Landgren 2008)
11	OSBPL5	9	10000	21.76404	2.93E-02	rs11025773	1.13E-03	AD (Edenberg, 2010)
11	PKNOX2	54	10000	104.4015	2.24E-02	rs7104304	4.76E-04	AD (Chen, 2009; Beirut, 2010; Zuo, 2010,Guo, 2012, Zuo, 2014), substance dependence (Chen 2011)
1	ST6GALNAC3	98	10000	168.6442	4.45E-02	rs12096789	5.76E-03	Nominal association theta oscillations (Hodgkinson, 2010)
2	TTL	7	100000	34.61967	4.01E-03	rs4849073	3.74E-04	AD with conduct disorder, depression and suicide (P=0.0302) (Wang, 2007)
2	HAAO	17	100000	79.99661	4.99E-03	rs7582883	1.86E-03	AD with conduct disorder, depression and suicide (P=0.0302) (Wang, 2007)
2	COBLL1	33	10000	88.48924	1.26E-02	rs355839	3.38E-03	AD with conduct disorder, depression and suicide (P=0.0302) (Wang, 2007)
14	DDHD1	17	10000	46.72375	3.38E-02	rs12435772	4.19E-03	AD (meta-analysis) (Wang, 2010)
11	PKNOX2	54	10000	104.4015	2.24E-02	rs7104304	4.76E-04	AD (Chen, 2009; Beirut, 2010; Zuo, 2010,Guo, 2012, Zuo, 2014), substance dependence (Chen 2011)
3	GPD1L	18	10000	57.55162	1.48E-02	rs1562878	7.76E-03	AD (suggestive) (Frank, 2012)

Table legend: SNP = single nucleotide polymorphism. N VEGAS2 Sims = number of VEGAS2 gene based simulations. BP = bipolar disorder. SZ = schizophrenia. AD = alcohol dependence.

Table 15.4 Meta-analysis vs normal controls: VEGAS2 pathway analysis

GO ID and pathway	Initial pathway length	Final pathway length	Pathway p	Empirical p	Pathway gene list
PANTHER Other transport	53	44	7.53E-05	2.20E-05	SLC35D1, HIAT1, AQP10, ERO1LB, VPS54, MFSD9, SLC5A7, SLC15A2, SEC61A1, MFSD1, LRPAP1, TRIM2, SLC35D3, AQP1, AQP3, HIATL1, SLC35D2, LCN9, SEC61A2, SLC22A18, TRIM3, CAPRIN1, SLC35C1, SLC15A3, CAPRIN2, AQP6, MIP, SLC35E3, NUP37, SLC15A4, SLC15A1, TMCO3, ERO1L, NGB, OCA2, AQP9, AQP8, SPNS3, TOM1L2, AQP4, SLC35E1, SYS1, SLC35C2, SLC35E4
BIOCARTA ARF PATHWAY	17	17	1.11E-04	3.20E-05	POLR1A, POLR1B, PIK3CA, PIK3R1, POLR1C, RAC1, TWIST1, PIK3CG, MYC, CDKN2A, ABL1, MDM2, POLR1D, RB1, TP53, TBX2, E2F1
GO:0050727 regulation inflammation response	71	63	1.00E-03	7.00E-04	TNFRSF1B, PLA2G2A, FCER1A, SELE, SERPINC1, PLA2G4A, ADORA1, IL20, AGT, CALCRL, GHRL, PPARG, TLR9, IL20RB, AGTR1, GHSR, ADIPOQ, BCL6, IL2, EDNRA, TLR2, NPY5R, KLKB1, OSMR, ITGA2, IL6ST, NDFIP1, ADRB2, F12, NT5E, IL6, TAC1, IDO1, FABP4, JAK2, TLR4, IL2RA, INS, SAA1, SERPING1, SCGB1A1, TNFRSF1A, A2M, CMA1, GPX2, UACA, CD276, SPN, CX3CL1, SERPINF1, ADORA2B, CCL5, STAT5A, PRKCA, SBNO2, C3, ZFP36, APOE, NLRP12, LBP, ADA, ADORA2A, OSM
GO:0010035 response to inorganic substance	138	127	1.06E-03	7.60E-04	PEF1, MTF1, TSHB, TXNIP, S100A16, PLA2G4A, TNNT2, ACTA1, RYR2, SDC1, APOB, KHK, XDH, HAAO, IL1A, LCT, NR4A2, SLC25A12, TTN, IGFBP2, SLC6A1, CTNNB1, TNNC1, CPOX, CASR, TRPC1, KCNMB3, TFRC, HTT, SOD3, PDGFRA, SEC31A, SNCA, LEF1, CASP6, FGG, PDCD6, SLC6A3, SLC1A3, SLC30A5, MAP1B, PCSK1, KCNIP1, DUSP1, SLC34A1, MAPK9, ITPR3, CDKN1A, GCLC, PGAM2, EGFR, SLC25A13, SRRT, CAV1, NOS3, PTK2B, STAR, ADAM9, GGH, IMPA1, CA2, TNFRSF11B, NDRG1, ACO1, AQP3, FXN, ALG2, ALDOB, GSN, RXRA, MAPK8, KCNMA1, ANXA11, PTEN, KCNIP2, AS3MT, UROS, STIM1, RELA, MTL5, CCND1, APOBEC1, A2M, MGP, SLC11A2, AVPR1A, KCNMB4, SYT1, HVCN1, PEBP1, P2RX7, ATP7B, MNAT1, SERPINA1, THBS1, CAPN3, SLC30A4, NEDD4, AQP9, CYP1A1, CHRNA3
GO:0042517 reg. tyrosine phosphorylation	11	10	1.06E-03	6.40E-04	IL6R, IL20, GHR, IL6ST, IL6, JAK2, CNTF, CLCF1, GH1, LIF
GO:0016605 PML body	29	29	1.23E-03	6.60E-04	SKI, HIPK1, SUMO1, SP100, TOPBP1, TDP2, LRCH4, HIPK2, EIF3E, SIRT1, TRIM8, ZBTB16, CBX5, CHFR, ZMYM2, RNF6, RB1, RAD51, PML, ISG20, BLM, UBE2I, RPA1, RPAIN, TP53, TOP3A, SPTBN4, MORC3, CHEK2

Table 15.3 contd

GO:0008593 regulation of Notch signaling pathway	10	10	1.53E- 03	7.80E- 04	SOX2, TP63, KIT, IL6ST, HEY2, ASCL1, DTX1, NFKBIA, DLL3, JAG1
GO:0048699 generation of neurons	455	382	1.76E- 03	1.00E- 03	HES5, TP73, HES3, EPHA2, PAX7, EPHB2, ID3, PPT1, ARTN, PCSK9, DAB1, DOCK7, RPE65, LHX8, USP33, MCOLN3, BARHL2, GF11, LPPR4, OLFM3, NTNG1, AMIGO1, NGF, S100A6, EFNA1, NTRK1, IGSF9, LMX1A, TNFR, LHX4, CRB1, BTG2, CNTN2, TGFB2, WNT3A, TBCE, GDF7, DPYSL5, STRN, NRXN1, RTN4, PEX13, B3GNT2, VAX2, ALMS1, SEMA4F, CTNNA2, TTL, ACTR3, EN1, GLI2, KIF5C, NR4A2, TBR1, DLX2, HOXD3, NEUROD1, SATB2, ALS2, ABI2, NRP2, CREB1, MAP2, XRCC5, EPHA4, DNER, GBX2, FARP2, CNTN4, ATP2B2, PPARG, WNT7A, TOP2B, CCR4, CCK, LAMB2, SEMA3B, FEZF2, ROBO2, ROBO1, RPL24, ALCAM, DRD3, KALRN, GATA2, EPHB1, SLITRK3, PRKCI, NLGN1, SOX2, EPHB3, DGKG, HES1, SPON2, HTT, SLIT2, CCKAR, PHOX2B, KIT, REST, RUFY3, NKX61, SPP1, ATOH1, UNC5C, NEUROG2, SCLT1, HHIP, SMAD1, POU4F2, VEGFC, HELT, SEMA5A, SLC1A3, ISL1, MAP1B, OTP, VCAN, GPR98, EFNA5, APC, SEMA6A, ACSL6, 4, DRD2, DSCAML1, PVRL1, FEZ1, OPCML, TULP3, CCND2, NTF3, PEX5, SOX5, WNT1
BIOCARTA HER2 PATHWAY	20	19	1.82E- 03	9.20E- 04	SHC1, SOS1, ERBB4, RAF1, PIK3CA, IL6ST, PIK3R1, ESR1, IL6, EGFR, PIK3CG, HRAS, ERBB3, GRIP1, MAP2K1, STAT3, GRB2, MAPK1, EP300
REACTOME KINESINS	21	20	1.97E- 03	1.08E- 03	KIF2C, KIFAP3, KIF3C, KIF15, KIF9, KIF2A, KIF20A, KIF4B, KLC4, KIF5B, KIF11, KIF18A, KLC2, RACGAP1, KIF5A, KIF26A, KIF23, KIF2B, KLC3, KIF3B
GO:00151- 03: inorganic anion trans- membrane transporter activity	31	27	2.91E- 03	1.58E- 03	SLC26A9, PCYOX1, SLC4A5, SLC20A1, SLC4A7, SLC26A1, SLC34A2, SLC4A4, ANKH, SLC26A2, SLC34A1, SLC17A2, SLC26A8, SLC26A5, SLC26A3, SLC13A1, SLC13A4, SLC20A2, SLC26A7, ADAMTS8, AQP6, SLC25A3, OCA2, SLC26A11, SLC5A5, SLC17A7, SLC4A11

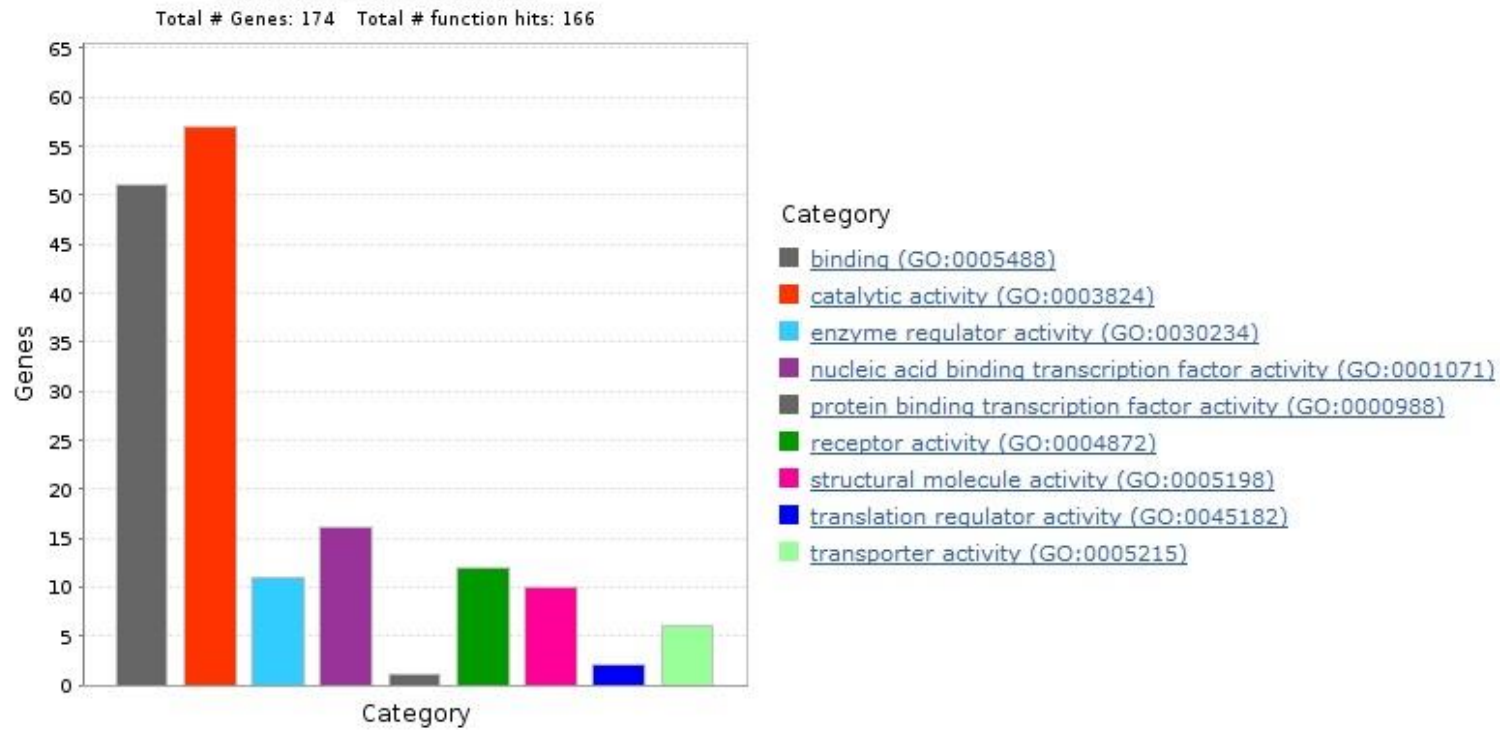


Figure legend: x axis = number of genes; x axis = GO molecular category

CHAPTER 16: RESULTS OF META-ANALYSIS (COMORBIDITY VS SUPERNORMAL CONTROLS): BPALC+SZALC vs SNC

This meta-analysis was used to try and increase the power to differentiate SZALC+BPALC genes from normal controls (Panagiotou et al. 2013). Unlike the previous chapter which sought AD genes as separate from BP, this approach looks at the combined severe mental illness phenotype. We used the PLINK meta-analysis function. To reduce bias we used half the normal controls in each analysis combined in the meta-analysis.

We meta-analysed 220 BPALC+SCZALC vs ancestrally matched normal controls. Again, despite the increased power, no SNP achieved conventional genome wide significance levels. Quantile-Quantile (Q-Q) plots for each analysis, illustrating the observed p-values for the autosomal associations in relation to the expected p-values (based on the number of tests, under the null hypothesis of no association), are presented in Figure 16.1. A Manhattan plot of our association analyses is shown in Fig.16.2. The lack of SNP data points near the top of each panel (i.e. $p \approx 10^{-8}$) indicates that there were no strong association signals. SNPs in the top 40 with lowest p-values are presented in Table 16.1.

I then checked whether any of the 40 SNPs with the lowest p-value for each analysis were in or near to a gene of known relevant function. None of the top SNPs themselves were previously related directly to AD or alcoholism-related phenotypes, however genes in or near these SNPs are briefly reported.

Our strongest SNP was rs743236 ($p = 4.14 \times 10^{-6}$; OR 1.72; gene based $P = 5.53 \times 10^{-2}$) which is in the LTBP1 (latent transforming growth factor beta binding protein 1, Fig. 16-3) gene. I have also plotted the same region from the PGC CDG which appears to show little evidence of association in comparison (Fig. 16-4).

Other interesting genes by nominal SNP association are in: Poly (ADP-Ribose) Polymerase Family, Member 16 (PARP16, catalyses the post-translational modification of proteins; rs1561889, $p = 4.17 \times 10^{-5}$, OR 1.7); RAD51 Paralog B (RAD51B, which may detect unstable DNA; rs11624333, $p = 8.42 \times 10^{-5}$, OR 1.7), Zinc Finger 33A (ZNF33A, functions include sequence-specific DNA binding transcription factor activity; rs645064, $p = 9.2 \times 10^{-5}$, OR 1.59), and LIM domain 7 (LMO7, which may be involved in protein-protein interactions and is implicated in Townes-Brocks syndrome; rs9544021, $p = 2.45 \times 10^{-5}$, OR 1.64). Gene Senataxin (SETX) was implicated by an intronic SNP rs10901156 ($p = 8.89 \times 10^{-5}$, OR 2.13) and gene based analysis ($p = 2.41 \times 10^{-3}$). CDH11 (cadherin 11, type 2, OB-cadherin (osteoblast) Fig. 16-5) which has numerous addiction signals in this study and elsewhere was ranked 24th by SNP p-value, and was also nominally significant by gene based analysis ($p = 1.44 \times 10^{-3}$). I plotted the same region using the PGC CDG GWAS data (Fig. 16-6) which suggests no evidence of association at this locus. CDH11 and other cadherins are discussed further in the Discussion section.

In order to confirm the involvement of genes previously implicated in alcoholism and addictions phenotypes, I employed gene based permutation tests on all SNPs within a 50 kb flank of all known genes.

In the gene-based tests, we found 202 genes with $p < 0.01$, strongest gene PDGFC on Chromosome 4, uncorrected $p = 2.7 \times 10^{-4}$ (Fig. 16-7). Cross referencing with the literature database of gene and SNP associations, I found 25 genes that were nominally associated ($p < 0.05$) and in common with our VEGAS2 gene based analysis list. The top 40 of these are listed in Table 16.2. No genes survived Bonferroni correction for multiple tests.

The most notable gene based results were strikingly similar to the previous meta-analysis. All gene replications of AD or related phenotypes and the top SNPs in that gene where $p < 0.05$ from our data are listed in Table 16.3.

PKNOX2 has been repeatedly associated with AD in numerous studies (Chen, 2009; Beirut, 2010; Zuo, 2010; Guo, 2012; Zuo, 2014), substance dependence (Chen 2011). UCL meta-analysis SNPs in PKNOX2 are plotted in Fig. 16-8 showing some SNP signal for possible association. The PGC CDG plot of the same region is shown in Fig. 16-9, which appears to show no association at this locus.

GABRB3 (GABA(A) receptor beta3 subunit, gene based $p = 1.5 \times 10^{-2}$) is of interest due to involvement of the GABA system in AD (also see Section 5.3.1) and a previous report of association with severity of alcoholism but not AD (Noble et al. 1998).

Of particular interest is KCNMA1 (Potassium Channel, Calcium Activated Large Conductance Subfamily M Alpha, Member 1) which has been associated with AD in SZ (Kendler, 2011). KCNMA1 codes for “MaxiK channels” which are voltage- and calcium-sensitive potassium channels fundamental to the control of smooth muscle tone and neuronal excitability. Diseases associated with KCNMA1 include generalized epilepsy, paroxysmal dyskinesia and impotence. Among its related pathways are Platelet homeostasis and Haemostasis. GO annotations related to this gene include actin binding and calcium-activated potassium channel activity.

OSBPL5 (Edenberg, 2010), Cancer Susceptibility Candidate 4 (CASC4) (Wang, 2010), and Glycerol-3-Phosphate Dehydrogenase 1-Like (GPD1L) (Frank, 2012) have all been implicated in AD.

TTL, HAAO, and COBLL1 are all associated with AD with conduct disorder, depression and suicide (Dick, 2010)

Further uncorrected gene based p-values <0.05 included GRINP1 (G Protein Regulated Inducer of Neurite Outgrowth 1) which has been linked to alcohol withdrawal seizures (Rejescu, 2005) and GHS-R1A (linked to body mass in heavy alcohol consuming individuals (Landgren, 2008).

GRM8 has been implicated in event-related theta oscillations during information processing and also in vulnerability to alcoholism using a family-based association test (FBAT) in the COGA sample (Chen, 2009). It is also a gene with strong evidence for involvement in SZ and depression but not bipolar disorder SZ (Takaki et al. 2004; Zhang et al. 2014; Li et al. 2015; Lee et al. 2012).

Pathway analysis using VEGAS2 revealed nearly identical similar genetic pathways to the previous meta-analysis. These involved generation of neurons, inflammation, reaction to inorganic substance, transportation, and tyrosine metabolism (See Table 15.4 for top ranked pathways results).

When compared with the PGC Cross Disorders Group Pathway analysis (Consortium 2015) there were few areas of overlap in the top ranked pathways. Only one PGC CDG pathway - GO:8601, Protein phosphatase type 2A regulator activity – was listed in our meta-analyses pathways with $p \leq 0.05$ (Table 15.5).

PANTHER analysis of our top genes implicated in gene based analysis ($P < 0.001$, $n=236$) revealed genes clustered in binding, catalytic, receptor and transporter activity among the main molecular pathways (Fig. 16-10).

In summary, this was the first GWAS comparing patients with BP or SZ and comorbid AD with normal controls in an attempt to tease out the severe mental illness + AD phenotype genetic effects (as opposed to AD alone). No marker reached genome-wide significance. However gene based, literature, and pathway approaches offer a number of biologically plausible genes.

16.1 DISCUSSION

This meta-analysis was of the combined the BPALC and SZALC groups compared with supernormal controls. Again, genome wide significance was not achieved, probably due to insufficient sample sizes.

Combining the BPALC and SZALC samples increases sample size (for “psychosis + AD”) but there are drawbacks: if BP and SZ are indeed biologically distinct disorders (and Kraepelin was correct in predicting this [see Section 15.1 for more on Kraepelin]) then power to detect a difference will be diluted by phenotypic heterogeneity. If, as we suspect, there is sufficient genetic overlaps between SZ and BP, then this dilution may be reversed.

A number of genes tentatively implicated by gene based analysis may offer some clues as to pathways differentiating addictions from psychotic disorders. Below I discuss the most pertinent.

Our strongest SNP was LTBP1 (latent transforming growth factor beta binding protein 1). This gene targets latent complexes of transforming growth factor beta to the extracellular matrix. Interestingly LTBP1 has been associated with alcohol drinking in a Han Chinese population by interaction with a nearby CNV (Pei, 2012). In comparison, the same region from the PGC CDG appears to show little evidence of association in comparison.

SETX (uncorrected gene based $p = 2.41 \times 10^{-3}$) is a gene which encodes for a protein which contains a DNA/RNA helicase domain at its C-terminal end, which suggests that it may be involved in both DNA and RNA processing. Mutations in this gene have been associated with ataxia-ocular apraxia-2 (AOA2) and an autosomal dominant form of juvenile amyotrophic lateral sclerosis (ALS4). However the gene has so far not been implicated in major mental illness or addiction.

CDH11 (cadherin 11, type 2, OB-cadherin (osteoblast) Fig. 16-5) which has numerous associations signals in this study and elsewhere was ranked 24th by SNP p-value, and was suggested by gene based analysis (uncorrected $p = 1.44 \times 10^{-3}$). I plotted same region using the PGC CDG GWAS data (Fig. 16-6) which suggests no regional evidence of association at this locus.

In the gene-based tests, we found 202 genes with $p < 0.01$, strongest gene PDGFC on Chromosome 24, $p = 2.7 \times 10^{-4}$ (Fig. 16-7). The most notable gene wise results were strikingly similar to the previous meta-analysis.

Cross referencing with the literature database of gene and SNP associations, I found 25 genes that were nominally associated ($p \leq 0.05$) and in common with our VEGAS2 gene wise list. No genes survived Bonferroni correction for multiple tests.

PKNOX2 has been repeatedly associated with AD in numerous studies AD (Chen et al. 2009; Zuo et al.; Wang et al. 2011; Mulligan et al. 2006; Bierut et al. 2010; Guo et al. 2012)(Chen, 2009; Beirut, 2010; Zuo, 2010; Guo, 2012; Zuo, 2014), substance dependence (Chen 2011). UCL meta-analysis SNPs in PKNOX2 are plotted in Fig. 16-8 showing some SNP signal for possible association. The PGC CDG plot of the same region is shown in Fig. 16-9, which appears to show no association at this locus.

