Bound by
Abbey Bookbinding
Unit 3 Gabalfa Workshops
Clos Menter
Excelsior Ind. Est.
Cardiff CF14 3AY
T: +44 (0) 29 2062 3290
F: +44 (0) 29 2062 5420
E: info@abbeybookbinding.co.uk
W: www.abbeybookbinding.co.uk

Sir Herbert Duthie Library

Llyfrgell Syr Herbert Duthie

University Hospital
of Wales
Heath Park
Cardiff
CF14 4XN

Ysbyty Athrofaol Cymru
Parc y Mynydd Bychan
Caerdydd
CF14 4XN

029 2074 2875
duthieliby@cardiff.ac.uk

The role of the Y chromosome in sex differences in ADHD and schizophrenia

Evangelia Stergiakouli

Thesis submitted for the degree of
Doctor of Philosophy at Cardiff University

Department of Psychological Medicine
and Neurology
Cardiff University

Supervisors:
Professor Anita Thapar and Professor Michael J Owen

April 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration and Statements

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date ...10/09/2010...

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ..... (candidate) Date ...10/09/2010.

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ..... (candidate) Date ...10/09/2010.

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date ...10/09/2010...

STATEMENT 4

This thesis is the result of my own work, unless otherwise stated.
I reviewed the literature and designed the Y chromosome panel genotyped in this study utilising advice on appropriate Y chromosome markers provided by Professor Mark Jobling.
I performed genotyping of the Y chromosome panel in the ADHD and schizophrenia samples and controls from the 1958 Birth Cohort. I selected appropriate phenotypic measures and performed phenotypic analysis on ADHD, schizophrenia and ALSPAC samples.
DNA from patients with ADHD and schizophrenia was collected by trained psychologists employed by the Department of Psychological Medicine and Neurology and they also performed patient assessments and collection of phenotypic measures.
DNA extractions were performed by departmental technicians.
Genotyping of Y chromosome markers in the ALSPAC samples was performed by KBiosciences.
I also designed the Tyrosine Hydroxylase panel and genotyped it in the ADHD and schizophrenia samples and controls from the 1958 Birth Cohort.
I performed all statistical analysis after consulting statisticians from the Biostatistics and Bioinformatics Unit where necessary.
Dr Hywel Williams supervised the experiments, performed double-genotyping and provided practical advice and guidance.

Acknowledgements

First of all, I would like to thank my supervisor, Anita Thapar, for the help and support but especially for believing in me and encouraging me to go a step further. Her enthusiasm for science was contagious! Also, I would like to thank my supervisor, Mike Owen, for always finding time for me and providing valuable advice and guidance. I would especially like to thank both my supervisors for the opportunity to do a PhD that I honestly enjoyed.

Special thanks have to go to my mentor, Hywel Williams, for all the support, advice and encouragement. There was no way I could have done my PhD without his help and I cannot thank him enough.

Huge thanks must also go to Kate Langley for her help with the phenotypic analysis and not only that and to the field team members for their huge efforts to collect so many patients. I must also thank Mick O'Donovan for his advice during the SAGE meetings, Will Davies for giving me the idea to focus on the Y chromosome, Stan Zammit for the help with the schizophrenia phenotypic variables and Mark Jobling from the University of Leicester for valuable advice on studying the Y chromosome.

The fact that I enjoyed my PhD so much has definitely something to do with having so many nice people around me and especially in the office. It is just great to have friends that understand your problems and are never tired of listening to you. So, thanks to Amy, Becky, Debbie, Denise, Didi, Irina, Jade and Sarah.

Thanks to my brother Apostolos and my good friend Ioanna for the patient and encouragement. I know that I didn't do anything else apart from complaining since I started writing this thesis, but you were there to remind me that there are more important things to life. Thanks also to my friend Eleni for always being available to distract me with shopping and drinks!

Huge thanks must go to my parents. I know that I could never thank you enough for what you have done for me but I have to thank you for supporting all my decisions and making sure that I never forget to smile. I also have to thank two special teachers of me, Mr Yannis Alvizos and my dad, Takis Stergiakoulis, for passing me their love for science and making me believe I could achieve anything I want. This PhD is your fault!