Of particular interest is KCNMA1 (Potassium Channel, Calcium Activated Large Conductance Subfamily M Alpha, Member 1) which has been associated with AD in SZ (Kendler, 2011). KCNMA1 codes for MaxiK channels which voltage and calcium-sensitive potassium channels fundamental to the control of smooth muscle tone and neuronal excitability. Diseases associated with KCNMA1 include generalized epilepsy and paroxysmal dyskinesia and impotence. Among its related pathways are Platelet homeostasis and Hemostasis. GO annotations related to this gene include actin binding and calcium-activated potassium channel activity. Calcium channels have been implicated repeatedly in both BP and SZ aetiology and neurobiology, and their genes were initially discovered by GWAS (Ferreira et al. 2008; Hamshere et al. 2013; Cross-Disorder Group of the Psychiatric Genomics Consortium 2013; Green et al. 2013; Sklar et al. 2008). Possible involvement of other electrolytes via calcium levels is therefore plausible.

GABRB3 (GABA(A) receptor beta3 subunit, gene based $p = 1.5 \times 10^{-2}$) is of interest due to involvement of the GABA system in AD (also see Section 6.3.1) and a previous report of association with severity of alcoholism (Noble et al. 1998). Additionally, OSBPL5 (Edenberg, 2010), Cancer Susceptibility Candidate 4 (CASC4) (Wang, 2010), and Glycerol-3-Phosphate Dehydrogenase 1-Like (GPD1L) (Frank, 2012) have all been implicated in AD. TTL, HAAO, and COBLL1 are genes all previously associated with AD with conduct disorder, depression and suicide (Dick, 2010).

Further gene based p -values <0.05 included the GRINP1 (G Protein Regulated Inducer of Neurite Outgrowth 1) gene which has been linked to alcohol withdrawal seizures (Rejescu, 2005) and GHS-R1A (Growth Hormone Stimulation Receptor 1A; heavy alcohol consumption; body mass in heavy alcohol consuming individuals (Landgren, 2008).

GRM8 has been implicated in event-related theta oscillations during information processing and also in vulnerability to alcoholism using a family-based association test (FBAT) in the COGA sample (Chen, 2009). It is also a gene with strong evidence for involvement in SZ and depression but not bipolar disorder SZ (Takaki et al. 2004; Zhang et al. 2014; Li et al. 2015; Lee et al. 2012).

Important negatives for this slightly better powered GWAS of comorbid psychotic illness and AD vs supernormal controls were the lack of signal in “classic” addictions genes, including dopamine, serotonin, COMT, and alcohol metabolism.

Despite possible ancestral commonalities with the Sulovari eye colour study, we found no evidence of LD between the AD-associated GABA receptor gene cluster and eye colour genes OCA2/HERC2, nor between AD-associated GRM5 and pigmentation-associated TYR (Sulovari et al. 2015).

Explanations of these important negative findings, aside from methodological differences and population heterogeneity are: i) our cases are of comorbid disorders vs supernormal controls, whereas the literature largely consists of alcoholism/AD as cases and a mix of controls; ii) most of the alcohol metabolism gene findings have been in Asian and American populations who therefore might have different risk alleles. iii) our study was underpowered.

Pathway analysis using VEGAS revealed very similar genetic pathways to the previous meta-analysis involving generation of neurons, inflammation, reaction to inorganic substance, transportation, and tyrosine metabolism (Table 16.4). PANTHER analysis of our top genes implicated in gene based

analysis ($P < 0.001$, $n = 236$) revealed binding, catalytic, receptor and transporter activity among the main molecular pathways (Fig. 16-10) – again a very similar picture to the first meta-analysis.

This is explainable for a number of reasons: 1) the psychosis+AD and AD genetic pathways are similar, which was not evident in the underpowered single (BPALC or SZALC) GWASs. 2) Despite the larger sample size these gene wise results are suggestive at best and may be false positives. 3) The methodology introduced other unknown confounders.

In the next chapters I critically appraise the strengths and weaknesses of this study, and look to future work directions.

Figure 16-1 Q-Q plot of meta-analysis vs supernormal controls GWAS

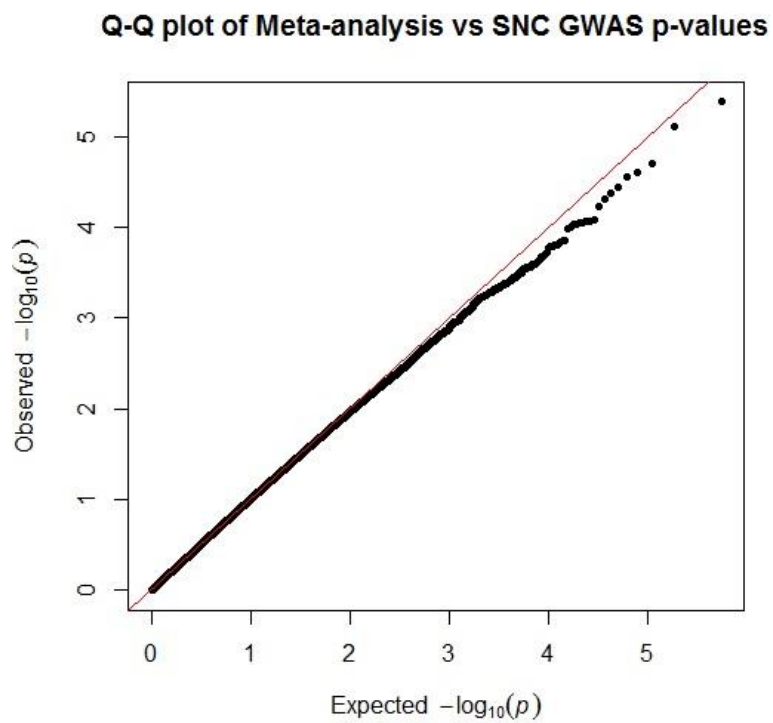


Figure legend: Q-Q plots of observed and expected $-\log_{10}(p)$ of the associations between SNPs

Figure 16-2 Manhattan plot of Meta-analysis vs supernormal controls

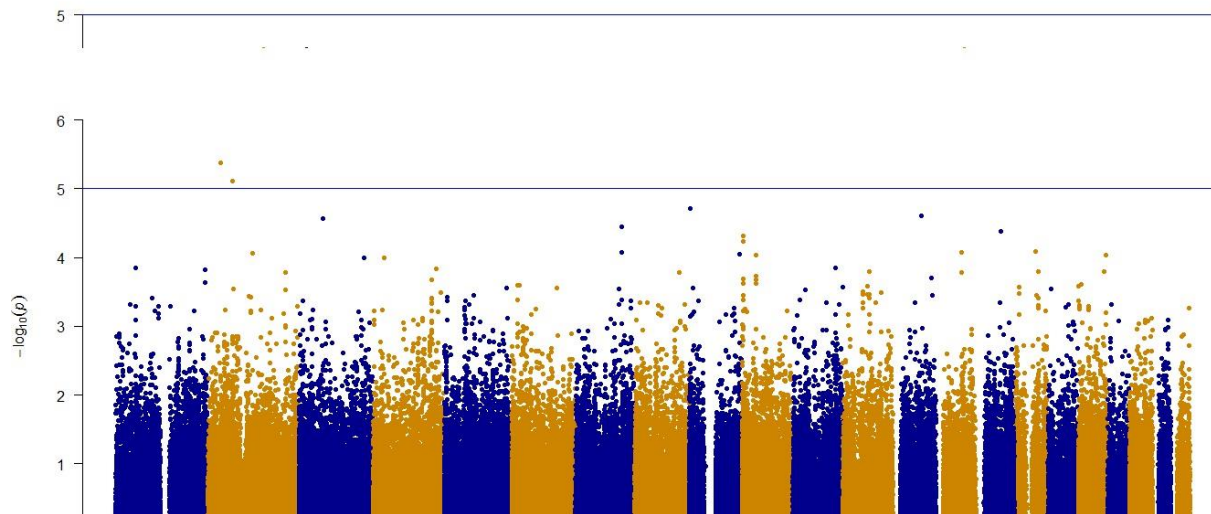


Figure legend: Results of the genome-wide association analyses for Meta-analysis vs SNC. The x -axis shows the chromosome numbers and the y -axis the significance of the association signals (i.e. $-\log_{10}(P)$ value)

Table 16.1 Meta-analysis vs supernormal controls: Top 40 SNPs and gene based results

CHR	SNP	Position (base pairs)	A1	A2	Q	P	OR	Position	Distance	Nearest gene code	Gene based p
2	rs743236	33346798	T	A	5.27E-01	4.14E-06	1.7223	intron	0	LTBP1	5.63E-02
2	rs17030670	65935157	T	C	6.00E-02	7.67E-06	2.2222	intron	0	LOC101927533	---
9	rs10967953	2756638	C	T	1.86E-01	1.96E-05	1.8603	intron	0	KIAA0020	3.17E-02
13	rs9544021	76253354	T	C	9.78E-01	2.45E-05	1.6443	intron	0	LMO7	2.06E-01
3	rs17430548	65484875	G	A	5.70E-01	2.74E-05	2.4615	intron	0	MAGI1	3.14E-01
7	rs1526408	125045427	T	A	1.12E-01	3.55E-05	2.0349	downstream	26052	LOC101928283	7.19E-02
15	rs1561889	65540826	T	C	5.04E-01	4.17E-05	1.6975	intron	0	PARP16	1.62E-02
10	rs2388268	2595331	T	C	8.19E-01	4.79E-05	1.9356	intron	0	---	---
10	rs2388263	2593161	T	G	8.51E-01	5.77E-05	1.9217	intron	0	---	---
16	rs711753	57105839	A	G	9.66E-01	8.22E-05	0.61	intron	0	NLRC5	5.90E-02
14	rs11624333	68979835	C	T	2.57E-01	8.42E-05	1.6466	intron	0	RAD51B	4.20E-02
7	rs17148220	124989897	T	G	1.25E-01	8.44E-05	1.9663	intron	0	LOC101928283	7.19E-02
2	rs10205049	119899933	G	A	1.28E-01	8.80E-05	1.585	downstream	13886	C1QL2	5.42E-03
9	rs10901156	135184237	G	A	3.92E-01	8.89E-05	2.1327	intron	0	SETX	2.41E-03
18	rs11665142	74380913	A	G	4.24E-01	9.14E-05	1.588	upstream	21073	FLJ44881	---
10	rs645064	38334341	T	C	7.15E-01	9.20E-05	1.5856	intron	0	ZNF33A	5.19E-04
4	rs260933	29373895	T	C	6.43E-01	1.00E-04	1.5809	downstream	151111	---	---
3	rs1549114	175121109	A	G	8.94E-01	1.02E-04	0.6158	intron	0	NAALADL2	7.29E-02
1	rs3013754	54061978	C	A	9.00E-01	1.42E-04	1.5911	intron	0	GLIS1	1.51E-02
11	rs620095	114061226	T	G	2.35E-01	1.42E-04	0.5049	intron	0	ZBTB16	6.04E-03
4	rs17661170	171175320	G	A	5.32E-01	1.47E-04	0.5947	upstream	19750	LOC101928223	---
1	rs4660017	241139151	C	T	5.44E-01	1.48E-04	1.9677	intron	0	RGS7	6.28E-02
18	rs17288584	69134662	A	G	4.07E-01	1.58E-04	2.4108	downstream	52536	LOC100505776	4.34E-01
16	rs4967886	65109912	A	G	3.33E-01	1.58E-04	1.5613	intron	0	CDH11	1.44E-03
12	rs4760786	71446789	C	T	9.00E-01	1.60E-04	1.5796	intron	0	---	---
2	rs3770979	207451594	C	A	3.32E-01	1.65E-04	2.2618	intron	0	ADAM23	2.06E-02
8	rs2514973	119253485	G	C	3.67E-01	1.65E-04	2.0776	intron	0	SAMD12	1.60E-01
14	rs11624164	68986157	T	C	2.76E-01	1.66E-04	1.6135	intron	0	RAD51B	4.20E-02
10	rs496627	38350398	G	A	6.77E-01	1.86E-04	1.5519	intron	0	ZNF33A	5.19E-04
13	rs9518449	102224923	A	G	2.99E-01	1.96E-04	2.3513	intron	0	ITGBL1	2.70E-01
10	rs10903846	2746750	C	T	9.83E-01	2.02E-04	1.7214	upstream	231463	---	---
4	rs2139980	157617198	A	G	4.49E-01	2.10E-04	0.6255	downstream	64408	PDGFC	2.72E-04
10	rs1208681	38092975	C	T	6.04E-01	2.11E-04	1.5446	intron	0	ZNF248	1.61E-02
1	rs4659586	241139692	A	T	4.85E-01	2.31E-04	1.9178	intron	0	RGS7	6.28E-02
10	rs11251437	2746118	G	A	8.39E-01	2.37E-04	1.7157	upstream	232095	---	---
10	rs7090858	38003670	C	A	6.23E-01	2.41E-04	1.5396	downstream	60629	---	---
18	rs4798677	8699988	A	G	8.88E-01	2.47E-04	1.5398	intron	0	GACAT2	5.07E-01
6	rs9366337	20252093	G	A	5.90E-01	2.51E-04	1.6021	intron	0	---	---
6	rs16877617	16089085	G	A	6.00E-01	2.55E-04	0.6177	upstream	18768	---	---

Key: CHR= Chromosome; BP= base pairs; A1= Allele 1; F A= Frequency of Allele 1; F U= Frequency of Allele 2; A2=Allele 2; CHISQ = chi squared. OR = odds ratio. Uncorrected p <0.05 in **bold**.

Table 16.2 Top 40 Genes Yielding Evidence for Association After VEGAS2 Permutation Testing

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Uncorrected gene based P value	Top SNP	Top SNP P
4	PDGFC	17	1000000	117.301571	2.72E-04	rs6827631	3.85E-04
5	IL6ST	4	1000000	32.79562082	3.68E-04	rs11956465	7.00E-04
12	AQP6	9	1000000	49.64338821	4.01E-04	rs7309966	2.14E-03
2	CHCHD5	8	1000000	44.12957422	4.13E-04	rs4849073	3.74E-04
12	AQP5	11	1000000	51.2609029	4.33E-04	rs7309966	2.14E-03
10	ZNF33A	15	1000000	94.03343212	5.19E-04	rs645064	9.20E-05
6	GCLC	11	1000000	46.81140748	5.65E-04	rs761141	6.69E-04
12	RACGAP1	10	1000000	48.96194208	5.68E-04	rs7309966	2.14E-03
2	POLR1B	9	1000000	45.64442756	7.20E-04	rs4849073	3.74E-04
12	AQP2	14	1000000	54.23592388	7.48E-04	rs7309966	2.14E-03
4	HGFAC	9	1000000	44.99965866	7.61E-04	rs13128489	5.98E-04
12	LOC101927318	14	1000000	54.23592388	7.76E-04	rs7309966	2.14E-03
3	GPR149	15	1000000	78.6904535	8.65E-04	rs701132	1.42E-03
20	GATA5	5	100000	24.86417385	9.70E-04	rs6061245	1.13E-03
10	ZNF37A	19	1000000	136.8578792	1.16E-03	rs645064	9.20E-05
18	LOC100652770	16	100000	81.60557849	1.19E-03	rs2096919	5.58E-04
12	KRT2	25	100000	116.7042829	1.30E-03	rs1464042	3.17E-04
11	ADAMTS8	13	100000	55.0290288	1.34E-03	rs11606448	2.11E-03
9	KCNV2	18	100000	64.31676007	1.37E-03	rs10967953	1.96E-05
16	CDH11	49	100000	193.6398273	1.44E-03	rs4967886	1.58E-04
12	MSI1	7	100000	33.42849535	1.95E-03	rs2701632	1.16E-03
17	EFNB3	5	100000	24.47388357	1.95E-03	rs2909430	2.83E-04
21	LINC00945	15	100000	66.52007164	2.10E-03	rs8128907	1.09E-03

Table 16.2 contd...

Chr	Gene	n SNPs	n Sims	VEGAS2 Test result	Gene based P value	Top SNP	Top SNP P
5	LOC102467073	6	100000	47.92295922	2.11E-03	rs7708515	1.95E-03
21	OLIG1	12	100000	55.40194253	2.14E-03	rs8128907	1.09E-03
10	TMEM26	15	100000	67.92123008	2.18E-03	rs10994750	2.14E-03
4	LRPAP1	14	100000	58.55143097	2.27E-03	rs13128489	5.98E-04
10	ITIH2	15	100000	52.84938343	2.31E-03	rs10508334	5.60E-04
4	DOK7	14	100000	58.70439946	2.32E-03	rs13128489	5.98E-04
9	SETX	10	100000	54.06500557	2.41E-03	rs10901156	8.89E-05
18	DSG2	23	100000	86.5474038	2.42E-03	rs2096919	5.58E-04
6	TXLNB	9	100000	30.5807926	2.49E-03	rs9484261	3.84E-03
5	PDCD6	7	100000	47.9370828	2.52E-03	rs7708515	1.95E-03
5	SDHA	7	100000	47.92308365	2.67E-03	rs7708515	1.95E-03
12	PLA2G1B	6	100000	29.19883618	2.70E-03	rs2701632	1.16E-03
3	PIK3CA	10	100000	47.45795547	2.78E-03	rs3975506	2.22E-03
20	PAX1	3	100000	27.08081683	2.80E-03	rs6035924	1.11E-03
3	MIR4789	11	100000	44.54119262	2.98E-03	rs1549114	1.02E-04
12	LOC101927292	17	100000	49.72407853	3.03E-03	rs2849266	4.12E-03
3	FOXP1	67	100000	164.3869724	3.17E-03	rs9878602	8.54E-04
9	LOC100507346	5	100000	31.07861393	3.24E-03	rs16909865	6.69E-04

Table 16.3 Meta-analysis vs normal controls: Genes Yielding Uncorrected Evidence for Association at P <0.05 After Permutation Testing and Replication

Chr	Gene symbol	n SNPs	n Sims	VEGAS Test result	Uncorrected gene based p-value	Top SNP	Top SNP p-value	AD/related phenotype replication (reference)
15	GABRB3	29	10000	61.62	1.55E-02	rs2114485	2.88E-03	Severity of alcoholism (Noble, 1998)
1	MTHFR	17	100000	69.76	4.29E-03	rs2075539	1.30E-03	Serum homocysteine levels in alcoholics (de La Vega 2001)
5	GPRIN1	6	10000	19.47	1.40E-02	rs10476198	3.17E-03	2108A allele/A-containing genotype over-represented in alcohol withdrawal seizures (Rejescu 2005)
7	GRM8	100	100000	262.99	7.49E-03	rs1211384	4.03E-03	Theta EEG and alcoholism (Chen, 2009)
3	GHSR	9	10000	32.70	9.00E-03	rs9881097	3.48E-03	Alcohol consumption and body weight (Landgren 2008)
11	OSBPL5	9	10000	21.76	2.93E-02	rs11025773	1.13E-03	AD (Edenberg, 2010)
11	PKNOX2	54	10000	104.40	2.24E-02	rs7104304	4.76E-04	AD (Chen, 2009; Beirut, 2010; Zuo, 2010, Guo, 2012, Zuo, 2014), substance dependence (Chen 2011)
1	ST6GALNAC3	98	10000	168.64	4.45E-02	rs12096789	5.76E-03	Nominal association theta oscillations (Hodgkinson, 2010)
15	CASC4	7	10000	29.12	2.23E-02	rs3809471	7.37E-03	AD (meta-analysis) (Wang, 2010)
2	TTL	7	100000	34.62	4.01E-03	rs4849073	3.74E-04	AD with conduct disorder, depression and suicide (P=0.0302) (Wang, 2007)
2	HAAO	17	100000	80.00	4.99E-03	rs7582883	1.86E-03	
2	COBLL1	33	10000	88.49	1.26E-02	rs355839	3.38E-03	
3	GPD1L	18	10000	57.55	1.48E-02	rs1562878	7.76E-03	AD (suggestive by nearby SNP) (Frank, 2012)
10	KCNMA1	157	10000	278.77	2.02E-02	rs16934688	3.16E-03	AD in SZ (Kendler, 2011)

Table legend: AD = alcohol dependence; BP = bipolar disorder; SZ = schizophrenia; nSims = number of simulations

Table 16.4 Top 16 PGC CDG and UCL VEGAS2 gene pathways

PGC CDG			UCL Meta vs SEMI		UCL Meta vs SNC		Pathway ID	Pathway Description
PGC Rank	Combined <i>P</i>	<i>q</i> -value	UCL Rank*	Empirical <i>P</i>	Rank*	Empirical <i>P</i>		
1	5.75×10^{-8}	0.0003	N/A	N/A	N/A	N/A	GO:51568	Histone H3-K4 methylation
2	1.46×10^{-5}	0.0362	907	0.11	896	0.111	GO:16571	Histone methylation
3	4.73×10^{-5}	0.0414	2172	0.33	N/A	N/A	GO:43414	Macromolecule methylation
4	5.10×10^{-5}	0.0414	2777	0.42	2738	0.424	GO:34968	Histone lysine methylation
5	5.58×10^{-5}	0.0414	837	0.1	829	0.103	GO:45216	Cell-cell junction organization
6	5.69×10^{-5}	0.0414	6010	0.96	5997	0.964	P00003	Alzheimer disease–amyloid secretase pathway
7	5.86×10^{-5}	0.0414	N/A	N/A	N/A	N/A	P04393	Ras pathway
8	7.12×10^{-5}	0.0422	313	0.048	400	0.033	GO:8601	Protein phosphatase type 2A regulator activity
9	7.83×10^{-5}	0.0422	N/A	N/A	N/A	N/A	GO:43297	Apical junction assembly
10	9.25×10^{-5}	0.0422	1113	0.14	1041	0.142	P00052	TGF- β signaling pathway
11	9.53×10^{-5}	0.0422	2477	0.37	2421	0.368	GO:14069	Postsynaptic density
12	0.0001	0.0422	898	0.11	855	0.11	GO:32869	Cellular response to insulin stimulus
13	0.0001	0.045	N/A	N/A	N/A	N/A	P00010	B cell activation
14	0.0001	0.045	3850	0.59	3935	0.588	GO:8757	S-adenosylmethionine–dependent methyltransferase activity
15	0.0001	0.0454	N/A	N/A	N/A	N/A	GO:23061	Signal release
16	0.0002	0.0473	2073	0.29	1984	0.296	GO:34330	Cell junction organization

Figure legend: PGC CDG = Psychiatric Genetics Consortium Cross Disorders Group. Meta = Meta analysis of BPALC + SZALC. SEMI = severe mental illness. SNC = supernormal controls. * of 6211 pathways reported by VEGAS2. UCL pathway **P** bold if $P \leq 0.05$.

Fig. 16-3 Meta-analysis vs SNC: strongest GWA SNP in *LTBP1*

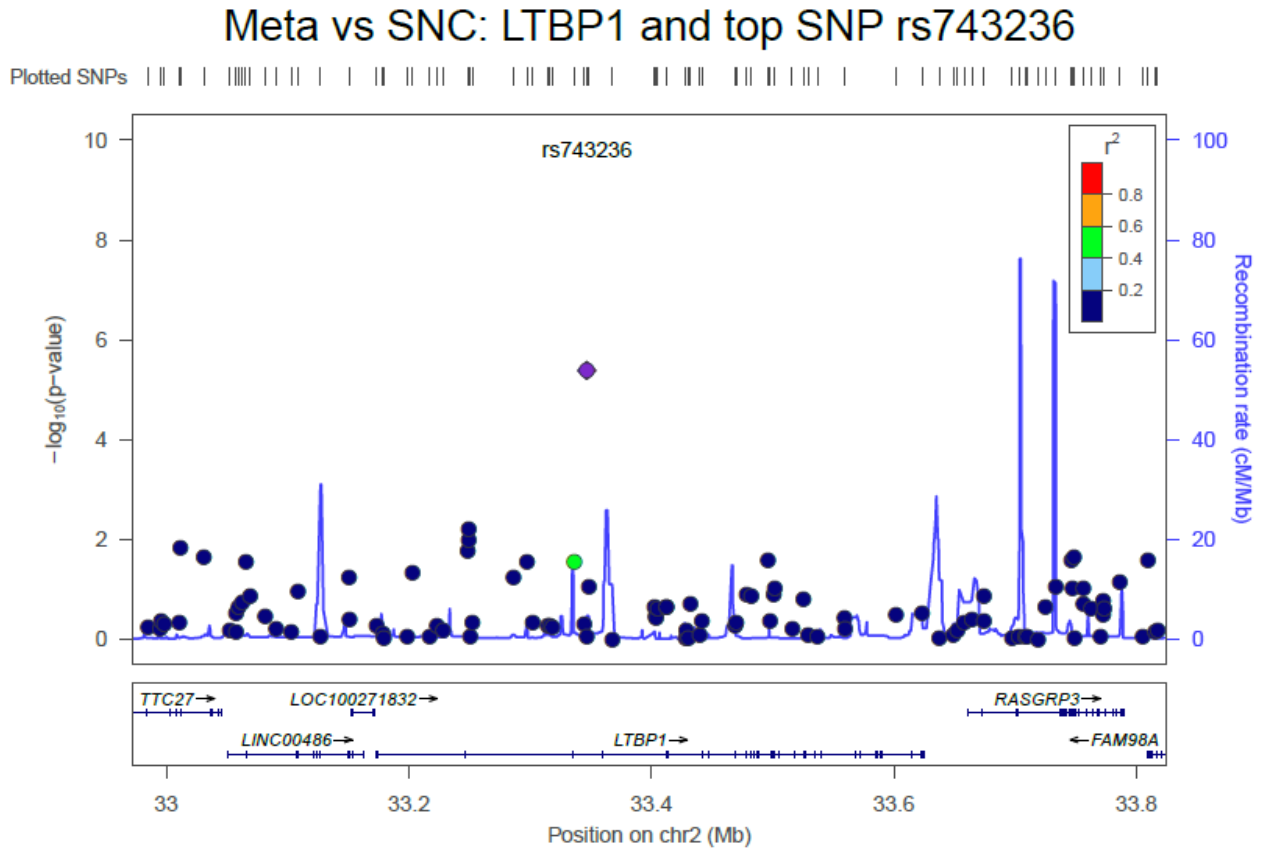


Figure key: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-4 PGC CDG: plotting the UCL meta-analysis strongest SNP in LTBP1
PGC CDG (Feb. 2013)

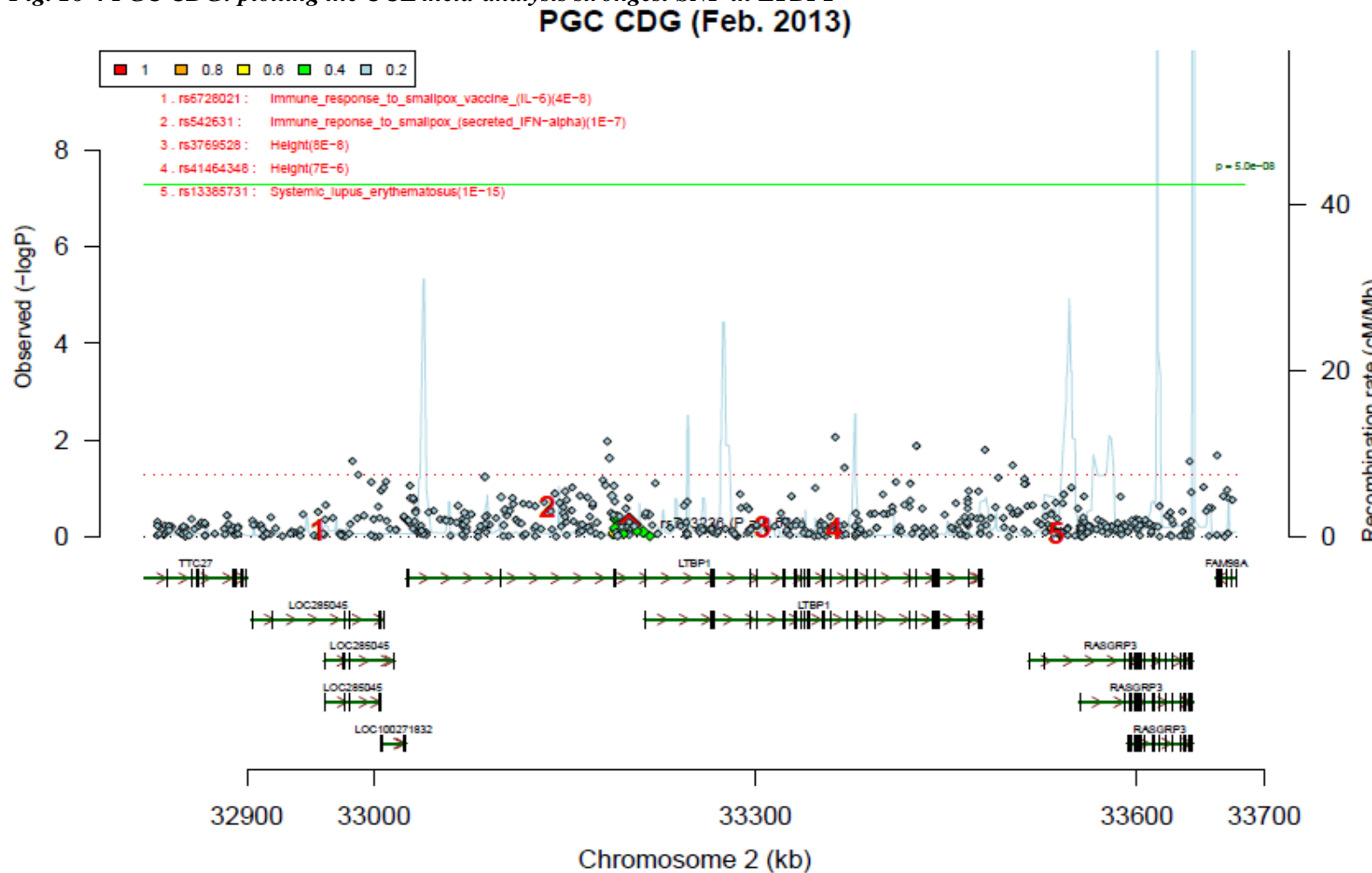


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position
 Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-5: Meta-analysis plot of UCL CDH11

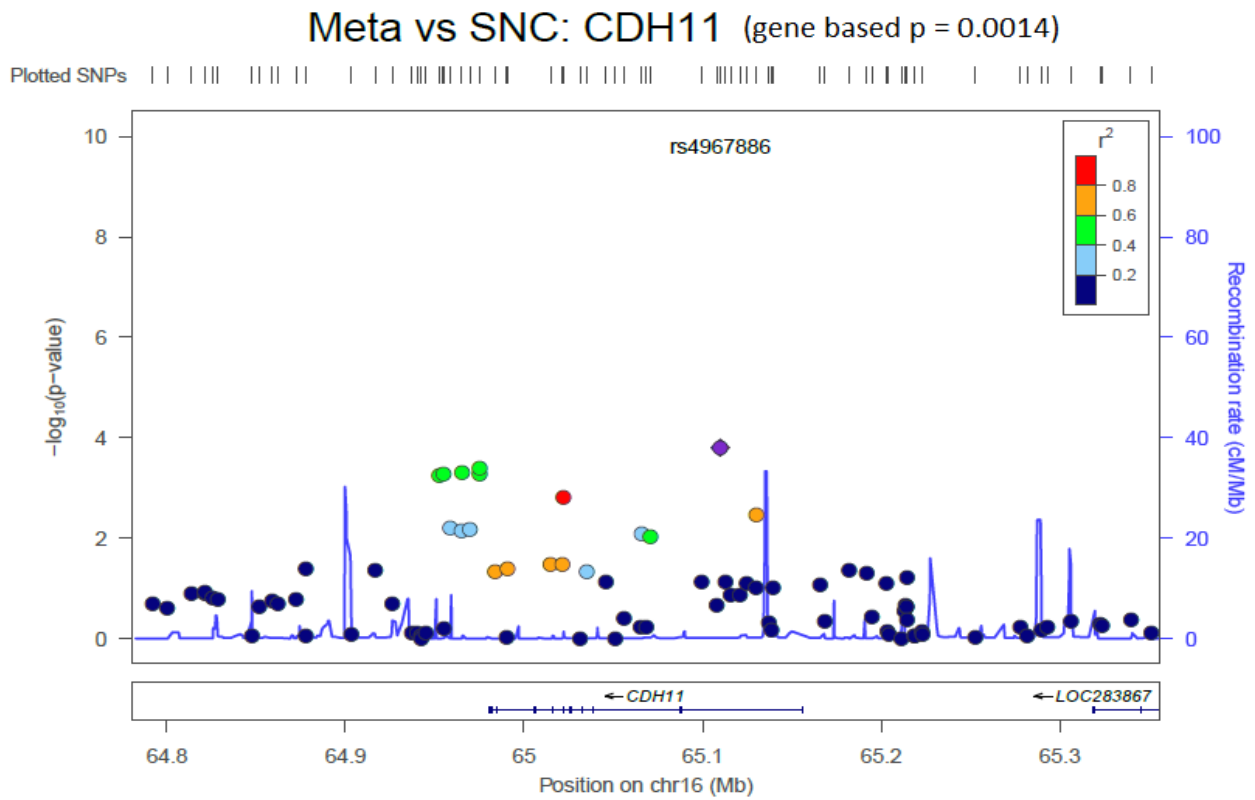


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-6 PGC CDG plot of CDH11 region

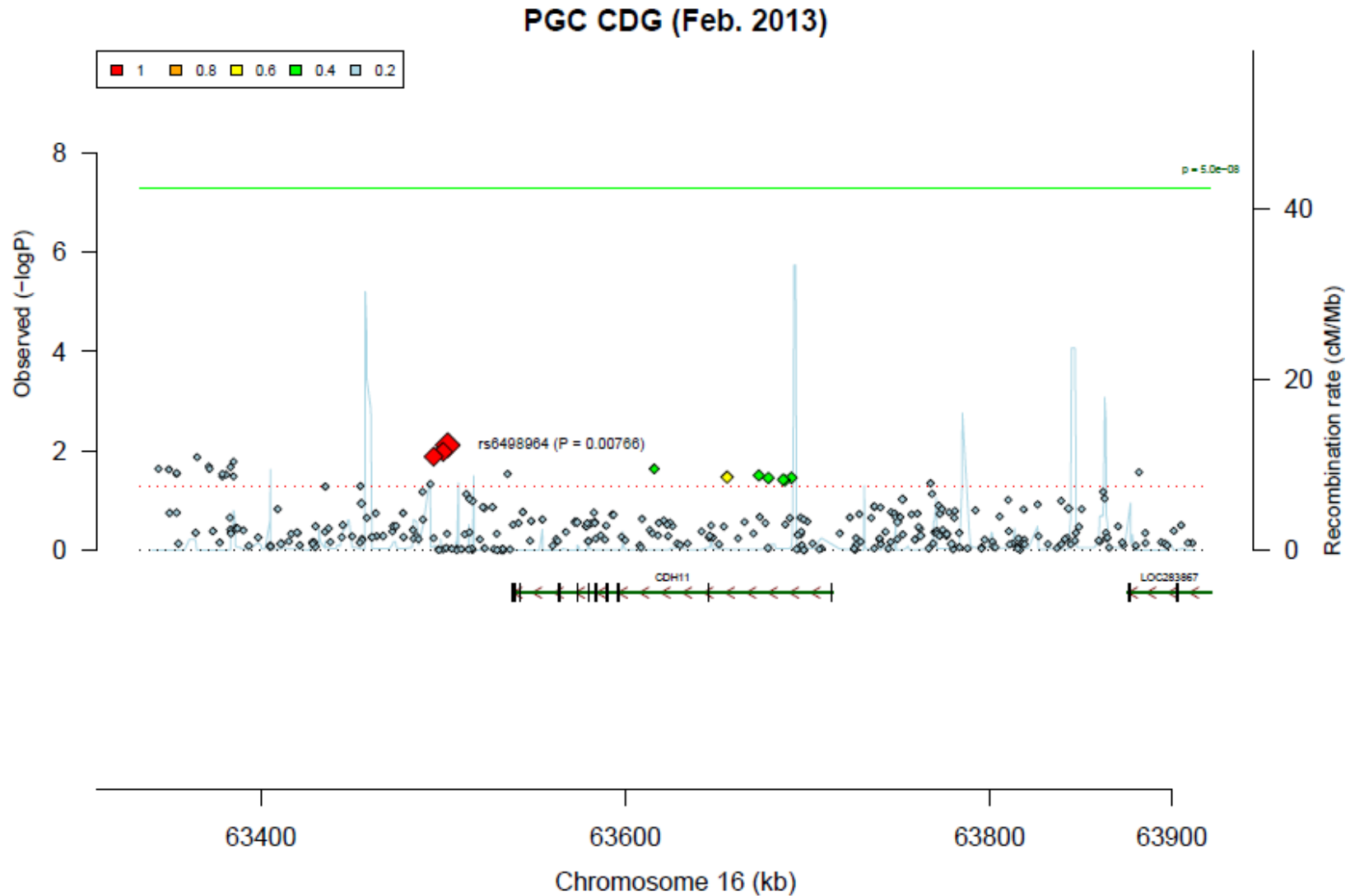


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position
Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-7 UCL meta-analysis vs SNC: Top gene based gene plot of PDGFC

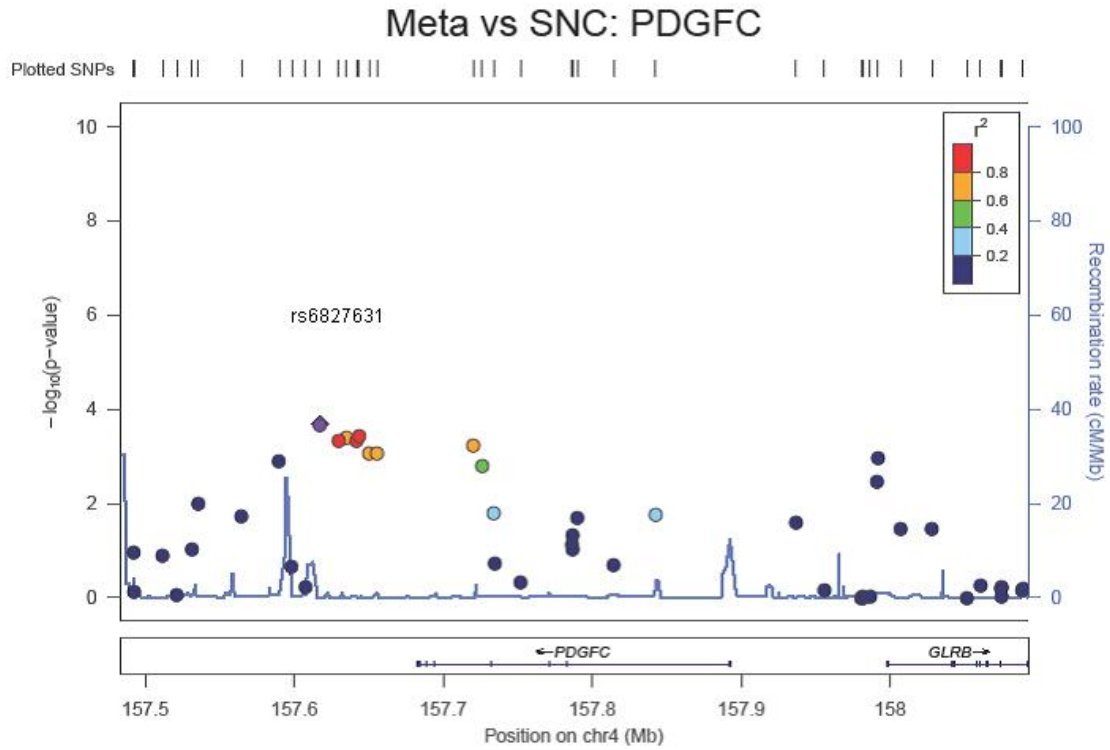


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-8 UCL meta-analysis vs SNC: UCL PKNOX2 region plot

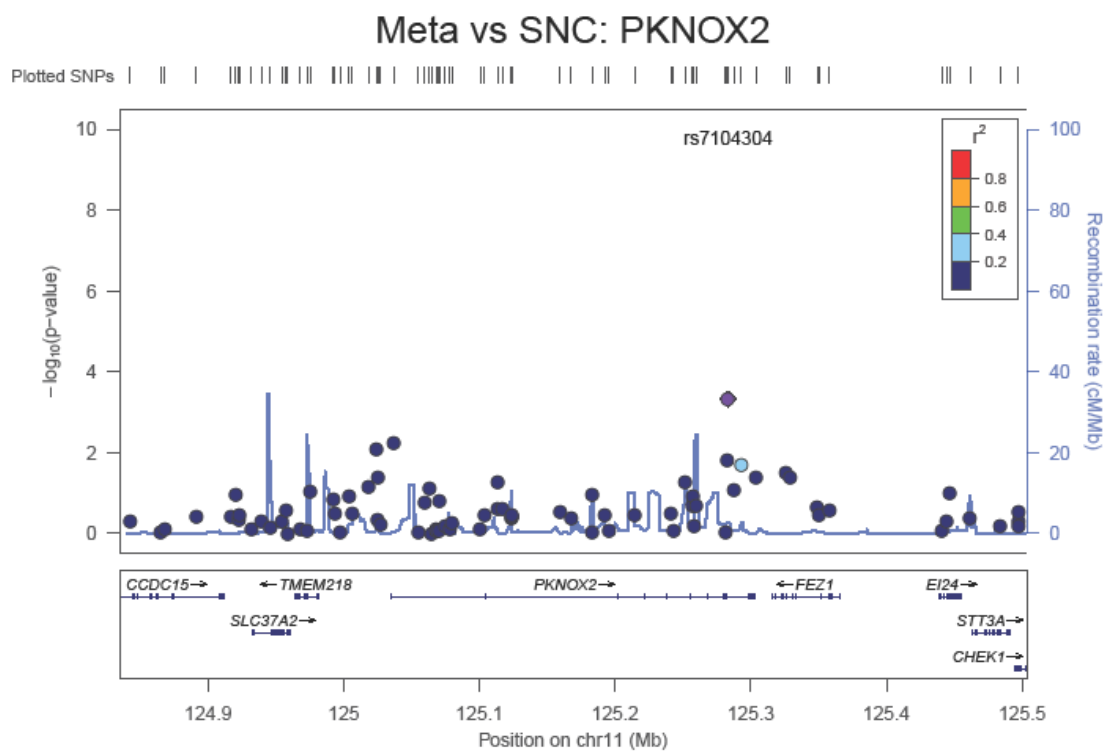


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-9 Meta-analysis vs SNC: PGC CDG PKNOX2 region plot

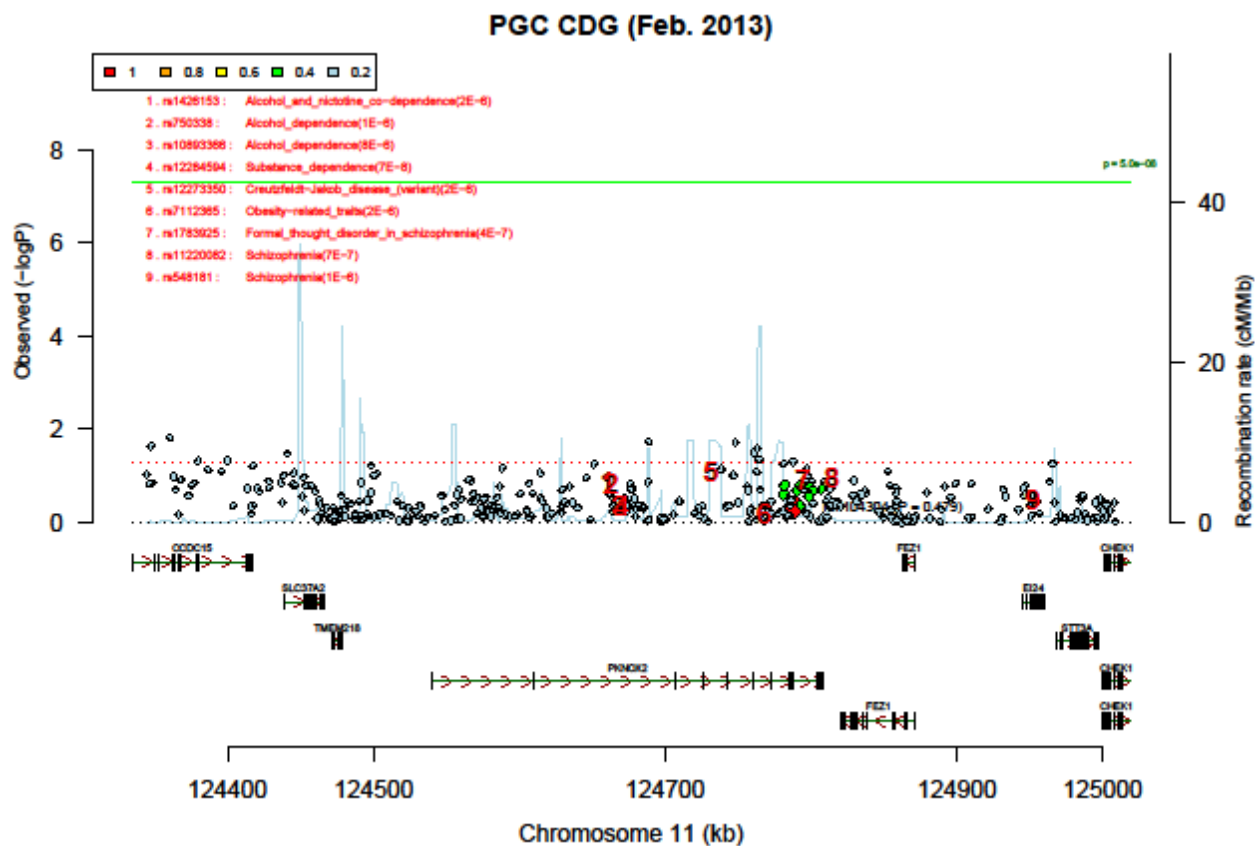


Figure legend: x-axis = association P-values on the $-\log_{10}$ scale; y-axis = chromosomal position

Purple diamond = index SNP. SNP colours are LD (r^2) estimates from 1000 genomes data.

Fig. 16-10 UCL meta-analysis vs SNC: number of genes in each GO pathway

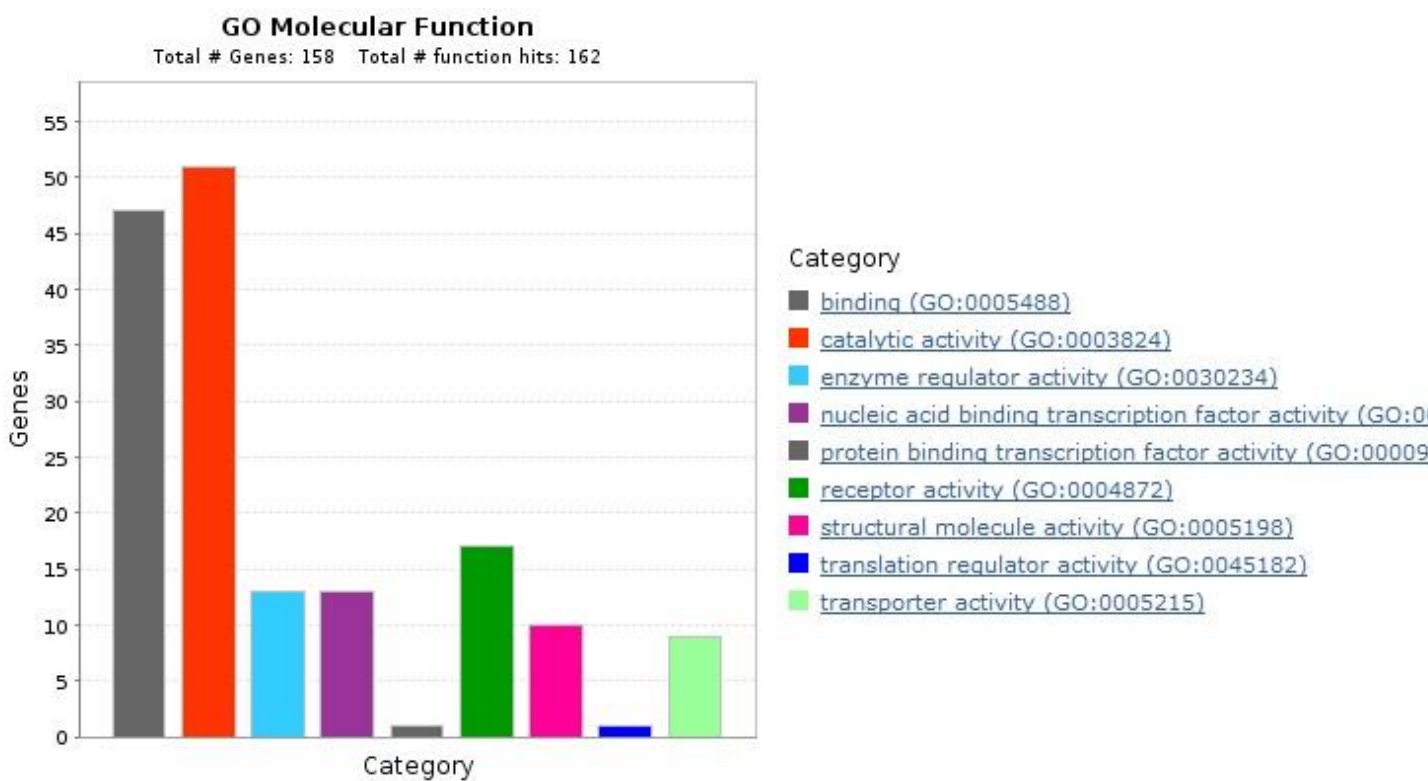


Figure legend: x axis = number of genes; x axis = GO molecular category

CHAPTER 17. DISCUSSION AND CONCLUSIONS

The new research component consisted of a case-control study of GABRA2 in AD and a series of GWAS of AD comorbid with severe psychotic illness. We performed the first genome wide association studies and meta-analyses of BPALC and SZALC compared with two different control groups. In addition to ancestrally matched supernormal controls, we uniquely used BP and/or SZ without AD as controls. The latter was a novel approach to “uncover” the underlying AD genes from the BP/SZ genes. To my knowledge this phenotype combination has not been published before (although SZ and AD versus normal controls (Kendler et al 2009) and SZ versus BP (Ruderfer, 2014) have been published).

17.1 GENERAL COMMENTS ON THESE GWAS

All these signals should be taken as tentative and biologically plausible candidate genes worthy of follow up in larger studies. With hindsight and the advancement of research during this degree, it is now not surprising that this study did not find markers which achieved gene-wise significance based on the small, by modern standards, sample size and the small effect sizes likely to impact on AD. However at the time of collection, the sample size was less modest by comparison to the literature of the time. For example Potkin et al published a GWAS on SZ with 64 cases and 77 controls (Potkin et al. 2009).

Strengths of the study include use of a well-characterised clinical sample assessed using validated research tools, ancestral homogeneity, and the use of ancestrally matched supernormal controls.

Validity of diagnoses was improved by cross checking with clinical notes and use of experienced recruiters. Technical quality control was rigorous. Numerous studies have been published from the UCL samples, many of which have been replicated by other groups in different populations, which may be

further evidence of validity (or common confounders). We conducted several levels of analysis to examine enrichment of signals in AD, including GWAS, gene-based and pathway analysis methods. These revealed plausible gene signals and pathways for AD, albeit not reaching genome wide or corrected gene wide significance. Furthermore the case-case approach to a complex disorder as a novel approach worthy of further exploration.

Weaknesses of the study include:

1. Modest sample size by modern standards. This power was vastly underpowered compared to modern GWAS and the chances of false negatives are high in low power studies (Sham & Purcell 2014). This was before larger international collaborations were commonplace and the recognition that much larger GWAS sample sizes would be needed to detect small genetic effect sizes in complex disorders. Another reason for the small sample size was pragmatic. While the genotyped BP and SZ samples on their own were of a good size for the time, the sizes shrunk dramatically when sub-dividing into ALC and nonALC groups. The UCL AD, BP and SZ samples have all grown considerably in the intervening time, thus future analyses will have greater power to detect a difference.
2. Different diagnostic tools were used for diagnosis of AD/alcoholism in each group. While each tool on its own was previously validated for the diagnosis of AD/alcoholism, BP and SZ, the OPCRIT, SADS-L and SSAGA-II may each differ sufficiently to introduce ascertainment bias. Ascertainment bias can induce spurious cross-phenotype (CP) effects and occurs when the recruitment of people with one phenotype increases the prevalence of another unrelated phenotype in the cohort and leads to a false correlation between them. An example is Berkson's bias which is common in clinically ascertained samples, as patients suffering from two conditions are often more likely to seek treatment than those with a single diagnosis (Westreich 2012).

3. Unaffected control individuals were shared across the BP and SZ studies. Common associations could occur if an artefact (such as population stratification or batch effects) systematically biased the shared controls and not the cases. In this study however I used half the normal controls in each half of the meta-analysis.

4. Lack of imputation. We used an older generation chip for genotyping and this was several years ago. It holds 500k SNP markers, which is low by modern standards but cutting edge at the time. This limits the density of coverage of the human genome, and imputation is one method to overcome this. The bulk of this study was done before imputation was widely in use, and this is a shortcoming of the study.

5. False effects between two phenotypes can also occur when subjects with one phenotype are systematically misclassified with a different phenotype. For example, patients with BP can sometimes be misdiagnosed with SZ, and vice versa, and this could result in spurious genetic correlation between the traits. In the cases of schizophrenia and bipolar disorder, the misclassification rate would need to be larger than 20% to generate the genetic correlation (0.60) observed between the traits if the true genetic correlation were zero (Solovieff et al. 2013). Thus the misclassification rate must be quite high to have a substantial impact on the genetic correlation, and it is hoped that the clinical ascertainment methods limited this bias.

These sources of bias can be avoided with careful study design. General guidelines for study design of GWAS should be followed, including appropriate control selection, adequate quality control and adjustment for population stratification. Population stratification is a major source of confounding in GWASs, and established methods of population stratification adjustment should be applied within studies. In addition, appropriate adjustments for multiple testing should be implemented to avoid false-positive associations. Similar populations should be used across studies, as the underlying linkage

disequilibrium structure varies across populations and the SNP of interest may differentially highlight the causal variant. A further consideration is that when combining phenotypes across studies, participants may have different phenotypes, or different sets of SNPs owing to differences in genotyping arrays. Genotype imputation can be used to obtain the same set of SNPs for all individuals (Solovieff et al. 2013).

17.2 GENE-BASED ANALYSES

The use of gene-based tests of association is worthy of further discussion. The gene-based test detects genes which show a greater signal of association than expected by chance given the LD between the SNPs and the gene's length, while also accounting for number of SNPs. This in effect gives a significance value at the level of the whole gene as opposed to the whole genome.

Gene-based analysis has several advantages. First, these analyses can account for multiple independent functional variants within a gene, with the potential to greatly increase the power to identify disease/trait-associated genes. Second, variations in protein-coding and adjacent regulatory regions are more likely to have functional relevance. Third, gene-based tests allow for direct comparison between different populations, despite the potential for different LD patterns (Neale & Sham 2004).

In gene-based replication, the gene identified by the initial study is consequently examined for association with effectively all genetic variants in the intragenic and regulatory regions. Neale & Sham describe gene-based replication as the "gold standard" for three reasons: 1) genes are the functional unit of the human genome, and the positions, sequence, and function of genes are highly consistent across diverse human populations. This universality is considerably greater than that of either a SNP or a haplotype. Gene-based analysis might therefore lead to more consistent results and

alleviates difficulties in replication. 2) gene-based replication implies that each population studied will have local allele frequencies and LD structure accounted for, which should address problems with non-replication due to population differences. 3) gene-based replication simplifies the multiple-testing problem by conveniently dividing it into two stages, dealing first with the multiple variants within a gene and then with the multiple genes in the genome (Neale & Sham 2004).

In the BPALC paper (Lydall et al 2011a) we used a method whereby we combined the p values of the SNPs in a gene by Fisher's combination test (Curtis et al. 2012). This method assumes that the constituent p values should be based on independent tests, which is unlikely to be true for SNPs in the same gene. Violation of this assumption is likely to inflate the type I error rate, unless use of a permutation procedure provides empirical statistical significance (Li et al. 2011). Of the multitude of gene-based analyses which use differing statistical approaches each with their advantages and disadvantages the VEGAS approach was robust and user-friendly. The more popular gene-based and pathway methods were recently compared by Wojcik et al (2015). The gene-level program VEGAS had the highest sensitivity (28.6%) with less than 1% false positives (Wojcik et al. 2015). VEGAS also supports robust pathway analyses in the same batch as gene-based analyses which represented efficient use of research time.

The results of our analysis are presented with correction for multiple testing. Bonferroni correction itself may be conservative if the tests are not independent – and we do not know enough about effects between blocks to be confident (Gao et al. 2010). Another reason why this correction may be conservative is that the literature replication analyses already give evidence of prior association (Sham & Purcell 2014). However our significance values should be interpreted with caution because the appropriate method for correcting these analyses is unclear (Johnson et al. 2010; Sham & Purcell 2014).

Conversely, using the literature as control increases the risk of confirmation bias (“seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand”), citation bias (Pannucci & Wilkins 2010) and positive outcome bias (Emerson et al. 2010) because of the higher likelihood of positive studies being published (MacCoun 1998).

17.3 PATHWAY ANALYSES

Statistical association, while important, is not the only consideration for concluding whether variation in a gene plays a role in causing individual differences in disease susceptibility. Other relevant considerations include biological plausibility, animal models, and gene-expression studies (Page et al. 2003). Since genetic variation must be ultimately translated into differences in gene function to affect disease risk, pathway-based approaches have been developed which use prior biological knowledge on gene function to facilitate more powerful analysis of GWAS data sets. These approaches typically examine whether a group of related genes in the same functional pathway are jointly associated with a trait of interest. Pathway-based analysis is considered an effective technique that overcomes the limitations of the current single-locus methods. This procedure aims to provide a comprehensive understanding of the molecular mechanisms underlying complex diseases (Wang et al. 2007; Wang et al. 2010).

Principally, a pathway-based approach is similar to the Gene Ontology (GO) analysis which, for example, has been used to provide insights into the biology of BP disorder (Holmans et al. 2009). The GO database contains large amounts of electronic annotations based on studies of human orthologues in model organisms, so that researchers can incorporate computationally predicted gene sets and pathways in the gene-enrichment analysis to improve power. Its capacity of capturing biological interaction among genes and improving power and robustness has been well recognized.

Jin and colleagues (2014) in an excellent recent review summarise the advantages of pathway-based approaches compared to single gene analysis methods. These include: 1) Pathway-based approaches appear to be well suited for analysis of massive GWAS data. 2) Since pathway-based approaches focus on sets of genes instead of individual genes, dimension reduction is automatically achieved and the issue of the multiple-test corrections is reduced when a large number of SNPs are examined. 3) Common diseases often arise from the joint action of multiple SNPs/genes within a pathway. Although each single SNP may confer only a small disease risk, their joint actions are likely to have a significant role in the development of disease. If one only considers the most significant SNPs, the genetic variants that jointly have significant risk effects but make only a small contribution if considered individually will be missed. 4) Locus heterogeneity, in which alleles at different loci cause disease in different populations, will increase the difficulty in replicating associations of a single marker with a disease. The list of significant SNPs from several studies may have little overlap leading to minimal replication of association findings at the SNP level. In comparison, pathway-based approaches that utilize information from multiple loci in a functional unit could produce more stable and robust results than single gene analyses do. 5) Fourth, the ultimate purpose of genetic studies of complex diseases is to decipher the path from genotype to phenotype. Despite the wealth of data, connections between DNA variations and complex phenotypes, which are essential for unravelling pathogeneses of complex diseases and predicting variation in human health, have remained elusive. In this sense, pathway-based approaches provide a complementary role to single-point analysis for interpreting the molecular paths underlying human diseases (Jin et al. 2014).

Several studies have demonstrated that pathway-based analysis is superior when it is applied to large-scale genetic datasets for rheumatoid arthritis (RA), type 2 diabetes (T2D), SZ, Parkinson's disease, etc. In addition, tracing the shared pathways among several pathologies tends to be an ongoing interest of disease pleiotropism, for example, the study of genetic links between RA and systemic lupus erythematosus, SZ and T2D (comprehensively reviewed in Jin et al. 2014).

As always in biological research, there are caveats. The current understanding of human gene function is incomplete, so the curated gene sets and pathways are not a comprehensive representation of functionally related gene cohorts in the human genome. A large number of genes in the human genome are uncharacterized or poorly characterized, so that there may not be any pathway information available. Information on SNPs near these genes may not be incorporated in the pathway-based approaches, so our approach can complement but not replace the single-SNP approach (Wang et al. 2007).

The reader should further be aware that, despite these advantages, gene pathway analyses do not specifically confirm individual gene associations. While an association with a gene may suggest potential association of another in the same biological pathway, it does not by itself confirm the involvement of the pathway gene, which would still arguably require an association signal between that and the phenotype.

17.4 “MISSING HERITABILITY”

It is known that most psychiatric conditions are generally highly heritable. However, at the time of writing, numerous GWAS appear to have limited power to predict the variation in psychiatric illness status at individual level, or develop novel pharmacological strategies, despite the huge amount of data

generated. Genetic risk profile scores (RPS) are constructed from alleles showing modest association with schizophrenia can have low sensitivity and specificity in predicting caseness. Combining all identified variants together may explain less than 7% of the total variance in disease risk between individuals (Ripke et al. 2014).

There are various theories of where the "missing heritability" lies: theories include gene-gene interactions (epistasis), larger effects in the truly causal variants than in the SNPs, or in rare variants of large effect, population heterogeneity, or underestimated environmental or epigenetic effects (both environment and epigenetic effects can be inherited) (Heard & Martienssen 2014).

The power of genome-wide association studies to detect genes depends on a strong link between the studied phenotype and genotype. There are several reasons why there is a more modest success in finding genes with replicable associations with psychiatric diseases in comparison to other complex human diseases.

Firstly, there is the precision to which we can diagnose psychiatric disorders today (Burmeister et al. 2008; Schulze 2010; Plomin et al. 2009). If the phenotype cannot be quantified accurately with resultant misclassification of cases and controls, the power of the study is reduced. Psychiatric disorders by their nature are arguable less easy to objectively "measure" in comparison to other diseases that have reasonably precise diagnostic criteria, such as Crohn's disease (Wellcome Trust, 2007; Webb et al. 2008; Shi et al. 2009; Eichler et al. 2010). Another challenge will come with the recent change in diagnostic criteria with the publication of DSM-V (this is explained in detail in Section 2.4). While merging the diagnostic criteria into a single 'substance use disorders' cluster with varying degrees of severity might ease the definition of endophenotypes or of semi-quantitative approaches like symptom count, care must be taken when comparing studies based on DSM-V with previously published DSM-

IV-based work, and appropriate approaches to quantitative rather than binary phenotypes used. Further formal cross-validation of different diagnostic systems and tools with respect to genetic studies is needed.

A more sophisticated characterization of the genetic susceptibility to AUD will most likely require (a) a more refined characterization of alcohol-related phenotypes (b) advanced analytic techniques that make use of both array and next-generation sequencing technologies and (c) new analytical techniques to model gene, gene-gene, gene x environment, epigenetic, and environmental influences, with sufficiently massive sample sizes. Ultimately it is hoped that we will move past descriptive syndromes in psychiatry towards a nosology informed by disease cause (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013).

Secondly is sample size and power calculation. Generally for genetic studies, bigger (and better classified phenotypes) is better. Innovations to increase sample size and offer longitudinal observation include the NIMH Research Domain Criteria (RDoC) project, which focuses on research integrating dimensions of behaviour and neurobiology (Cuthbert & Insel 2013). Longitudinal studies, e.g. following up large cohorts of twins from early childhood through adult life, have potential to give further insights into how genetic and environmental influences on development link with the emergence of clinical disorders, and also the aetiological relationships between clinical disorders and the occurrence of subclinical symptoms in the general population (Fagnani et al. 2011; Lin et al. 2007; Linney et al. 2003; van Os & Kapur 2009). Development of an Alcohol Dependence Biobank, similar to the one for BP (Frye et al. 2015) and an international AD consortium would serve as a resources for larger sample sizes and clinical and biomarker studies of disease risk and treatment response.

The predominant use of current main-lifetime diagnostic categories may change to a more flexible approach (Owen 2011) in which endophenotypes, categorical/dimensional syndromes, or networks of correlated symptoms are used. Studies of simpler phenotypes, such as e.g. human height or body mass index, have now reached more than 250,000 participants in the world-wide endeavour to uncover the “genetic architecture” of complex human traits. Future larger sample sizes will increase power to detect a difference but with appropriate statistical techniques. GWA of endophenotypes will increase precision further compared with the complex and rather broad phenotypes of human psychiatric diseases (Burmeister et al. 2008; Plomin et al. 2009; Schulze 2010).

Third is genetic heterogeneity. For AD, functional variations in genes involved in alcohol metabolism and taste modulate alcohol-related behaviours and thereby affect the individual risk for AD. In populations of European ancestry, in which the coding variations in ALDH and TASR238 are far less common, the greatest proportion of the individual genetic risk is yet unexplained and probably reflects the accumulating effects of multiple genes with minor individual impact.

Large, rare CNVs may have a greater influence on the risk of one disorder than another, but the current evidence for the role of CNVs in AD is limited compared to SZ and BP. It is also likely that there are common risk variants that are not shared between the disorders studied. While there is considerable genetic overlap between SZ and BP, the molecular overlap with these diagnoses and AD is not well established. Lastly, between the commonly occurring variants detectable by GWAS and rarer CNVs, there are likely to be risk variants of intermediate frequency and effect size, and also rare single-nucleotide mutations (Sullivan et al. 2012). CNVs in AD need significant further investigation in order to make meaningful comparisons with other complex brain disorders.

High-throughput next-generation sequencing studies are underway, aimed at uncovering both inherited and de novo genetic variants in these ranges (Girard et al. 2011; Xu et al. 2012; Need et al. 2012; Gulsuner et al. 2013) .

17.5 FUTURE DIRECTIONS

There are a number of future avenues I would like to explore beyond this work.

Firstly I would like to reanalyse the data using imputation methods which would help overcome some of the limitations of the design. Secondly we would like to secure funding for genotyping the full UCL ADS sample (which has doubled in size since I started) – unfortunately, GWAS funding has been unsuccessful thus far but is ongoing. Prof Gurling had started discussions about “iCOGA” (the International Collaboration on the Genetics of Alcoholism) and I think this would be an enormously progressive step. I would be honoured if our AD sampled could join in with a global GWAS or consortium like the PGC substance use disorders (SUD) group if we are able to complete genotyping. The Psychiatric Genomics Consortium has entered a growth phase that will, by the end of 2015, result in a tripling of the number of psychiatric cases with GWAS data from ~80 000 to ~240 000. This is expected to accelerate gene discovery and provide more clarity into interrelationships between complex psychiatric disorders (Crowley & Sakamoto 2015).

I would also like to repeat the case vs case (phenotype A vs phenotype B) approach with larger samples of BPALC and SZALC, and also other psychiatric phenotypes commonly co-occurring with AD, once sample sizes are sufficient.

The use of sequential gene based scanning of genome wide data (as in VEGAS2) intrigues me because it may allow for smaller studies to contribute to science (an appropriate p-threshold will need to be decided). Risk of such an approach is missing influences in non-genic regions, which GWAS overcome, albeit at the expense of very large sample sizes. Approaches may include:

1. Protein-protein interaction analysis of gene wide findings.
2. Looking at transcriptomics analysis of data from either post mortem ADS/blood leucocytes of living subjects/from animal models and overlaying that data with that from the genetics.

Lastly, translational approaches eg convergent functional genomics (CFG) have been used to generate a genetic risk prediction score in AD, SZ and BP (Levey et al. 2014). This requires GWAS data and other genetic and gene expression data, from human and animal model studies. I would be interested to see whether gene expression data can be extracted from the UCL blood samples and if so to apply these techniques.

As we learn more about the genetic framework of AUD, we find that different genetic variants lead to varying responses to treatment options, which might improve our ability to design individualized therapies – under the banner of personalized medicine. One example for the potential of pharmacogenetic approaches is the association of the OPRM1 As- n40Asp variation with the efficacy of the opioid antagonist naltrexone on relapse prevention (Ray et al. 2012). In the era of self-genotyping, more patients will be informed about their own genotypes and interested in their family histories, which presents with interesting clinical conundra, but also opportunities. Critical to the validity of this process is robust genomic, epistatic and epigenetic data.

17.6 FINAL REMARKS

I hope that this work will go some small way to contributing to science and in time improved patient care. I am hugely grateful to all of the patients and their families who gave so much to this study. I am immensely thankful for researchers and colleagues upon whose shoulders I have stood this thesis, and I thank you the reader for bearing with me on this journey.

REFERENCES

- A Widiger, T., 2011. Personality and psychopathology. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 10(2), pp.103–6.
- Abdolmaleky, H.M. et al., 2006. Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. *Human molecular genetics*, 15(21), pp.3132–45.
- Adegbola, A. et al., 2008. A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *American journal of medical genetics. Part A*, 146A(4), pp.505–11.
- Agrawal, A. 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 and Alcohol Dependence*, 93(1-2), pp.12–20.
- Ait-Daoud, N. et al., 2012. Preliminary Evidence for cue-induced Alcohol Craving Modulated by Serotonin Transporter Gene Polymorphism rs1042173. *Frontiers in psychiatry*, 3, p.6.
- Alaerts, M. & Del-Favero, J., 2009. Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future. *Human Mutation*, 30(8), pp.1139–1152.
- Ali, M.A. et al., 2015. Phenotypic heterogeneity in study populations may significantly confound the results of genetic association studies on alcohol dependence. *Psychiatric genetics*, 25(6), pp.234–40.
- Alloy, L.B. et al., 2005. The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors. *Clinical psychology review*, 25(8), pp.1043–75.
- De Alwis, D. et al., 2014. ADHD symptoms, autistic traits, and substance use and misuse in adult Australian twins. *Journal of studies on alcohol and drugs*, 75(2), pp.211–21.
- Ament, S. a et al., 2015. Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), pp.3576–81.
- American Psychiatric Association, 2013. *Diagnostic and statistical manual of mental disorders : DSM-5*, Washington D.C.: American Psychiatric Association.
- Anstee, Q.M. et al., 2013. Mutations in the Gabrb1 gene promote alcohol consumption through increased tonic inhibition. *Nature communications*, 4, p.2816.
- Comuzzie, A.G. et al., 2012. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *PLoS ONE*, 7(12).
- Anton, R.F. et al., 2008. An evaluation of mu-opioid receptor (OPRM1) as a predictor of naltrexone response in the treatment of alcohol dependence: results from the Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (COMBINE) study. *Archives of general psychiatry*, 65(2), pp.135–44.
- American Psychiatric Association, 2000. *Diagnostic and statistical manual of mental disorders : DSM-IV*, Washington D.C.: American Psychiatric Association.
- Arnone, D. et al., 2009. Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. *The British journal of psychiatry : the journal of mental science*, 195(3), pp.194–201.
- Axelrod, J. & Tomchick, R., 1958. Enzymatic O-methylation of epinephrine and other catechols. *The Journal of biological chemistry*, 233(3), pp.702–5.
- Azorin, J.-M. et al., 2010. Possible new ways in the pharmacological treatment of bipolar disorder and comorbid alcoholism. *Neuropsychiatric disease and treatment*, 6, pp.37–46.
- Badner, J. a & Gershon, E.S., 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Molecular psychiatry*, 7(4), pp.405–411.
- Baethge, C. et al., 2008. Sequencing of substance use and affective morbidity in 166 first-episode bipolar I disorder patients. *Bipolar disorders*, 10(6), pp.738–41.
- Baker, G.R. et al., 2004. The Canadian Adverse Events Study: the incidence of adverse events among

- hospital patients in Canada. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, 170(11), pp.1678–86.
- Barnett, J.H. & Smoller, J.W., 2009. The genetics of bipolar disorder. *Neuroscience*, 164(1), pp.331–343.
- Bart, G. et al., 2005. Increased attributable risk related to a functional mu-opioid receptor gene polymorphism in association with alcohol dependence in central Sweden. *Neuropsychopharmacology* 30(2), pp.417–422.
- Bass, N.J. et al., 2009. Evidence for the association of the DAOA (G72) gene with schizophrenia and bipolar disorder but not for the association of the DAO gene with schizophrenia. *Behavioral and brain functions: BBF*, 5(1), p.28.
- Bassett, A.S. et al., 1998. 22q11 deletion syndrome in adults with schizophrenia. *American journal of medical genetics*, 81(4), pp.328–37.
- Bassett, A.S. & Chow, E.W., 1999. 22q11 deletion syndrome: a genetic subtype of schizophrenia. *Biological psychiatry*, 46(7), pp.882–91.
- Baum, a E. et al., 2008. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Molecular psychiatry*, 13(5), pp.466–467.
- Begleiter, H. et al., 1998. Quantitative trait loci analysis of human event-related brain potentials: P3 voltage. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 108(3), pp.244–250.
- Begleiter, H. & Porjesz, B., 2006. Genetics of human brain oscillations. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, 60(2), pp.162–71.
- Belcher, A.M. et al., 2014. Personality traits and vulnerability or resilience to substance use disorders. *Trends in cognitive sciences*, 18(4), pp.211–7.
- Benson, D.L. & Tanaka, H., 1998. N-cadherin redistribution during synaptogenesis in hippocampal neurons. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 18(17), pp.6892–6904.
- Benyamina, A. et al., 2009. Association between MTHFR 677C-T polymorphism and alcohol dependence according to Lesch and Babor typology. *Addiction biology*, 14(4), pp.503–5.
- Bhaskar, L.V.K.S. et al., 2013. Association between neuropeptide Y gene polymorphisms and alcohol dependence: a case-control study in two independent populations. *European addiction research*, 19(6), pp.307–13.
- Bidwell, L.C. et al., 2015. Novelty Seeking as a Phenotypic Marker of Adolescent Substance Use. *Substance abuse: research and treatment*, 9(Suppl 1), pp.1–10.
- Bierut, L.J. et al., 2010. A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), pp.5082–5087.
- Blackwood, D.H. et al., 2001. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *American journal of human genetics*, 69(2), pp.428–433.
- Blöink, R. et al., 2005. Factorial structure and internal consistency of the German TEMPS-A scale: validation against the NEO-FFI questionnaire. *Journal of affective disorders*, 85(1-2), pp.77–83.
- Blomqvist, O., Gelernter, J. & Kranzler, H.R., 2000. Family-based study of DRD2 alleles in alcohol and drug dependence. *American journal of medical genetics*, 96(5), pp.659–64.
- Blum, K. et al., 1990. Allelic association of human dopamine D2 receptor gene in alcoholism. *JAMA: The Journal of the American Medical Association*, 263(15), pp.2055–2060.
- Bouby, L., Boissinot, P. & Marinval, P., 2011. Never Mind the Bottle. Archaeobotanical Evidence of Beer-brewing in Mediterranean France and the Consumption of Alcoholic Beverages During the 5th Century BC. *Human Ecology*, 39(3), pp.351–360.
- Boutte, D. et al., 2012. Association of genetic copy number variations at 11 q14.2 with brain regional volume differences in an alcohol use disorder population. *Alcohol (Fayetteville, N.Y.)*, 46(6),

- pp.519–27.
- Brady, K.T. & Lydiard, R.B., 1992. Bipolar affective disorder and substance abuse. *Journal of clinical psychopharmacology*, 12(1 Suppl), p.17S–22S.
- Bucholz, K.K. et al., 1994. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, 55(2), pp.149–158.
- Bucholz, K.K. et al., 1996. Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcoholism, clinical and experimental research*, 20(8), pp.1462–71.
- Buck, K.J. et al., 1997. Quantitative Trait Loci Involved in Genetic Predisposition to Acute Alcohol Withdrawal in Mice. *J. Neurosci.*, 17(10), pp.3946–3955.
- Buckley, P.F., 1998. Substance abuse in schizophrenia: a review. *The Journal of clinical psychiatry*, 59 Suppl 3, pp.26–30.
- Bulwa, Z.B. et al., 2011. Increased consumption of ethanol and sugar water in mice lacking the dopamine D2 long receptor. *Alcohol (Fayetteville, N.Y.)*, 45(7), pp.631–9.
- Burmeister, M., McInnis, M.G. & Zöllner, S., 2008. Psychiatric genetics: progress amid controversy. *Nature reviews. Genetics*, 9(7), pp.527–40.
- Cadore, R.J. et al., 1995. Genetic-environmental interaction in the genesis of aggressivity and conduct disorders. *Archives of general psychiatry*, 52(11), pp.916–24.
- Cadore, R.J., Troughton, E. & O’Gorman, T.W., 1987. Genetic and environmental factors in alcohol abuse and antisocial personality. *Journal of studies on alcohol*, 48(1), pp.1–8.
- Cao, Q. et al., 1997. Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. *Genomics*, 43(1), pp.1–8.
- Caporaso, N. et al., 2009. Genome-Wide and Candidate Gene Association Study of Cigarette Smoking Behaviors P. H. Reitsma, ed. *PLoS ONE*, 4(2), p.e4653.
- Carlson, C.S. et al., 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *American journal of human genetics*, 74(1), pp.106–120.
- Carmiol, N. et al., 2014. Shared genetic factors influence risk for bipolar disorder and alcohol use disorders. *European psychiatry : the journal of the Association of European Psychiatrists*, 29(5), pp.282–7.
- Carrigan, M.A. et al., 2015. Hominids adapted to metabolize ethanol long before human-directed fermentation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2), pp.458–63.
- Caspi, a., 2002. Role of Genotype in the Cycle of Violence in Maltreated Children. *Science*, 297(5582), pp.851–854.
- Centre, Health & Social Care., 2015. *Statistics on Smoking - England 2015*,
- de Chaldée, M. et al., 1999. Linkage disequilibrium on the COMT gene in French schizophrenics and controls. *American journal of medical genetics*, 88(5), pp.452–7.
- Chambers, J.C. et al., 2011. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics*, 43(11), pp.1131–8.
- Chamorro, A.-J. et al., 2012. Association of μ -opioid receptor (OPRM1) gene polymorphism with response to naltrexone in alcohol dependence: a systematic review and meta-analysis. *Addiction biology*, 17(3), pp.505–12.
- Chanock, S.J. et al., 2007. Replicating genotype-phenotype associations. *Nature*, 447(7145), pp.655–60.
- Chartier, K.G., Hesselbrock, M.N. & Hesselbrock, V.M., 2009. Ethnicity and adolescent pathways to alcohol use. *Journal of studies on alcohol and drugs*, 70(3), pp.337–45.
- Cheah, S.-Y. et al., 2014. BDNF SNPs are implicated in comorbid alcohol dependence in schizophrenia but not in alcohol-dependent patients without schizophrenia. *Alcohol and alcoholism (Oxford, Oxfordshire)*, 49(5), pp.491–7.

- Chen, C.-H. et al., 2014. Variant GADL1 and response to lithium therapy in bipolar I disorder. *The New England journal of medicine*, 370(2), pp.119–28.
- Chen, J. et al., 2004. Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *American journal of human genetics*, 75(5), pp.807–21.
- Chen, X. et al., 2009. PKNOX2 gene is significantly associated with substance dependence in European-origin women. *Proceedings of the National Academy of Sciences of the United States of America*.
- Cheng, Y., Quinn, J.F. & Weiss, L.A., 2013. An eQTL mapping approach reveals that rare variants in the SEMA5A regulatory network impact autism risk. *Human molecular genetics*, 22(14), pp.2960–72.
- Di Chiara, G. & Imperato, A., 1988. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proceedings of the National Academy of Sciences*, 85(14), pp.5274–5278.
- Choi, K.-Y., Yoon, H.-K. & Kim, Y.-K., 2010. Association between Serotonin-Related Polymorphisms in 5HT2A, TPH1, TPH2 Genes and Bipolar Disorder in Korean Population. *Psychiatry investigation*, 7(1), pp.60–7.
- Chopra, K. & Tiwari, V., 2012. Alcoholic neuropathy: possible mechanisms and future treatment possibilities. *British journal of clinical pharmacology*, 73(3), pp.348–62.
- Cichon, S. et al., 2011. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American journal of human genetics*, 88(3), pp.372–81.
- Cichon, S. et al., 1996. Systematic screening for mutations in the 5'-regulatory region of the human dopamine D1 receptor (DRD1) gene in patients with schizophrenia and bipolar affective disorder. *American journal of medical genetics*, 67(4), pp.424–8.
- Cichoz-Lach, H. et al., 2007. Alcohol dehydrogenase and aldehyde dehydrogenase gene polymorphism in alcohol liver cirrhosis and alcohol chronic pancreatitis among Polish individuals. *Scandinavian journal of gastroenterology*, 42(4), pp.493–498.
- Clark, D.M., 2011. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International review of psychiatry (Abingdon, England)*, 23(4), pp.318–27.
- Cloninger, C.R. et al., 1981. Inheritance of Alcohol Abuse. *Archives of General Psychiatry*, 38(8), p.861.
- Conrod, P.J. et al., 2013. Effectiveness of a selective, personality-targeted prevention program for adolescent alcohol use and misuse: a cluster randomized controlled trial. *JAMA psychiatry*, 70(3), pp.334–42.
- Cotton, N.S., 1979. The familial incidence of alcoholism: a review. *Journal of studies on alcohol*, 40(1), pp.89–116.
- Covault, J. et al., 2004. Allelic and haplotypic association of GABRA2 with alcohol dependence. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 129B(1), pp.104–9.
- Crabbe, J.C. et al., 1999. Identifying genes for alcohol and drug sensitivity: recent progress and future directions. *Trends in neurosciences*, 22(4), pp.173–9.
- Craddock, N. et al., 1995. Mathematical limits of multilocus models: the genetic transmission of bipolar disorder. *American journal of human genetics*, 57(3), pp.690–702.
- Craddock, N. & Sklar, P., 2009. Genetics of bipolar disorder: successful start to a long journey. *Trends in genetics : TIG*, 25(2), pp.99–105.
- Crafts, M.W.F., Leitch, M.W. & Leitch, M., 1900. *Intoxicating drinks & drugs in all lands and times: a twentieth century survey of intemperance, based on a symposium of testimony from one hundred missionaries and travelers*, The International reform bureau.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Identification of risk loci with

- shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381(9875), pp.1371–9.
- Crowley, J.J. & Sakamoto, K., 2015. Psychiatric genomics: outlook for 2015 and challenges for 2020. *Current Opinion in Behavioral Sciences*, 2, pp.102–107.
- Cui, W.-Y. et al., 2012. Genetics of GABAergic signaling in nicotine and alcohol dependence. *Human Genetics*, 131(6), pp.843–855.
- Curtis, D. et al., 2011. Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. *Psychiatric genetics*, 21(1), pp.1–4.
- Curtis, D., Vine, A.E. & Knight, J., 2012. A simple method for assessing the strength of evidence for association at the level of the whole gene. *Advances and Applications in Bioinformatics and Chemistry*, 5(1), pp.115–120.
- Cuthbert, B.N. & Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC medicine*, 11, p.126.
- Dahl, J.P. et al., 2005. Confirmation of the association between a polymorphism in the promoter region of the prodynorphin gene and cocaine dependence. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 139B(1), pp.106–8.
- Davis, L. et al., 2008. Major depression and comorbid substance use disorders. *Current opinion in psychiatry*, 21(1), pp.14–18.
- Dedman, A. et al., 2012. Sequencing of the ANKYRIN 3 gene (ANK3) encoding ankyrin G in bipolar disorder reveals a non-conservative amino acid change in a short isoform of ankyrin G. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 159B(3), pp.328–35.
- Department of Health UK, 2016. Alcohol Guidelines Review – Report from the Guidelines development group to the UK Chief Medical Officers. , (January), pp.1–44.
- Dick, D.M. et al., 2008. A Systematic single nucleotide polymorphism screen to fine-map alcohol dependence genes on chromosome 7 identifies association with a novel susceptibility gene ACN9. *Biological psychiatry*, 63(11), pp.1047–53.
- Dick, D.M. et al., 2010. Evidence for genes on chromosome 2 contributing to alcohol dependence with conduct disorder and suicide attempts. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 153B(6), pp.1179–88.
- Dick, D.M. et al., 2006. Marital status, alcohol dependence, and GABRA2: evidence for gene-environment correlation and interaction. *Journal of studies on alcohol*, 67(2), pp.185–194.
- Dick, D.M. et al., 2011. Measures of current alcohol consumption and problems: two independent twin studies suggest a complex genetic architecture. *Alcoholism, clinical and experimental research*, 35(12), pp.2152–61.
- Dick, D.M. et al., 2009. Role of GABRA2 in trajectories of externalizing behavior across development and evidence of moderation by parental monitoring. *Archives of general psychiatry*, 66(6), pp.649–57.
- Dick, D.M. et al., 2002. Suggestive linkage on chromosome 1 for a quantitative alcohol-related phenotype. *Alcoholism, clinical and experimental research*, 26(10), pp.1453–1460.
- Dixon, L., 1999. Dual diagnosis of substance abuse in schizophrenia: prevalence and impact on outcomes. In *Schizophrenia research*. pp. S93–100.
- Djurovic, S. et al., 2010. A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *Journal of affective disorders*, 126(1-2), pp.312–6.
- Dmitrzak-Weglarz, M. et al., 2006. Dopamine receptor D1 gene -48A/G polymorphism is associated with bipolar illness but not with schizophrenia in a Polish population. *Neuropsychobiology*, 53(1),

- pp.46–50.
- Drgon, T., D'Addario, C. & Uhl, G.R., 2006. Linkage disequilibrium, haplotype and association studies of a chromosome 4 GABA receptor gene cluster: candidate gene variants for addictions. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 141B(8), pp.854–60.
- Easton, A.C. et al., 2013. α CaMKII autophosphorylation controls the establishment of alcohol drinking behavior. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 38(9), pp.1636–47.
- Easton, A.C. et al., 2014. α CaMKII controls the establishment of cocaine's reinforcing effects in mice and humans. *Translational psychiatry*, 4, p.e457.
- Edenberg, H.J. et al., 2010a. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism: Clinical and Experimental Research*, 34(5), pp.840–852.
- Edenberg, H.J. et al., 2010b. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism, clinical and experimental research*, 34(5), pp.840–852.
- Edenberg, H.J. et al., 2004. Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *American journal of human genetics*, 74(4), pp.705–14.
- Edwards, G. & Gross, M.M., 1976. Alcohol dependence: provisional description of a clinical syndrome. *British medical journal*, 1(6017), pp.1058–61.
- Egan, M.F. et al., 2004. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), pp.12604–9.
- Ehlers, C.L. & Wilhelmsen, K.C., 2006. Genomic screen for loci associated with tobacco usage in Mission Indians. *BMC medical genetics*, 7, p.9.
- Eichler, E.E. et al., 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics*, 11(6), pp.446–50.
- Elbers, C.C. et al., 2009. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genetic epidemiology*, 33(5), pp.419–31.
- Elkins, I.J. et al., 2006. Personality traits and the development of nicotine, alcohol, and illicit drug disorders: prospective links from adolescence to young adulthood. *Journal of abnormal psychology*, 115(1), pp.26–39.
- Elvidge, G. et al., 2001. Allelic variation of a BalI polymorphism in the DRD3 gene does not influence susceptibility to bipolar disorder: results of analysis and meta-analysis. *American journal of medical genetics*, 105(4), pp.307–11.
- Emerson, G.B. et al., 2010. Testing for the Presence of Positive-Outcome Bias in Peer Review. *Archives of Internal Medicine*, 170(21), pp.254–257.
- Enoch, M.A. et al., 2010. The Influence of GABRA2, Childhood Trauma, and Their Interaction on Alcohol, Heroin, and Cocaine Dependence. *Biological Psychiatry*, 67(1), pp.20–27.
- Esslinger, C. et al., 2009. Neural mechanisms of a genome-wide supported psychosis variant. *Science (New York, N.Y.)*, 324(5927), p.605.
- Fagnani, C. et al., 2011. Investigation of shared genetic effects for psychotic and obsessive symptoms in young adult twins. *Psychiatry research*, 188(2), pp.276–82.
- Fan, J.-B. et al., 2005. Catechol-O-methyltransferase gene Val/Met functional polymorphism and risk of schizophrenia: a large-scale association study plus meta-analysis. *Biological psychiatry*, 57(2), pp.139–44.
- Fanous, A.H. et al., 2012. Genome-wide association study of clinical dimensions of schizophrenia: polygenic effect on disorganized symptoms. *The American journal of psychiatry*, 169(12), pp.1309–17.
- Fatemi, S.H. et al., 2013. Expression of GABAA α 2-, β 1- and ϵ -receptors are altered significantly in the lateral cerebellum of subjects with schizophrenia, major depression and bipolar disorder.

- Translational psychiatry*, 3, p.e303.
- Ferreira, M. a R. et al., 2008. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature genetics*, 40(9), pp.1056–1058.
- Fitt, E., Pell, D. & Cole, D., 2013. Assessing caffeine intake in the United Kingdom diet. *Food Chemistry*, 140(3), pp.421–426.
- Flister, M.J. et al., 2013. Identifying multiple causative genes at a single GWAS locus. *Genome research*, 23(12), pp.1996–2002.
- Flory, J.D. et al., 2011. Alcohol dependence, disinhibited behavior and variation in the prodynorphin gene. *Biological Psychology*, 88(1), pp.51–56.
- Flory, J.D. et al., 2011. Alcohol dependence, disinhibited behavior and variation in the prodynorphin gene. *Biological psychology*, 88(1), pp.51–6.
- Forero, D.A. et al., 2015. Meta-analysis of six genes (BDNF, DRD1, DRD3, DRD4, GRIN2B and MAOA) involved in neuroplasticity and the risk for alcohol dependence. *Drug and alcohol dependence*, 149, pp.259–63.
- Foroud, T. et al., 2007. Association of alcohol craving with alpha-synuclein (SNCA). *Alcoholism, clinical and experimental research*, 31(4), pp.537–545.
- Foroud, T. et al., 2000. Identification of quantitative trait loci influencing alcohol consumption in the high alcohol drinking and low alcohol drinking rat lines. *Behavior Genetics*, 30(2), pp.131–140.
- Francès, F. et al., 2015. Association between Opioid Receptor mu 1 (OPRM1) Gene Polymorphisms and Tobacco and Alcohol Consumption in a Spanish Population. *Bosnian journal of basic medical sciences / Udruženje osnovnih medicinskih znanosti = Association of Basic Medical Sciences*, 15(2), pp.31–6.
- Frank, J. et al., 2012. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addiction biology*, 17(1), pp.171–80.
- Frazer, K.A. et al., 2009. Human genetic variation and its contribution to complex traits. *Nature reviews. Genetics*, 10(4), pp.241–51.
- Fromme, K. et al., 2004. Biological and behavioral markers of alcohol sensitivity. *Alcoholism, clinical and experimental research*, 28(2), pp.247–56.
- Frye, M.A. et al., 2015. Development of a bipolar disorder biobank: differential phenotyping for subsequent biomarker analyses. *International journal of bipolar disorders*, 3(1), p.30.
- Furlong, R.A. et al., 1998. No association of a functional polymorphism in the dopamine D2 receptor promoter region with bipolar or unipolar affective disorders. *American journal of medical genetics*, 81(5), pp.385–7.
- Gao, B. & Bataller, R., 2011. Alcoholic liver disease: pathogenesis and new therapeutic targets. *Gastroenterology*, 141(5), pp.1572–85.
- Gao, J. et al., 2016. TPH2 gene polymorphisms and bipolar disorder: A meta-analysis. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171(2), pp.145–152.
- Gao, X. et al., 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic epidemiology*, 34(1), pp.100–5.
- Garner, C. et al., 2001. Linkage analysis of a complex pedigree with severe bipolar disorder, using a Markov chain Monte Carlo method. *American journal of human genetics*, 68(4), pp.1061–4.
- Gejman, P. V, Sanders, A.R. & Duan, J., 2010. The role of genetics in the etiology of schizophrenia. *The Psychiatric clinics of North America*, 33(1), pp.35–66.
- Gelernter, J. et al., 2014. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Molecular psychiatry*, 19(1), pp.41–9.
- Gelernter, J. & Kranzler, H., 1999. D2 dopamine receptor gene (DRD2) allele and haplotype frequencies in alcohol dependent and control subjects: no association with phenotype or severity of phenotype. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 20(6), pp.640–9.
- George, A.W., 2013. Controlling type 1 error rates in genome-wide association studies in plants.

- Heredity*, 111(1), pp.86–7.
- Georgiades, K. et al., 2011. Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. *PLoS one*, 6(9), p.e24857.
- Gershon, E.S., 1982. A Family Study of Schizoaffective, Bipolar I, Bipolar II, Unipolar, and Normal Control Proband. *Archives of General Psychiatry*, 39(10), p.1157.
- Gershon, E.S. et al., 1975. Transmitted factors in the morbid risk of affective disorders: A controlled study. *Journal of Psychiatric Research*, 12(4), pp.283–299.
- Girard, S.L. et al., 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics*, 43(9), pp.860–3.
- Girault, J. & Greengard, P., 2004. The neurobiology of dopamine signaling. *Archives of neurology*.
- Gizer, I.R. et al., 2011. Linkage scan of alcohol dependence in the UCSF Family Alcoholism Study. *Drug and alcohol dependence*, 113(2-3), pp.125–32.
- Glatt, S.J., Faraone, S. V & Tsuang, M.T., 2003. Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: meta-analysis of case-control and family-based studies. *The American journal of psychiatry*, 160(3), pp.469–76.
- Goghari, V.M. & Sponheim, S.R., 2008. Differential association of the COMT Val158Met polymorphism with clinical phenotypes in schizophrenia and bipolar disorder. *Schizophrenia research*, 103(1-3), pp.186–91.
- Goodwin, D.W. et al., 1973. Alcohol problems in adoptees raised apart from alcoholic biological parents. *Archives of general psychiatry*, 28(2), pp.238–43.
- Goodwin, D.W., 1974. Drinking Problems in Adopted and Nonadopted Sons of Alcoholics. *Archives of General Psychiatry*, 31(2), p.164.
- Gorwood, P. et al., 2000. Reappraisal of the association between the DRD2 gene, alcoholism and addiction. *European Psychiatry*, 15(2), pp.90–96.
- Gothelf, D. et al., 1997. Velocardiofacial manifestations and microdeletions in schizophrenic inpatients. *American journal of medical genetics*, 72(4), pp.455–61.
- Gottesman, I.I. et al., 2010. Severe mental disorders in offspring with 2 psychiatrically ill parents. *Archives of general psychiatry*, 67(3), pp.252–7.
- Gottesman, I.I. & Shields, J., 1967. A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1), pp.199–205.
- Gottesman, I.I. & Shields, J., 1973. Genetic theorizing and schizophrenia. *The British journal of psychiatry: the journal of mental science*, 122(566), pp.15–30.
- Grant, B.F. & Dawson, D.A., 1997. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of substance abuse*, 9, pp.103–10.
- Green, E.K. et al., 2013. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Molecular psychiatry*, 18(12), pp.1302–7.
- Greenson, A.J., 1983. The “Holiday Heart”: Electrophysiologic Studies of Alcohol Effects in Alcoholics. *Annals of Internal Medicine*, 98(2), p.135.
- Greenwood, T.A. et al., 2013. Heritability and genome-wide SNP linkage analysis of temperament in bipolar disorder. *Journal of affective disorders*, 150(3), pp.1031–40.
- Guerrini, I. et al., 2005. Genetic linkage analysis supports the presence of two susceptibility loci for alcoholism and heavy drinking on chromosome 1p22.1-11.2 and 1q21.3-24.2. *BMC Genet*, 6(1), p.11.
- Gulacsi, A.A. & Anderson, S.A., 2008. Beta-catenin-mediated Wnt signaling regulates neurogenesis in the ventral telencephalon. *Nature neuroscience*, 11(12), pp.1383–91.
- Gulsuner, S. et al., 2013. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*, 154(3), pp.518–29.

- Guo, R. & Ren, J., 2010. Alcohol and acetaldehyde in public health: from marvel to menace. *International journal of environmental research and public health*, 7(4), pp.1285–301.
- Guo, X. et al., 2012. Large scale association analysis for drug addiction: results from SNP to gene. *TheScientificWorldJournal*, 2012, p.939584.
- Gurling, H., 1986. Candidate genes and favoured loci: strategies for molecular genetic research into schizophrenia, manic depression, autism, alcoholism and Alzheimer's disease. *Psychiatr Dev*, 4(4), pp.289–309.
- Hamilton, S.P. et al., 2002. Evidence for a susceptibility locus for panic disorder near the catechol-O-methyltransferase gene on chromosome 22. *Biological psychiatry*, 51(7), pp.591–601.
- Hamshere, M.L. et al., 2013. Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Molecular psychiatry*, 18(6), pp.708–12.
- Han, S. et al., 2013. Ordered subset linkage analysis based on admixture proportion identifies new linkage evidence for alcohol dependence in African-Americans. *Human genetics*, 132(4), pp.397–403.
- Hansell, N.K. et al., 2008. Long-term stability and heritability of telephone interview measures of alcohol consumption and dependence. *Twin Res Hum Genet*, 11(3), pp.287–305.
- Harnic, D. et al., [Temperament and attachment in alcohol addicted patients of type 1 and 2]. *Rivista di psichiatria*, 45(5), pp.311–9.
- Harrison, P.J. & Weinberger, D.R., 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular psychiatry*, 10(1), pp.40–68; image 5.
- Hasin, D. et al., 2006. Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and International Classification of Diseases, tenth edition (ICD-10). *Addiction (Abingdon, England)*, 101 Suppl , pp.59–75.
- Hayden, E.P. & Nurnberger, J.I., 2006. Molecular genetics of bipolar disorder. *Genes, brain, and behavior*, 5(1), pp.85–95.
- Hayes, D.J. & Greenshaw, A.J., 2011. 5-HT receptors and reward-related behaviour: a review. *Neuroscience and biobehavioral reviews*, 35(6), pp.1419–49.
- Health and Social Care Information Centre, 2014. *Statistics on Alcohol - England, 2014*.
- Heard, E. & Martienssen, R.A., 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, 157(1), pp.95–109.
- Heath, A.C. et al., 2012. Gene–Environment Interaction Effects on Behavioral Variation and Risk of Complex Disorders: The Example of Alcoholism and Other Psychiatric Disorders. *Twin Research*, 5(01), pp.30–37.
- Heath, A.C. et al., 1997. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychological medicine*, 27(6), pp.1381–96.
- Heath, K.E. et al., 1999. The type of mutation in the low density lipoprotein receptor gene influences the cholesterol-lowering response of the HMG-CoA reductase inhibitor simvastatin in patients with heterozygous familial hypercholesterolaemia. *Atherosclerosis*, 143(1), pp.41–54.
- Heinz, A. et al., 2000. A relationship between serotonin transporter genotype and in vivo protein expression and alcohol neurotoxicity. *Biological psychiatry*, 47(7), pp.643–9.
- Hesselbrock, M.N. et al., 2003. Ethnicity and psychiatric comorbidity among alcohol-dependent persons who receive inpatient treatment: African Americans, Alaska natives, Caucasians, and Hispanics. *Alcoholism, clinical and experimental research*, 27(8), pp.1368–73.
- Heston, L.L., 1966. Psychiatric Disorders in Foster Home Reared Children of Schizophrenic Mothers. *The British Journal of Psychiatry*, 112(489), pp.819–825.
- Hettema, J.M. et al., 2008. Catechol-O-methyltransferase contributes to genetic susceptibility shared among anxiety spectrum phenotypes. *Biological psychiatry*, 64(4), pp.302–10.
- Heutink, P., 2002. Gene finding in genetically isolated populations. *Human Molecular Genetics*, 11(20),

- pp.2507–2515.
- Hicks, B.M. et al., 2013. Gene-environment correlation in the development of adolescent substance abuse: selection effects of child personality and mediation via contextual risk factors. *Development and psychopathology*, 25(1), pp.119–32.
- Hicks, B.M., Iacono, W.G. & McGue, M., 2010. Consequences of an adolescent onset and persistent course of alcohol dependence in men: adolescent risk factors and adult outcomes. *Alcoholism, clinical and experimental research*, 34(5), pp.819–33.
- Hill, S.Y. et al., 2004. A genome wide search for alcoholism susceptibility genes. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 128B(1), pp.102–113.
- Hines, L.M. et al., 2005. Alcoholism: the dissection for endophenotypes. *Dialogues in clinical neuroscience*, 7(2), pp.153–63.
- Hofvander, B. et al., 2009. Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC psychiatry*, 9, p.35.
- Holmans, P. et al., 2009. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *American journal of human genetics*, 85(1), pp.13–24.
- Holmes, M. V et al., 2014. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ (Clinical research ed.)*, 349(jul10_6), p.g4164.
- Hong, J.P. et al., 2008. New functional single nucleotide polymorphism (Ala72Ser) in the COMT gene is associated with aggressive behavior in male schizophrenia. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 147B(5), pp.658–60.
- Hu, T. et al., 2011. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC bioinformatics*, 12(1), p.364.
- Hutchison, K.E. et al., 2003. Olanzapine reduces craving for alcohol: a DRD4 VNTR polymorphism by pharmacotherapy interaction. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 28(10), pp.1882–8.
- Hutchison, K.E. et al., 2002. The DRD4 VNTR polymorphism moderates craving after alcohol consumption. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, 21(2), pp.139–46.
- Hyman, S.E. & Malenka, R.C., 2001. Addiction and the brain: The neurobiology of compulsion and its persistence. *Nature Reviews Neuroscience*, 2(10), pp.695–703.
- Iacono, W.G., Malone, S.M. & McGue, M., 2008. Behavioral disinhibition and the development of early-onset addiction: common and specific influences. *Annual review of clinical psychology*, 4, pp.325–48.
- Iatropoulos, P. et al., 2009. Association study and mutational screening of SYNGR1 as a candidate susceptibility gene for schizophrenia. *Psychiatric genetics*, 19(5), pp.237–43.
- International Schizophrenia Consortium., 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455(7210), pp.237–41.
- Irvine, E.E. et al., 2006. alphaCaMKII autophosphorylation: a fast track to memory. *Trends in neurosciences*, 29(8), pp.459–65.
- Irwin, M., Schuckit, M. & Smith, T.L., 1990. Clinical importance of age at onset in type 1 and type 2 primary alcoholics. *Archives of general psychiatry*, 47(4), pp.320–4.
- Ishiguro, H. et al., 1998. Association study between the -141C Ins/Del and TaqI A polymorphisms of the dopamine D2 receptor gene and alcoholism. *Alcoholism, clinical and experimental research*, 22(4), pp.845–8.
- Iwamoto, K. et al., 2004. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Molecular psychiatry*, 9(4), pp.406–16.
- Jellinek, E., 1960. *The disease concept of alcoholism.*, New Haven: Hillhouse Press.
- Jin, L. et al., 2014. Pathway-based analysis tools for complex diseases: a review. *Genomics, proteomics*

- & *bioinformatics*, 12(5), pp.210–20.
- Johnson, B.A. et al., 2013. Determination of genotype combinations that can predict the outcome of the treatment of alcohol dependence using the 5-HT(3) antagonist ondansetron. *The American journal of psychiatry*, 170(9), pp.1020–31.
- Johnson, C. et al., 2006. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. *Am J Med Genet B Neuropsychiatr Genet*, 141B(8), pp.844–853.
- Johnson, E.O., van den Bree, M.B. & Pickens, R.W., 1996. Subtypes of alcohol-dependent men: a typology based on relative genetic and environmental loading. *Alcoholism, clinical and experimental research*, 20(8), pp.1472–80.
- Johnson, R.C. et al., 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC genomics*, 11, p.724.
- Johnstone, M. et al., 2015. Copy Number Variations in DISC1 and DISC1-Interacting Partners in Major Mental Illness. *Molecular Neuropsychiatry*, 1(3), pp.175–190.
- Kandaswamy, R. et al., 2013. Genetic Association, Mutation Screening, and Functional Analysis of a Kozak Sequence Variant in the Metabotropic Glutamate Receptor 3 Gene in Bipolar Disorder. *JAMA Psychiatry*, 70(6), p.591.
- Kato, T., 2007. Molecular genetics of bipolar disorder and depression. *Psychiatry and Clinical Neurosciences*, 61(1), pp.3–19.
- Katz, S.H. & Voigt, M.M., 1986. Bread and Beer: The Early Use of Cereals in the Human Diet. *Expedition*, 2, pp.23–34.
- Kauhanen, J. et al., 2000. Association between the functional polymorphism of catechol-O-methyltransferase gene and alcohol consumption among social drinkers. *Alcoholism, clinical and experimental research*, 24(2), pp.135–139.
- Kendler, K.S., 2008. Explanatory models for psychiatric illness. *The American journal of psychiatry*, 165(6), pp.695–702.
- Kendler, K.S. et al., 2008. Genetic and Environmental Influences on Alcohol, Caffeine, Cannabis, and Nicotine Use From Early Adolescence to Middle Adulthood. *Archives of General Psychiatry*, 65(6), p.674.
- Kendler, K.S. et al., 2011. Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. *Alcoholism, clinical and experimental research*, 35(5), pp.963–75.
- Kendler, K.S. et al., 1992. Major depression and generalized anxiety disorder. Same genes, (partly) different environments? *Archives of general psychiatry*, 49(9), pp.716–22.
- Kendler, K.S. et al., 1997. Temperance board registration for alcohol abuse in a national sample of Swedish male twins, born 1902 to 1949. *Arch Gen Psychiatry*, 54(2), pp.178–184.
- Kendler, K.S. et al., 1995. The structure of the genetic and environmental risk factors for six major psychiatric disorders in women. Phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of general psychiatry*, 52(5), pp.374–83.
- Kendler, K.S., Gardner, C. & Dick, D.M., 2011. Predicting alcohol consumption in adolescence from alcohol-specific and general externalizing genetic risk factors, key environmental exposures and their interaction. *Psychological medicine*, 41(7), pp.1507–16.
- Kendler, K.S., Gruenberg, A.M. & Kinney, D.K., 1994. Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish Adoption Study of Schizophrenia. *Archives of general psychiatry*, 51(6), pp.456–68.
- Kessler, R.C. et al., 1997. Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. *Archives of General Psychiatry*, 54(4), p.313.
- Kety, S.S. et al., 1994. Mental illness in the biological and adoptive relatives of schizophrenic adoptees.

- Replication of the Copenhagen Study in the rest of Denmark. *Archives of general psychiatry*, 51(6), pp.442–55.
- Khazaal, Y. et al., 2013. Affective temperaments in alcohol and opiate addictions. *The Psychiatric quarterly*, 84(4), pp.429–38.
- Kieseppä, T. et al., 2005. Memory and verbal learning functions in twins with bipolar-I disorder, and the role of information-processing speed. *Psychological Medicine*, 35(2), pp.205–215.
- Koob, G. & Bloom, F., 1988. Cellular and molecular mechanisms of drug dependence. *Science*, 242(4879), pp.715–723.
- Koob, G.F., 1997. Drug Abuse: Hedonic Homeostatic Dysregulation. *Science*, 278(5335), pp.52–58.
- Koob, G.F. & Volkow, N.D., 2010. Neurocircuitry of addiction. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 35(1), pp.217–38.
- Kosobud, A.E.K. et al., 2015. Adaptation of Subjective Responses to Alcohol is Affected by an Interaction of GABRA2 Genotype and Recent Drinking. *Alcoholism, clinical and experimental research*, 39(7), pp.1148–57.
- Kraepelin, E., 1921. *Manic-Depressive Insanity and Paranoia.*, Salem N.H.: Ayer Company, Publishers, Inc..
- Kraft, P. & Hunter, D.J., 2009. Genetic risk prediction--are we there yet? *The New England journal of medicine*, 360(17), pp.1701–3.
- Kremer, I. et al., 2003. Family-based and case-control study of catechol-O-methyltransferase in schizophrenia among Palestinian Arabs. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 119B(1), pp.35–9.
- Krenz, M. & Korthuis, R.J., 2012. Moderate ethanol ingestion and cardiovascular protection: From epidemiologic associations to cellular mechanisms. *Journal of Molecular and Cellular Cardiology*, 52(1), pp.93–104.
- Kuo, P.H. et al., 2006. The temporal relationship of the onsets of alcohol dependence and major depression: using a genetically informative study design. *Psychol Med*, 36(8), pp.1153–1162.
- de la Vega, M.J. et al., 2001. High prevalence of hyperhomocysteinemia in chronic alcoholism: the importance of the thermolabile form of the enzyme methylenetetrahydrofolate reductase (MTHFR). *Alcohol (Fayetteville, N.Y.)*, 25(2), pp.59–67.
- Laird, N.M., 2001. Family-based Association Test (FBAT). In *eLS*. John Wiley & Sons, Ltd.
- Lappalainen, J. et al., 2004. Confirmation and fine mapping of the chromosome 1 alcohol dependence risk locus. *Mol Psychiatry*, 9(3), pp.312–319.
- Laursen, T.M. et al., 2005. Family history of psychiatric illness as a risk factor for schizoaffective disorder: a Danish register-based cohort study. *Archives of general psychiatry*, 62(8), pp.841–8.
- Lee, P.H. et al., 2012. Multi-locus genome-wide association analysis supports the role of glutamatergic synaptic transmission in the etiology of major depressive disorder. *Translational psychiatry*, 2, p.e184.
- Lee, S. et al., 2014. Rare-Variant Association Analysis: Study Designs and Statistical Tests. *American journal of human genetics*, 95(1), pp.5–23.
- Lee, S.-G. et al., 2005. Association of Ala72Ser polymorphism with COMT enzyme activity and the risk of schizophrenia in Koreans. *Human genetics*, 116(4), pp.319–28.
- Levander, E. et al., 2007. Alcoholism and anxiety in bipolar illness: Differential lifetime anxiety comorbidity in bipolar I women with and without alcoholism. *Journal of Affective Disorders*, 101(1-3), pp.211–217.
- Levey, D.F. et al., 2014. Genetic risk prediction and neurobiological understanding of alcoholism. *Translational psychiatry*, 4(5), p.e391.
- Levinson, D.F. et al., 2000. Multicenter linkage study of schizophrenia candidate regions on chromosomes 5q, 6q, 10p, and 13q: schizophrenia linkage collaborative group III. *American journal of human genetics*, 67(3), pp.652–63.

- Levy, S. et al., 2007. The diploid genome sequence of an individual human. *PLoS biology*, 5(10), p.e254.
- Lewis, C.M. & Knight, J., 2012. Introduction to Genetic Association Studies. *Cold Spring Harbor Protocols*, 2012(3), p.pdb.top068163–pdb.top068163.
- Li, D. et al., 2014. Association of gamma-aminobutyric acid A receptor $\alpha 2$ gene (GABRA2) with alcohol use disorder. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 39(4), pp.907–918.
- Li, M., Li, C. & Guan, W., 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. *European journal of human genetics : EJHG*, 16(5), pp.635–43.
- Li, M.-X. et al., 2011. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *American journal of human genetics*, 88(3), pp.283–93.
- Li, T. et al., 1999. Association analysis between dopamine receptor genes and bipolar affective disorder. *Psychiatry research*, 86(3), pp.193–201.
- Li, T. et al., 1996. Preferential transmission of the high activity allele of COMT in schizophrenia. *Psychiatric genetics*, 6(3), pp.131–3.
- Li, T.K., 2000. Pharmacogenetics of responses to alcohol and genes that influence alcohol drinking. *Journal of studies on alcohol*, 61(1), pp.5–12.
- Li, W. et al., 2015. Significant association of GRM7 and GRM8 genes with schizophrenia and major depressive disorder in the Han Chinese population. *European Neuropsychopharmacology*.
- Lichtenstein, P. et al., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, 373(9659), pp.234–9.
- Limosin, F. et al., 2003. Association Between Dopamine Receptor D1 Gene *Dde I* polymorphism and Sensation Seeking in Alcohol-Dependent Men. *Alcoholism: Clinical and Experimental Research*, 27(8), pp.1226–1228.
- Limosin, F. et al., 2005. Association between dopamine receptor D3 gene *BaII* polymorphism and cognitive impulsiveness in alcohol-dependent men. *European psychiatry : the journal of the Association of European Psychiatrists*, 20(3), pp.304–6.
- Lin, C.C.H. et al., 2007. Genetic and environmental influences on schizotypy among adolescents in Taiwan: a multivariate twin/sibling analysis. *Behavior genetics*, 37(2), pp.334–44.
- Lin, M.W. et al., 1995. Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. *Psychiatric genetics*, 5(3), pp.117–26.
- Lin, P. et al., 2012. Copy number variations in 6q14.1 and 5q13.2 are associated with alcohol dependence. *Alcoholism, clinical and experimental research*, 36(9), pp.1512–8.
- Lind, P.A. et al., 2010. A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations. *Twin Res Hum Genet*, 13(1), pp.10–29.
- Linney, Y.M. et al., 2003. A quantitative genetic analysis of schizotypal personality traits. *Psychological medicine*, 33(5), pp.803–16.
- Litt, M. & Luty, J.A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American journal of human genetics*, 44(3), pp.397–401.
- Liu, J. et al., 2006. Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 31(7), pp.1574–82.
- Liu, J.Z. et al., 2010. A versatile gene-based test for genome-wide association studies. *American journal of human genetics*, 87(1), pp.139–145.
- Long, J.C. & Goldman, D., 1998. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an american indian population. *American Journal of Medical Genetics*, 81(3), pp.216–221.
- López León, S. et al., 2005. The dopamine D4 receptor gene 48-base-pair-repeat polymorphism and mood disorders: a meta-analysis. *Biological psychiatry*, 57(9), pp.999–1003.

- Lovinger, D.M. & Crabbe, J.C., 2005. Laboratory models of alcoholism: treatment target identification and insight into mechanisms. *Nature neuroscience*, 8(11), pp.1471–80.
- Lu, L., 2015. The impact of genetic variation on sensitivity to opioid analgesics in patients with postoperative pain: a systematic review and meta-analysis. *Pain physician*.
- Luo, X. et al., 2005. CHRM2 gene predisposes to alcohol dependence, drug dependence and affective disorders: results from an extended case-control structured association study. *Human molecular genetics*, 14(16), pp.2421–34.
- Lusher, J.M., Chandler, C. & Ball, D., 2001. Dopamine D4 receptor gene (DRD4) is associated with Novelty Seeking (NS) and substance abuse: the saga continues... *Molecular psychiatry*, 6(5), pp.497–9.
- Lydall, G.J. et al., 2011. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. *Psychiatric genetics*, 21(6), pp.294–306.
- Lydall, G.J. et al., 2011. Genetic association study of GABRA2 single nucleotide polymorphisms and electroencephalography in alcohol dependence. *Neuroscience letters*, 500(3), pp.162–6.
- MacCoun, R.J., 1998. BIASES IN THE INTERPRETATION AND USE OF RESEARCH RESULTS. *Annual Review of Psychology*, 49(1), pp.259–287.
- Macgregor, D.R. et al., 2008. Root system architecture in Arabidopsis grown in culture is regulated by sucrose uptake in the aerial tissues. *The Plant cell*, 20(10), pp.2643–60.
- Maier, W. et al., 1993. Continuity and discontinuity of affective disorders and schizophrenia. Results of a controlled family study. *Archives of general psychiatry*, 50(11), pp.871–83.
- Mandelli, L. et al., 2007. Interaction between serotonin transporter gene, catechol-O-methyltransferase gene and stressful life events in mood disorders. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, 10(4), pp.437–47.
- Mann, J.J. et al., 2014. Anxiety in major depression and cerebrospinal fluid free gamma-aminobutyric acid. *Depression and Anxiety*, 31(10), pp.814–821.
- Manzo-Avalos, S. & Saavedra-Molina, A., 2010. Cellular and mitochondrial effects of alcohol consumption. *International journal of environmental research and public health*, 7(12), pp.4281–304.
- Maremmani, I. et al., 2009. Affective temperaments in heroin addiction. *Journal of affective disorders*, 117(3), pp.186–92.
- Maremmani, I. et al., 2005. The relationship of Kraepelian affective temperaments (as measured by TEMPS-I) to the tridimensional personality questionnaire (TPQ). *Journal of affective disorders*, 85(1-2), pp.17–27.
- Martin, N.G., Perl, J., et al., 1985. A twin study of ethanol metabolism. *Behavior genetics*, 15(2), pp.93–109.
- Martin, N.G., Oakeshott, J.G., et al., 1985. A twin study of psychomotor and physiological responses to an acute dose of alcohol. *Behavior genetics*, 15(4), pp.305–47.
- Massat, I. et al., 2005. Association between COMT (Val158Met) functional polymorphism and early onset in patients with major depressive disorder in a European multicenter genetic association study. *Molecular psychiatry*, 10(6), pp.598–605.
- Matsuzaki, H. et al., 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome research*, 14(3), pp.414–25.
- McBride, W.J. & Li, T.K., 1998. Animal models of alcoholism: neurobiology of high alcohol-drinking behavior in rodents. *Critical reviews in neurobiology*, 12(4), pp.339–69.
- McCarthy, M.I. et al., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews. Genetics*, 9(5), pp.356–69.
- McClellan, J. & King, M.-C., 2010. Genetic heterogeneity in human disease. *Cell*, 141(2), pp.210–7.
- McGovern, P.E., 2009. *Uncorking the Past: The Quest for Wine, Beer, and Other Alcoholic Beverages*,

- University of California Press.
- McGue, M., 1999. The Behavioral Genetics of Alcoholism. *Current Directions in Psychological Science*, 8(4), pp.109–115.
- McGue, M., Pickens, R.W. & Svikis, D.S., 1992. Sex and age effects on the inheritance of alcohol problems: a twin study. *J Abnorm Psychol*, 101(1), pp.3–17.
- McGuffin, P. et al., 2003. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of general psychiatry*, 60(5), pp.497–502.
- McQueen, M.B. et al., 2005. Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *American journal of human genetics*, 77(4), pp.582–595.
- McQuillin, A. et al., 2011. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *European journal of human genetics : EJHG*, 19(5), pp.588–92.
- Moses, M.J. Genetic refinement and physical mapping of a 2.3 Mb probable disease region associated with a bipolar affective disorder susceptibility locus on chromosome 4q35. *Am Genet Genet*, 117(1), pp.23–32.
- Miles, J.H. et al., 2003. Autism Families with a High Incidence of Alcoholism. *Journal of Autism and Developmental Disorders*, 33(4), pp.403–415.
- Miñana, R. et al., 2002. Alcohol Exposure Alters the Expression Pattern of Neural Cell Adhesion Molecules During Brain Development. *Journal of Neurochemistry*, 75(3), pp.954–964.
- Mirnic, K. et al., 2001. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends in Neurosciences*, 24(8), pp.479–486.
- Moore, D.J. et al., 2005. Temperament and risky behaviors: a pathway to HIV? *Journal of affective disorders*, 85(1-2), pp.191–200.
- Moreau, M.P. et al., 2011. Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. *Biological psychiatry*, 69(2), pp.188–93.
- Mortensen, P.B. et al., 2003. Individual and familial risk factors for bipolar affective disorders in Denmark. *Archives of general psychiatry*, 60(12), pp.1209–15.
- Mortensen, P.B., Pedersen, M.G. & Pedersen, C.B., 2010. Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? *Psychological medicine*, 40(2), pp.201–10.
- Morton, N.E., 1955. Sequential tests for the detection of linkage. *American journal of human genetics*, 7(3), pp.277–318.
- Moskvina, V. et al., 2005. Design of case-controls studies with unscreened controls. *Annals of human genetics*, 69(Pt 5), pp.566–76.
- Moskvina, V. et al., 2009. Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry*, 14(3), pp.252–60.
- Mottagui-Tabar, S. et al., 2005. A Novel Single Nucleotide Polymorphism of the Neuropeptide Y (NPY) Gene Associated With Alcohol Dependence. *Alcoholism, clinical and experimental research*, 29(5), pp.702–707.
- Mulligan, M.K. et al., 2006. Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), pp.6368–73.
- Murphy, D.L. et al., 1971. Regular induction of hypomania by L-dopa in “bipolar” manic-depressive patients. *Nature*, 229(5280), pp.135–6.
- Murphy, J.M. et al., 2002. Phenotypic and genotypic characterization of the Indiana University rat lines selectively bred for high and low alcohol preference. *Behavior genetics*, 32(5), pp.363–88.
- Narendran, R. et al., 2014. Decreased prefrontal cortical dopamine transmission in alcoholism. *The American journal of psychiatry*, 171(8), pp.881–8.
- Neale, B.M. & Sham, P.C., 2004. The Future of Association Studies: Gene-Based Analysis and

- Replication. *The American Journal of Human Genetics*, 75(3), pp.353–362.
- Need, A.C. et al., 2012. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *American journal of human genetics*, 91(2), pp.303–12.
- Nery, F.G. et al., 2010. Bipolar disorder comorbid with alcoholism: a 1H magnetic resonance spectroscopy study. *Journal of psychiatric research*, 44(5), pp.278–285.
- Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium, 2015. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature neuroscience*, 18(2), pp.199–209.
- Newlin, D.B. & Thomson, J.B., 1999. Chronic tolerance and sensitization to alcohol in sons of alcoholics: II. Replication and reanalysis. *Experimental and clinical psychopharmacology*, 7(3), pp.234–43.
- Ngwa, J.S. et al., 2011. Pathway analysis following association study. *BMC proceedings*, 5 Suppl 9(Suppl 9), p.S18.
- Nitsche, C. et al., 2011. Environmental risk factors for chronic pancreatitis and pancreatic cancer. *Digestive diseases (Basel, Switzerland)*, 29(2), pp.235–42.
- Noble, E.P. et al., 1998. D2 dopamine receptor and GABA(A) receptor beta3 subunit genes and alcoholism. *Psychiatry research*, 81(2), pp.133–47.
- Noble, E.P. et al., 2000. Haplotypes at the DRD2 locus and severe alcoholism. *American journal of medical genetics*, 96(5), pp.622–31.
- Nomura, A. et al., 2006. Genetic variant of prodynorphin gene is risk factor for methamphetamine dependence. *Neuroscience letters*, 400(1-2), pp.158–62.
- Nöthen, M.M. et al., 1992. Lack of association between dopamine D1 and D2 receptor genes and bipolar affective disorder. *The American journal of psychiatry*, 149(2), pp.199–201.
- Nurnberger Jr., J.I. et al., 2001. Evidence for a Locus on Chromosome 1 That Influences Vulnerability to Alcoholism and Affective Disorder. *Am J Psychiatry*, 158(5), pp.718–724.
- Nurnberger, J.I. et al., 2004. A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. *Archives of general psychiatry*, 61(12), pp.1246–1256.
- Nyegaard, M. et al., 2010. Support of association between BRD1 and both schizophrenia and bipolar affective disorder. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 153B(2), pp.582–91.
- O'Donovan, M.C. et al., 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature genetics*, 40(9), pp.1053–5.
- O'Rourke, D.H. et al., 1982. Refutation of the general single-locus model for the etiology of schizophrenia. *American journal of human genetics*, 34(4), pp.630–49.
- Ohannessian, C.M. & Hesselbrock, V.M., 2008. A comparison of three vulnerability models for the onset of substance use in a high-risk sample. *Journal of studies on alcohol and drugs*, 69(1), pp.75–84.
- Okubo, T. et al., 2003. Association analyses between polymorphisms of the phase II detoxification enzymes (GSTM1, NQO1, NQO2) and alcohol withdrawal symptoms. *Alcohol Clin Exp Res*, 27(8 Suppl), p.68S–71S.
- Omidvar, M. et al., 2009. The effect of catechol-O-methyltransferase Met/Val functional polymorphism on smoking cessation: retrospective and prospective analyses in a cohort study. *Pharmacogenetics and genomics*, 19(1), pp.45–51.
- Oosterhuis, B.E. et al., 2008. Catechol-O-methyltransferase (COMT) gene variants: possible association of the Val158Met variant with opiate addiction in Hispanic women. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 147B(6), pp.793–8.
- Oquendo, M.A. et al., 2010. Increased risk for suicidal behavior in comorbid bipolar disorder and alcohol use disorders: results from the National Epidemiologic Survey on Alcohol and Related

- Conditions (NESARC). *The Journal of clinical psychiatry*, 71(7), pp.902–9.
- van Os, J. & Kapur, S., 2009. Schizophrenia. *Lancet*, 374(9690), pp.635–645.
- Ott, J., Wang, J. & Leal, S.M., 2015. Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics*, 16(5), pp.275–84.
- Owen, M.J., 2011. Is there a schizophrenia to diagnose? *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 10(1), pp.34–5.
- Page, G.P. et al., 2003. “Are We There Yet?”: Deciding When One Has Demonstrated Specific Genetic Causation in Complex Diseases and Quantitative Traits. *The American Journal of Human Genetics*, 73(4), pp.711–719.
- Panagiotou, O.A. et al., 2013. The power of meta-analysis in genome-wide association studies. *Annual review of genomics and human genetics*, 14, pp.441–65.
- Pannucci, C.J. & Wilkins, E.G., 2010. Identifying and avoiding bias in research. *Plastic and reconstructive surgery*, 126(2), pp.619–25.
- Park, B.L. et al., 2013. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human genetics*, 132(6), pp.657–68.
- Parsian, A. & Cloninger, C.R., 1995. Genetic components of alcoholism and subtypes of alcoholism. In *Biological aspects of alcoholism*. pp. pp. 63–83.
- Pearson, T.A. & Manolio, T.A., 2008. How to interpret a genome-wide association study. *JAMA*, 299(11), pp.1335–44.
- Perlis, R.H. et al., 2008. Family-based association study of lithium-related and other candidate genes in bipolar disorder. *Arch Gen Psychiatry*, 65(1), pp.53–61.
- Phillips, R., 2014. *Alcohol: A History* 1st ed., The University of North Carolina Press.
- Piasecki, T.M., Robertson, B.M. & Epler, A.J., 2010. Hangover and risk for alcohol use disorders: existing evidence and potential mechanisms. *Current drug abuse reviews*, 3(2), pp.92–102.
- Piccardi, M.P. et al., 1997. No evidence of association between dopamine D3 receptor gene and bipolar affective disorder. *American journal of medical genetics*, 74(2), pp.137–9.
- Plomin, R., Haworth, C.M.A. & Davis, O.S.P., 2009. Common disorders are quantitative traits. *Nature reviews. Genetics*, 10(12), pp.872–8.
- Pollock, V.E., 1992. Meta-analysis of subjective sensitivity to alcohol in sons of alcoholics. *The American journal of psychiatry*, 149(11), pp.1534–8.
- Porjesz, B. et al., 2002. Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), pp.3729–33.
- Potkin, S.G. et al., 2009. A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia bulletin*, 35(1), pp.96–108.
- Power, R.A. & Pluess, M., 2015. Heritability estimates of the Big Five personality traits based on common genetic variants. *Translational psychiatry*, 5, p.e604.
- Preissig, C.M. & Rigby, M.R., 2009. Hyperglycaemia results from beta-cell dysfunction in critically ill children with respiratory and cardiovascular failure: a prospective observational study. *Critical care (London, England)*, 13(1), p.R27.
- Prescott, C.A. & Kendler, K.S., 1999. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *The American journal of psychiatry*, 156(1), pp.34–40.
- Preuss, U.W. et al., 2002. No association between metabotropic glutamate receptors 7 and 8 (mGlu7 and mGlu8) gene polymorphisms and withdrawal seizures and delirium tremens in alcohol-dependent individuals. *Alcohol Alcohol*, 37(2), pp.174–178.
- Pritchard, J.K., 2001. Are Rare Variants Responsible for Susceptibility to Complex Diseases? *The American Journal of Human Genetics*, 69(1), pp.124–137.
- Pruim, R.J. et al., 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)*, 26(18), pp.2336–7.

- Pulst, S.M., 1999. Genetic Linkage Analysis. *Archives of Neurology*, 56(6), p.667.
- Pulver, A.E. et al., 1995. Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *American journal of medical genetics*, 60(3), pp.252–60.
- Pulver, A.E. et al., 1994. Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: Part 1. *American journal of medical genetics*, 54(1), pp.36–43.
- Purcell, S.M. et al., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), pp.748–52.
- Quertemont, E., 2004. Genetic polymorphism in ethanol metabolism: acetaldehyde contribution to alcohol abuse and alcoholism. *Molecular psychiatry*, 9(6), pp.570–81.
- Quillen, E.E. et al., 2014. ALDH2 is associated to alcohol dependence and is the major genetic determinant of “daily maximum drinks” in a GWAS study of an isolated rural Chinese sample. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 165B(2), pp.103–10.
- Rabbee, N. & Speed, T.P., 2006. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics (Oxford, England)*, 22(1), pp.7–12.
- Rakofsky, J.J. & Dunlop, B.W., 2013. Do alcohol use disorders destabilize the course of bipolar disorder? *Journal of affective disorders*, 145(1), pp.1–10.
- Ray, L.A. et al., 2012. The Role of the Asn40Asp Polymorphism of the Mu Opioid Receptor Gene (OPRM1) on Alcoholism Etiology and Treatment: A Critical Review. *Alcoholism: Clinical and Experimental Research*, 36(3), pp.385–394.
- Ray, L.A. & Hutchison, K.E., 2004. A polymorphism of the mu-opioid receptor gene (OPRM1) and sensitivity to the effects of alcohol in humans. *Alcoholism, clinical and experimental research*, 28(12), pp.1789–95.
- Ray, R. et al., 2005. A functional prodynorphin promoter polymorphism and opioid dependence. *Psychiatric genetics*, 15(4), pp.295–8.
- Regier, D.A. et al., 1990. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA*, 264(19), pp.2511–2518.
- Rehm, J. et al., 2009. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet*, 373(9682), pp.2223–33.
- Reich, D.E. & Lander, E.S., 2001. On the allelic spectrum of human disease. *Trends in genetics : TIG*, 17(9), pp.502–10.
- Reich, T. et al., 1998a. Genome-wide search for genes affecting the risk for alcohol dependence. *American Journal of Medical Genetics*, 81(3), pp.207–215.
- Reich, T. et al., 1998b. Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet*, 81(3), pp.207–215.
- Reich, T. et al., 1988. Secular Trends in the Familial Transmission of Alcoholism. *Alcoholism: Clinical and Experimental Research*, 12(4), pp.458–464.
- Rice, J. et al., 1987. The familial transmission of bipolar illness. *Archives of general psychiatry*, 44(5), pp.441–7.
- Rietschel, M. et al., 1993. A serine to glycine substitution at position 9 in the extracellular N-terminal part of the dopamine D3 receptor protein: no role in the genetic predisposition to bipolar affective disorder. *Psychiatry research*, 46(3), pp.253–9.
- Riley, B. & Kendler, K.S., 2005. *Genetics of schizophrenia: linkage and association studies* K. Kendler & L. Eaves, eds., Washington, DC: Psychiatric Genetics. Washington DC: American Psychiatric Publishing Incorporated.
- Ripke, S. et al., 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), pp.421–427.
- Ripke, S. et al., 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics*, 45(10), pp.1150–9.

- Ripke, S. et al., 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.*, 43(10), pp.969–976.
- Risch, N., 1990. Linkage strategies for genetically complex traits. I. Multilocus models. *American journal of human genetics*, 46(2), pp.222–8.
- Risch, N. & Merikangas, K., 1996. The future of genetic studies of complex human diseases. *Science*, 273, pp.1516–1517.
- Risso, D. et al., 2014. Genetic variation in taste receptor pseudogenes provides evidence for a dynamic role in human evolution. *BMC evolutionary biology*, 14(1), p.198.
- Rivero, O. et al., 2013. Impact of the ADHD-susceptibility gene CDH13 on development and function of brain networks. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 23(6), pp.492–507.
- Roerecke, M. & Rehm, J., 2012. The cardioprotective association of average alcohol consumption and ischaemic heart disease: a systematic review and meta-analysis. *Addiction (Abingdon, England)*, 107(7), pp.1246–60.
- Room, R., Babor, T. & Rehm, J., 2005. Alcohol and public health. *Lancet (London, England)*, 365(9458), pp.519–30.
- Rosenthal, D. et al., 1971. The adopted-away offspring of schizophrenics. *The American journal of psychiatry*, 128(3), pp.307–11.
- Rösner, S. et al., 2010. Opioid antagonists for alcohol dependence. In S. Rösner, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd.
- Ruderfer, D.M. et al., 2013. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular Psychiatry*, 19(9), pp.1017–1024.
- Rudgley, R., 1994. *Essential Substances: A Cultural History of Intoxicants in Society*, Kodansha International.
- Rujescu, D. et al., 2005. GRIN1 locus may modify the susceptibility to seizures during alcohol withdrawal. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 133B(1), pp.85–7.
- Saccone, N.L. et al., 2000. A genome screen of maximum number of drinks as an alcoholism phenotype. *Am J Med Genet*, 96(5), pp.632–637.
- Saccone, S.F. & Rice, J.P., 2005. Alcoholism and related traits: a summary of Group 13 contributions. *Genetic Epidemiology*, 29(S1), pp.S96–S102.
- Saffroy, R. et al., 2008. Protective effect against alcohol dependence of the thermolabile variant of MTHFR. *Drug and alcohol dependence*, 96(1-2), pp.30–6.
- Samochowicz, J. et al., 2014. Genetics of alcohol dependence: a review of clinical studies. *Neuropsychobiology*, 70(2), pp.77–94.
- Samochowicz, J. et al., 2000. Predominant influence of the 3'-region of dopamine D2 receptor gene (DRD2) on the clinical phenotype in German alcoholics. *Pharmacogenetics*, 10(5), pp.471–5.
- Sander, T. et al., 1999. Association analysis of sequence variants of GABA(A) alpha6, beta2, and gamma2 gene cluster and alcohol dependence. *Alcohol Clin Exp Res*, 23(3), pp.427–431.
- Sander, T. et al., 1995. Dopamine D1, D2 and D3 receptor genes in alcohol dependence. *Psychiatric genetics*, 5(4), pp.171–6.
- Santosh, P.J. & Mijovic, A., 2006. Does pervasive developmental disorder protect children and adolescents against drug and alcohol use? *European Child & Adolescent Psychiatry*, 15(4), pp.183–188.
- Saxena, R. et al., 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science (New York, N.Y.)*, 316(5829), pp.1331–6.
- Schadt, E.E., 2009. Molecular networks as sensors and drivers of common human diseases. *Nature*, 461(7261), pp.218–23.
- Schinka, J.A., Letsch, E.A. & Crawford, F.C., 2002. DRD4 and novelty seeking: results of meta-analyses. *American journal of medical genetics*, 114(6), pp.643–8.

- Schork, A.J. et al., 2013. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS genetics*, 9(4), p.e1003449.
- Schuckit, M., 2002. Vulnerability factors for alcoholism. In E. Davis K, Charney DS, Coyle J, Nemeroff C, ed. *Neuropsychopharmacology, The Fifth Generation of Progress*. Philadelphia, Pa: Lippincott Williams & Wilkins, pp. 1399–1411.
- Schuckit, M. a, 2009. An overview of genetic influences in alcoholism. *Journal of substance abuse treatment*, 36(1), pp.S5–S14.
- Schuckit, M.A., 2014. A brief history of research on the genetics of alcohol and other drug use disorders. *Journal of studies on alcohol and drugs. Supplement*, 75 Suppl 1, pp.59–67.
- Schuckit, M.A., 1994. A clinical model of genetic influences in alcohol dependence. *Journal of Studies on Alcohol*, 55(1), pp.5–17.
- Schuckit, M.A. et al., 2001. A Genome-Wide Search for Genes That Relate to a Low Level of Response to Alcohol. *Alcoholism: Clinical and Experimental Research*, 25(3), pp.323–329.
- Schuckit, M.A. et al., 1996. Alcohol challenges in young men from alcoholic pedigrees and control families: a report from the COGA project. *Journal of studies on alcohol*, 57(4), pp.368–77.
- Schuckit, M.A. et al., 1995. An evaluation of type A and B alcoholics. *Addiction (Abingdon, England)*, 90(9), pp.1189–203.
- Schuckit, M.A., Goodwin, D.A. & Winokur, G., 1972. A study of alcoholism in half siblings. *The American journal of psychiatry*, 128(9), pp.1132–6.
- Schuckit, M.A., Smith, T.L. & Tipp, J.E., 1997. The Self-Rating of the Effects of alcohol (SRE) form as a retrospective measure of the risk for alcoholism. *Addiction (Abingdon, England)*, 92(8), pp.979–88.
- Schulze, T.G., 2010. Genetic research into bipolar disorder: the need for a research framework that integrates sophisticated molecular biology and clinically informed phenotype characterization. *The Psychiatric clinics of North America*, 33(1), pp.67–82.
- Schulze, T.G. & McMahon, F.J., 2003. Genetic linkage and association studies in bipolar affective disorder: a time for optimism. *American journal of medical genetics. Part C, Seminars in medical genetics*, 123C(1), pp.36–47.
- Schumann, G. et al., 2011. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), pp.7119–24.
- Schwartz, S. & Susser, E., 2011. The use of well controls: an unhealthy practice in psychiatric research. *Psychological medicine*, 41(6), pp.1127–31.
- Scott, L.J. et al., 2009. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proceedings of the National Academy of Sciences*, 106(18), pp.7501–7506.
- Segurado, R. et al., 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *American journal of human genetics*, 73(1), pp.49–62.
- Seneviratne, C. et al., 2013. Association, interaction, and replication analysis of genes encoding serotonin transporter and 5-HT₃ receptor subunits A and B in alcohol dependence. *Human genetics*, 132(10), pp.1165–76.
- Serretti, A. & Mandelli, L., 2008. The genetics of bipolar disorder: genome “hot regions,” genes, new potential candidates and future directions. *Molecular Psychiatry*, 13(8), pp.742–771.
- Severino, G. et al., 2005. A48G polymorphism in the D1 receptor genes associated with bipolar I disorder. *American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*, 134B(1), pp.37–8.
- Sham, P.C. & Purcell, S.M., 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15(5), pp.335–346.
- Sharp, S.I. et al., 2014. Genetic association of the tachykinin receptor 1 TACR1 gene in bipolar disorder, attention deficit hyperactivity disorder, and the alcohol dependence syndrome. *American*

- journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 165B(4), pp.373–80.
- Shi, J. et al., 2009. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, 460(7256), pp.753–7.
- Shi, J. et al., 2008. Neurotransmission and bipolar disorder: a systematic family-based association study. *Am J Med Genet B Neuropsychiatr Genet*, 147B(7), pp.1270–1277.
- Shi, Y., 2007. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nature reviews. Genetics*, 8(11), pp.829–33.
- Sigvardsson, S. et al., 1982. Predisposition to petty criminality in Swedish adoptees. III. Sex differences and validation of the male typology. *Arch Gen Psychiatry*, 39(11), pp.1248–1253.
- Sigvardsson, S., Bohman, M. & Cloninger, C.R., 1996. Replication of the Stockholm Adoption Study of alcoholism. Confirmatory cross-fostering analysis. *Arch Gen Psychiatry*, 53(8), pp.681–687.
- Sivakumaran, S. et al., 2011. Abundant pleiotropy in human complex diseases and traits. *American journal of human genetics*, 89(5), pp.607–18.
- Sklar, P. et al., 2008. Whole-genome association study of bipolar disorder. *Molecular psychiatry*, 13(6), pp.558–69.
- Slutske, W.S. et al., 1998. Common genetic risk factors for conduct disorder and alcohol dependence. *Journal of abnormal psychology*, 107(3), pp.363–74.
- Slutske, W.S. et al., 2008. Searching for an environmental effect of parental alcoholism on offspring alcohol use disorder: a genetically informed study of children of alcoholics. *J Abnorm Psychol*, 117(3), pp.534–551.
- Slutske, W.S., Piasecki, T.M. & Hunt-Carter, E.E., 2003. Development and initial validation of the Hangover Symptoms Scale: prevalence and correlates of Hangover Symptoms in college students. *Alcoholism, clinical and experimental research*, 27(9), pp.1442–50.
- Smith, E.N. et al., 2009. Genome-wide association study of bipolar disorder in European American and African American individuals. *Molecular psychiatry*, 14(8), pp.755–63.
- Sokolov, B.P. et al., 2003. Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. *J Neurosci Res*, 72(6), pp.756–67.
- Solovieff, N. et al., 2013. Pleiotropy in complex traits: challenges and strategies. *Nature reviews. Genetics*, 14(7), pp.483–95.
- Span, S.A. & Earleywine, M., 1999. Familial risk for alcoholism and hangover symptoms. *Addictive Behaviors*, 24(1), pp.121–125.
- Spielman, R.S. et al., 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics*, 52(3), pp.506–516.
- Spitzer, R.L. & Endicott, J., 1977. *The Schedule for Affective Disorders and Schizophrenia, Lifetime Version, 3rd edition.*, New York: New York State Psychiatric Institute.
- Stefansson, H. et al., 2009. Common variants conferring risk of schizophrenia. *Nature*, 460(7256), pp.744–7.
- Stefansson, H. et al., 2002. Neuregulin 1 and Susceptibility to Schizophrenia. *The American Journal of Human Genetics*, 71(4), pp.877–892.
- Steinberg, S. et al., 2014. Common variant at 16p11.2 conferring risk of psychosis. *Molecular psychiatry*, 19(1), pp.108–14.
- Stöber, G. et al., 1998. Insertion/deletion variant (-141C Ins/Del) in the 5' regulatory region of the dopamine D2 receptor gene: lack of association with schizophrenia and bipolar affective disorder. Short communication. *Journal of neural transmission (Vienna, Austria : 1996)*, 105(1), pp.101–9.
- Strakowski, S.M. et al., 2000. The impact of substance abuse on the course of bipolar disorder. *Biological psychiatry*, 48(6), pp.477–85.
- Suarez, B.K. & Washington, D.C., 1994. C. L. Hampe, . Problems of replicating linkage claims in

- psychiatry. *Genetic approaches to mental disorders* E S Gershon and C R Coninger Am Psychiatric Press.
- Sulik, K.K., 2005. Genesis of alcohol-induced craniofacial dysmorphism. *Experimental biology and medicine* (Maywood, N.J.), 230(6), pp.366–75.
- Sullivan, P.F., Daly, M.J. & O'Donovan, M., 2012. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature reviews. Genetics*, 13(8), pp.537–51.
- Sullivan, P.F., Kendler, K.S. & Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry*, 60(12), pp.1187–92.
- Sulovari, A. et al., 2015. Eye color: A potential indicator of alcohol dependence risk in European Americans. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 168B(5), pp.347–53.
- Szczepankiewicz, A. et al., 2007. Study of dopamine receptors genes polymorphisms in bipolar patients with comorbid alcohol abuse. *Alcohol and Alcoholism*, 42(2), pp.70–74.
- Takaki, H. et al., 2004. Positive associations of polymorphisms in the metabotropic glutamate receptor type 8 gene (GRM8) with schizophrenia. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 128B(1), pp.6–14.
- Takeuchi, F. et al., 2011. Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circulation journal : official journal of the Japanese Circulation Society*, 75(4), pp.911–8.
- Tanda, G. et al., 1997. Contribution of blockade of the noradrenaline carrier to the increase of extracellular dopamine in the rat prefrontal cortex by amphetamine and cocaine. *The European journal of neuroscience*, 9(10), pp.2077–85.
- Taqi, M.M. et al., 2011. Prodynorphin CpG-SNPs associated with alcohol dependence: elevated methylation in the brain of human alcoholics. *Addiction biology*, 16(3), pp.499–509.
- Tarter, R.E. et al., 2003. Neurobehavioral Disinhibition in Childhood Predicts Early Age at Onset of Substance Use Disorder. *American Journal of Psychiatry*, 160(6), pp.1078–1085.
- Tarter, R.E. et al., 1990. Temperament Deviation and Risk for Alcoholism. *Alcoholism: Clinical and Experimental Research*, 14(3), pp.380–382.
- Tarter, R.E. & Vanyukov, M., 1994. Alcoholism: a developmental disorder. *Journal of Consulting and Clinical Psychology*, 62(6), pp.1096–1107.
- Thiselton, D.L. et al., 2004. No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). *Molecular psychiatry*, 9(8), pp.777–83; image 729.
- Thome, J. et al., 1999. Dopamine D3 receptor gene polymorphism and alcohol dependence: relation to personality rating. *Psychiatric genetics*, 9(1), pp.17–21.
- Tienari, P. et al., 2003. Genetic boundaries of the schizophrenia spectrum: evidence from the Finnish Adoptive Family Study of Schizophrenia. *The American journal of psychiatry*, 160(9), pp.1587–94.
- Tiihonen, J. et al., 1999. Association between the functional variant of the catechol-O-methyltransferase (COMT) gene and type 1 alcoholism. *Molecular psychiatry*, 4(3), pp.286–289.
- Todd, R.D. et al., 1996. Increased prevalence of alcoholism in relatives of depressed and bipolar children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(6), pp.716–724.
- Treutlein, J. et al., 2009. Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry*, 66(7), pp.773–784.
- Tsuang, M. & Faraone, S., 1990. *The genetics of mood disorders.*,
- Tsuang, M.T., Faraone, S. V & Lyons, M.J., 1993. Identification of the phenotype in psychiatric genetics. *European archives of psychiatry and clinical neuroscience*, 243(3-4), pp.131–42.
- Tsuang, M.T., Winokur, G. & Crowe, R.R., 1980. Morbidity risks of schizophrenia and affective

- disorders among first degree relatives of patients with schizophrenia, mania, depression and surgical conditions. *The British Journal of Psychiatry*, 137(6), pp.497–504.
- Tung, J.Y. et al., 2011. Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS one*, 6(8), p.e23473.
- Turner, S.D., 2014. *qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots*, Cold Spring Harbor Labs Journals.
- Uhl, G.R. et al., 2008. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch Gen Psychiatry*, 65(6), pp.683–693.
- Uhl, G.R. & Drgonova, J., 2014. Cell adhesion molecules: druggable targets for modulating the connectome and brain disorders? *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 39(1), p.235.
- Uhl, G.R. & Grow, R.W., 2004. The burden of complex genetics in brain disorders. *Archives of general psychiatry*, 61(3), pp.223–9.
- Ulloa, A.E. et al., 2014. Association between copy number variation losses and alcohol dependence across African American and European American ethnic groups. *Alcoholism, clinical and experimental research*, 38(5), pp.1266–74.
- Üstün, B. et al., 1997. WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug and Alcohol Dependence*, 47(3), pp.161–169.
- Uusi-Oukari, M. et al., 2000. Long-range interactions in neuronal gene expression: evidence from gene targeting in the GABA(A) receptor beta2-alpha6-alpha1-gamma2 subunit gene cluster. *Molecular and cellular neurosciences*, 16(1), pp.34–41.
- Venken, T. et al., 2008. Chromosome 10q harbors a susceptibility locus for bipolar disorder in Ashkenazi Jewish families. *Molecular psychiatry*, 13(4), pp.442–50.
- Verma, R. et al., 2005. SYNGR1 is associated with schizophrenia and bipolar disorder in southern India. *Journal of human genetics*, 50(12), pp.635–40.
- Verweij, K.J.H. et al., 2010. A genome-wide association study of Cloninger's temperament scales: implications for the evolutionary genetics of personality. *Biological psychology*, 85(2), pp.306–17.
- Voisey, J. et al., 2011. A novel SNP in COMT is associated with alcohol dependence but not opiate or nicotine dependence: a case control study. *Behavioral and brain functions : BBF*, 7(1), p.51.
- Vyssoki, B. et al., 2011. Inpatient treatment of major depression in Austria between 1989 and 2009: Impact of downsizing of psychiatric hospitals on admissions, suicide rates and outpatient psychiatric services. *Journal of Affective Disorders*, 133(1-2), pp.93–96.
- Walters, J.T.R. et al., 2013. The Role of the Major Histocompatibility Complex Region in Cognition and Brain Structure: A Schizophrenia GWAS Follow-Up. *American Journal of Psychiatry*, 170(8), pp.877–885.
- Wang, D.G., 1998. Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science*, 280(5366), pp.1077–1082.
- Wang, H. et al., 2008. CaMKII activation state underlies synaptic labile phase of LTP and short-term memory formation. *Current biology : CB*, 18(20), pp.1546–54.
- Wang, J.C. et al., 2004. Evidence of common and specific genetic effects: association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome. *Human molecular genetics*, 13(17), pp.1903–11.
- Wang, J.-C. et al., 2013. A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. *Molecular psychiatry*, 18(11), pp.1218–24.
- Wang, K., Li, M. & Bucan, M., 2007. Pathway-based approaches for analysis of genomewide association studies. *American journal of human genetics*, 81(6), pp.1278–83.
- Wang, K., Li, M. & Hakonarson, H., 2010. Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics*, 11(12), pp.843–854.
- Wang, K.-S. et al., 2011. A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. *Journal of psychiatric research*, 45(11), pp.1419–25.

- Wang, Y. et al., 2009. No association between bipolar disorder and syngr1 or synapsin II polymorphisms in the Han Chinese population. *Psychiatry research*, 169(2), pp.167–8.
- Way, M. et al., 2015. Genetic variants in or near ADH 1 B and ADH 1 C affect susceptibility to alcohol dependence in a British and Irish population. *Addiction Biology*, 20(3), pp.594–604.
- Webb, A. et al., 2008. Genetic approaches to addiction: genes and alcohol E. N. and F. L. de Silva, ed. *Alcohol Clin Exp Res*, 8(1), pp.207–215.
- Weber, J.L. & May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American journal of human genetics*, 44(3), pp.388–96.
- Weiss, F. et al., 1993. Oral alcohol self-administration stimulates dopamine release in the rat nucleus accumbens: genetic and motivational determinants. *J. Pharmacol. Exp. Ther.*, 267(1), pp.250–258.
- Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), pp.661–678.
- Wender, P.H. et al., 1974. Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia. *Archives of general psychiatry*, 30(1), pp.121–8.
- Westreich, D., 2012. Berkson's bias, selection bias, and missing data. *Epidemiology (Cambridge, Mass.)*, 23(1), pp.159–64.
- Wetherill, L. et al., 2008. Neuropeptide Y receptor genes are associated with alcohol dependence, alcohol withdrawal phenotypes, and cocaine dependence. *Alcoholism, clinical and experimental research*, 32(12), pp.2031–40.
- Whatley, V.J., Johnson, T.E. & Erwin, V.G., 1999. Identification and confirmation of quantitative trait loci regulating alcohol consumption in congenic strains of mice. *Alcoholism, clinical and experimental research*, 23(7), pp.1262–71.
- Whitfield, J.B. et al., 2001. Variation in Alcohol Pharmacokinetics as a Risk Factor for Alcohol Dependence. *Alcoholism: Clinical and Experimental Research*, 25(9), pp.1257–1263.
- Whitfield, J.B. & Martin, N.G., 1994. Alcohol Consumption and Alcohol Pharmacokinetics: Interactions Within the Normal Population. *Alcoholism: Clinical and Experimental Research*, 18(2), pp.238–243.
- van den Wildenberg, E. et al., 2007. A functional polymorphism of the mu-opioid receptor gene (OPRM1) influences cue-induced craving for alcohol in male heavy drinkers. *Alcoholism, clinical and experimental research*, 31(1), pp.1–10.
- Williams, H.J. et al., 2011. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Molecular psychiatry*, 16(4), pp.429–41.
- Williams, H.J. et al., 2005. No association between schizophrenia and polymorphisms in COMT in two large samples. *The American journal of psychiatry*, 162(9), pp.1736–8.
- Williams, T.J. et al., 2007. Prodynorphin gene promoter repeat associated with cocaine/alcohol codependence. *Addiction biology*, 12(3-4), pp.496–502.
- Winokur, G. et al., 1993. Alcoholism in manic depressive (bipolar) patients. *Journal of studies on alcohol*, 54(5), pp.574–576.
- Winterer, G. et al., 2000. Association analysis of GABAAbeta2 and gamma2 gene polymorphisms with event-related prefrontal activity in man. *Human genetics*, 107(5), pp.513–8.
- Withington, P. & McShane, A., 2009. *Memorandum by Dr Phil Withington and Dr Angela McShane (AL 57): FLUCTUATIONS IN ENGLISH DRINKING HABITS: AN HISTORICAL OVERVIEW*.
- Wodarz, N. et al., 2003. The candidate gene approach in alcoholism: are there gender-specific differences? *Archives of women's mental health*, 6(4), pp.225–30.
- Wojcik, G.L. et al., 2015. Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies. *BMC Genetics*, 16(1), p.34.
- Wonodi, I. et al., 2003. Association between Val108/158 Met polymorphism of the COMT gene and schizophrenia. *American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*, 120B(1), pp.47–50.

- World Health Organization, 1992. *ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines*, Geneva: World Health Organisation.
- World Health Organization, 2014. *WHO Global status report on alcohol and health 2014*, Geneva: World Health Organization.
- Wrzosek, M. et al., 2011. Association of polymorphisms in HTR2A, HTR1A and TPH2 genes with suicide attempts in alcohol dependence: a preliminary report. *Psychiatry research*, 190(1), pp.149–51.
- Wyszynski, D.F. et al., 2003. Genome-wide screen for heavy alcohol consumption. *BMC genetics*, 4 Suppl 1, p.S106.
- Xu, B. et al., 2012. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics*, 44(12), pp.1365–9.
- Xu, W. et al., 2014. Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. *BMC medical genetics*, 15(1), p.2.
- Yan, T. et al., 2009. NK1 (TACR1) receptor gene “knockout” mouse phenotype predicts genetic association with ADHD. *J Psychopharmacol*.
- Zhang, L. et al., 2014. Association analysis of the GRM8 gene with schizophrenia in the Uygur Chinese population. *Hereditas*, 151(6), pp.140–4.
- Zhong, X. et al., 2015. Human cell adhesion molecules: annotated functional subtypes and overrepresentation of addiction-associated genes. *Annals of the New York Academy of Sciences*, 1349(1), pp.83–95.
- Zill, P. et al., 2008. Analysis of single nucleotide polymorphisms and haplotypes in the neuropeptide Y gene: no evidence for association with alcoholism in a German population sample. *Alcoholism, clinical and experimental research*, 32(3), pp.430–4.
- Zlojutro, M. et al., 2010. Genome-wide association study of theta band event-related oscillations identifies serotonin receptor gene HTR7 influencing risk of alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet*, 156(1), pp.44–58.
- Zuo, L. et al., 2015. A New Genomewide Association Meta-Analysis of Alcohol Dependence. *Alcoholism, clinical and experimental research*, 39(8), pp.1388–95.
- Zuo, L. et al., 2011. A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. *PloS one*, 6(11), p.e26726.
- Zuo, L. et al., Genome-wide association discoveries of alcohol dependence. *The American journal on addictions / American Academy of Psychiatrists in Alcoholism and Addictions*, 23(6), pp.526–39.
- Zuo, L. et al., 2012. Genome-wide search for replicable risk gene regions in alcohol and nicotine co-dependence. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 159B(4), pp.437–44.

APPENDICES

18.1 LIST OF TABLES

- Table 4.1. Comorbidity of Alcoholism and Mood Disorders in relatives
- Table 5.1 Chromosomal Regions with Evidence of Linkage to AD and related phenotypes (adapted from Gizer et al. 2011).
- Table 8.1. Genome-Wide Association Study Findings for SZ and BP from the review by Sullivan et al, combined with Specificity of Genome-Wide Association Study Findings for SZ and BP in the Psychiatric Genomics Consortium Cross-Disorder Group Study of 5 Psychiatric Disorders
- Table 8.2 Relative risks for developing SZ and BP with affected probands (from Lichtenstein et al 2009)
- Table 10.1 GABRA2 SNPs genotyped in the UCL sample
- Table 11.1. GABRA2 SNP results in the UCL AD sample
- Table 12.1 BPALC vs BPnonALC: Top 40 SNPs and gene based results
- Table 12.2 BPALC vs BPnonALC Top 40 gene based analysis genes
- Table 12.3 BPALC vs BPnonALC gene wise replications of AD or related phenotypes from the literature
- Table 12.4: Top 10 BPALC vs BP VEGAS2 pathways
- Table 12.1 BPALC vs SNC: Top 40 SNPs and gene based results
- Table 12.2 BPALC vs SNC Top 40 Genes After VEGAS2 Permutation Testing
- Table 12.3 BPALC vs normal controls: Genes Yielding Evidence for Association at P
- Table 12.4 VEGAS2 pathway analysis results
- Table 13.1 SZALC vs SZnonALC: Top 40 SNPs and gene based results
- Table 13.2 SZALC vs SZ Top 40 Genes After VEGAS2 Permutation Testing
- Table 13.3 SZALC vs SZ Genes Yielding Evidence for Association at $P < 0.05$ After Permutation Testing and Literature Replication
- Table 14.1 SZALC vs SNC: Top 40 SNPs and gene based significance values
- Table 14.2 SZALC vs SNC: Top 40 gene based significance values
- Table 14.3 SZALC vs SNC Genes Yielding Evidence for Association at $P < 0.05$ After Permutation Testing and Literature Replication
- Table 15.1 Meta vs SEMI Top 40 SNPs and gene based results
- Table 15.2 Top 40 Genes Yielding Evidence for Association After VEGAS2 Permutation Testing
- Table 15.3 Meta-analysis vs BP+SZ Genes Yielding Evidence for Association at $P < 0.05$ After Permutation Testing and replication
- Table 16.1 Meta-analysis vs supernormal controls: Top 40 SNPs and gene based results
- Table 16.2 Top 40 Genes Yielding Evidence for Association After VEGAS2 Permutation Testing
- Table 16.3 Meta-analysis vs normal controls: Genes Yielding Uncorrected Evidence for Association at $P < 0.05$ After Permutation Testing and Replication
- Table 16.4 Top 16 PGC CDG and UCL VEGAS2 gene pathways

18.2 LIST OF FIGURES

Figure 10-1 GABRA2 comparison of A. HapMap CEU Linkage Disequilibrium (LD) and B. LD from the UCL AD data.

Figure 11-1 Q-Q plot of BPALC vs BPnonALC GWAS p-values

Figure 11-2 Manhattan Plot of BPALC vs BPnonALC

Figure 11-3 BPALC vs BPnonALC: LocusZoom regional plot of top SNP and nearest gene SGOL1

Fig. 11-4 BPALC vs BPnonALC: Region plot of SNP nearest GADL1

Figure 11-5 UCL BPALC vs BPnonALC plot of SEMA5A and strongest SNP

Figure 11-6 PGC Bipolar GWAS plot of same region as Fig. 11-5

Fig. 11-7 BPALC vs BPnonALC: ADH5 region plot

Fig. 11-8 BPALC vs BPnonALC: CDH13 region plot

Fig. 11-9 BPALC vs BPnonALC gene pathway analysis

Fig. 12-1 Q-Q Plot of BPALC vs SNC

Fig. 12-2 Manhattan plot of BPALC vs SNC

Fig. 12-3 BPALC vs SNC: TAS2R38 region plot

Fig. 12-4 BPALC vs SNC: CDH11 region plot

Fig. 12-5 BPALC vs SNC: HAAO region plot

Fig. 12-6 BPALC vs SNC: GABRB3 region plot

Fig. 12-7 BPALC vs SNC: PANTHER histogram of numbers of genes in each GO pathway

Fig. 13-1 Q-Q plot of SZALC vs SZ

Fig. 13-2 Manhattan plot of SZALC vs SZ

Fig. 13-3 SZALC vs SZnonALC: Top SNP in LAMB1

Fig. 13-4 SZALC vs SZnonALC: CADM2 region

Fig. 13-5 SZALC vs SZnonALC: LOC257358

Fig. 13-6 SZALC vs SZnonALC: CAMK2A

Fig. 14-1 Q-Q plot of SZALC vs SNC

Fig. 14-2 Manhattan plot of SZALC vs SNC

Fig. 14-3 SZALC vs SNC: uncharacterised gene LOC101927533 region plot

Fig. 14-4 SZALC vs SNC: uncharacterised gene LOC101927166 region plot

Fig. 14-5 SZALC vs SNC: PDYN region plot (gene based $P = 2.03 \times 10^{-2}$)

Fig. 15-1 Q-Q plot of meta-analysis vs “pure” BP+SZ (SEMI)

Fig. 15-2 Manhattan plot of meta-analysis vs “pure” BP+SZ (SEMI)

Fig. 15-3 UCL meta-analysis vs SEMI: strongest SNP near TLN2

Fig. 15-4 UCL meta-analysis vs SEMI: AUTS2 region

Fig. 15-5 UCL meta-analysis vs pure mental illness: PANTHER histogram of numbers of genes in each GO pathway

Fig. 16-1 Q-Q plot of meta-analysis vs supernormal controls GWAS

Fig. 16-2 Manhattan plot of Meta-analysis vs supernormal controls

Fig. 16-3 Meta-analysis vs SNC: strongest GWA SNP in LTBP1

Fig. 16-4 PGC CDG: plotting the UCL meta-analysis strongest SNP in LTBP1

Fig. 16-5 Meta-analysis plot of UCL CDH11

Fig. 16-6 PGC CDG plot of CDH11 region

Fig. 16-7 UCL meta-analysis vs SNC: Top gene based gene plot of PDGFC

Fig. 16-8 UCL meta-analysis vs SNC: UCL PKNOX2 region plot

Fig. 16-9 Meta-analysis vs SNC: PGC CDG PKNOX2 region plot

Fig. 16-10 UCL meta-analysis vs SNC: number of genes in each GO pathway

18.3 LIST OF PUBLISHED PAPERS (WHERE FIRST AUTHOR IN BOLD):

- Nyegaard M, Severinsen JE, Als TD, Hedemand A, Straarup S, Nordentoft M, McQuillin A, Bass N, Lawrence J, Thirumalai S, Pereira AC, Kandaswamy R, Lydall GJ, Sklar P, Scolnick E, Purcell S, Curtis D, Gurling HM, Mortensen PB, Mors O, Børnglum AD. Support of association between BRD1 and both schizophrenia and bipolar affective disorder. *Am J Med Genet B Neuropsychiatr Genet.* 2009 Aug 19
- McQuillin A, Bass N, Anjorin A, Lawrence J, Kandaswamy R, Lydall G, Moran J, Sklar P, Purcell S, Gurling H. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *Eur J Hum Genet.* 2011 May;19(5):588-92. Epub 2011 Jan 5
- **Lydall GJ**, Saini J, Ruparelia K, Montagnese S, McQuillin A, Guerrini I, Rao H, Reynolds G, Ball D, Smith I, Thomson AD, Morgan MY, Gurling HM. Genetic association study of GABRA2 single nucleotide polymorphisms and electroencephalography in alcohol dependence. *Neurosci Lett.* 2011 Aug 18;500(3):162-6. Epub 2011 Jun 12
- Kasiakogia-Worlley K, McQuillin A, Lydall GJ, Patel S, Kottalgi G, Gunwardena P, Cherian R, Rao H, Hillman A, Gobikrishnan N, Douglas E, Qureshi SY, Jauhar S, Ball D, Okane A, Owens L, Dedman A, Sharp SI, Kandaswamy R, Guerrini I, Thomson AD, Smith I, Dar K, Morgan MY, Gurling HM. Lack of allelic association between markers at the DRD2 and ANKK1 gene loci with the alcohol-dependence syndrome and criminal activity. *Psychiatr Genet.* 2011 Dec;21(6):323-4. doi: 10.1097/YPG.0b013e3283458a68.
- **Lydall GJ**, Bass NJ, McQuillin A, Lawrence J, Anjorin A, Kandaswamy R, Pereira A, Guerrini I, Curtis D, Vine AE, Sklar P, Purcell SM, Gurling HM. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. *Psychiatr Genet.* 2011 Dec;21(6):294-306.
- Quadri et al. Evidence for genetic susceptibility to the alcohol dependence syndrome from the thiamine transporter 2 gene solute carrier SLC19A3. *Psychiatr Genet.* Jun 2014; 24(3): 122–123.
- Way M, et al. Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addict Biol.* 2014 Apr 16.
- Sharp et al. Genetic association of the tachykinin receptor 1 TACR1 gene in bipolar disorder, attention deficit hyperactivity disorder, and the alcohol dependence syndrome. *J Med Genet B Neuropsychiatr Genet.* 2014 Jun;165B(4):373-80.
- O'Brien NL et al. The functional GRM3 Kozak sequence variant rs148754219 affects the risk of schizophrenia and alcohol dependence as well as bipolar disorder. *Psychiatr Genet.* 2014 Dec;24(6):277-8.

Copies of the published papers follow.