The biggest thanks have to go to Dimitris and this thesis is dedicated to him. I know that I have been extremely lucky to have him next to me during these years. He provided the inspiration and the motive to go further and a patience and understanding that I could have never shown.

Summary

ADHD and schizophrenia are neurodevelopmental disorders that are more prevalent in males and show sex differences in age of onset or severity. The Y chromosome is potentially an important influence on male susceptibility to neuropsychiatric disorders. The way the Y chromosome could increase risk to neuropsychiatric disorders is directly or indirectly by interacting with autosomal genes expressed in the brain. In addition, it could modify the disease phenotype. However, due to difficulties arising from the lack of recombination and widely accepted nomenclature, the Y chromosome has been largely excluded from genetic and genomic studies of neuropsychiatric disorders.

To overcome this lack of knowledge, 9 Y chromosome markers were selected to represent the main Y chromosome haplogroups that are present in the U.K. and they were genotyped in a sample of 210 cases with ADHD, 313 cases with schizophrenia and 637 U.K. controls.

Statistical analysis of Y chromosome haplogroups revealed that although there was no significantly increased representation of any haplogroup in cases with ADHD or schizophrenia compared to controls, there was evidence of a possible modifying effect on the phenotype of ADHD and schizophrenia. Y chromosome haplogroup 3 was associated with higher performance and full scale IQ within the sample of patients with ADHD. Haplogroup 1 was associated with better outcome and higher educational level within the sample of patients with schizophrenia. There was no association of Y chromosome haplogroups with IQ in a population sample of 3,749 individuals.

Y chromosome haplogroups were also tested for interaction with tyrosine hydroxylase SNPs because animal studies suggest this is biologically plausible. Although there was no evidence of interaction, three tyrosine hydroxylase SNPs showed nominally significant association in the sample of male patients with schizophrenia.

This study is one of the largest Y chromosome studies in the UK. It suggests that although Y chromosome variation does not appear to be associated with ADHD or schizophrenia, it may modify cognitive performance and clinical features.

Table of contents

Index of Figures

Index of Tables

Chapter 1 Sex differences

Being male or female is a fixed factor which can have an effect not only on morphological traits but also on the organisation of the brain. In addition, sex is one of the few externally observed factors that can be measured accurately and easily. Research on sex differences can not only provide an answer to the fundamental question "Why do men and women differ in susceptibility to a large number of disorders?" but more importantly, it can provide clues to the underlying mechanisms of devastating disorders.

It is well-established that quantitative traits such as blood pressure, obesity and lipid levels are sexually dimorphic (Weiss et al. 2006). The sex-specific nature of quantitative traits has been investigated in members of the Hutterite community, because they practise a communal lifestyle with remarkable similarities for both men and women. For 16 out of 19 quantitative traits, including High-Density Lipoprotein (HDL), triglyceride levels, blood pressure and asthma, sex was a significant predictor (Pan et al. 2007; Weiss et al. 2006). Most importantly, complex diseases with a large impact on health tend to have different prevalence rates and courses in men and women. It is striking that it is actually difficult to find a human disease with no sex difference in prevalence, development, age of onset or severity. For example, cardiovascular disease and type 2 diabetes are more common in men compared to pre-menopausal women (Choi and McLaughlin 2007; Woods et al. 2003). Asthma and autoimmune diseases also show age-sex interactions with a higher prevalence in females which is attenuated after puberty (Lockshin 2006; Postma 2007).

1.1 Sex differences in psychopathology

Psychiatric disorders are no exception to the rule of sex differences in diseases. The pattern of sex differences in psychopathology is that of two groups of disorders: early-onset neurodevelopmental disorders show a male preponderance, while emotional disorders with an onset during adolescence or later show a female preponderance. Following this pattern, autism, Asperger's syndrome, Tourette's syndrome, ADHD, dyslexia, developmental language disorders, conduct disorder and early-onset

schizophrenia are more common in males. Females are more likely to be affected by depression, eating disorders and anxiety (Holden 2005; Ober et al. 2008). Figure 1-1 demonstrates the difference in prevalence of major psychiatric disorders between males and females. Antisocial behaviour appears to be an exception to this pattern, since it affects more males while peaking in adolescence. However, early-onset antisocial behaviour has an even higher male incidence with female antisocial behaviour rarely manifesting before adolescence (Rutter et al. 2003).

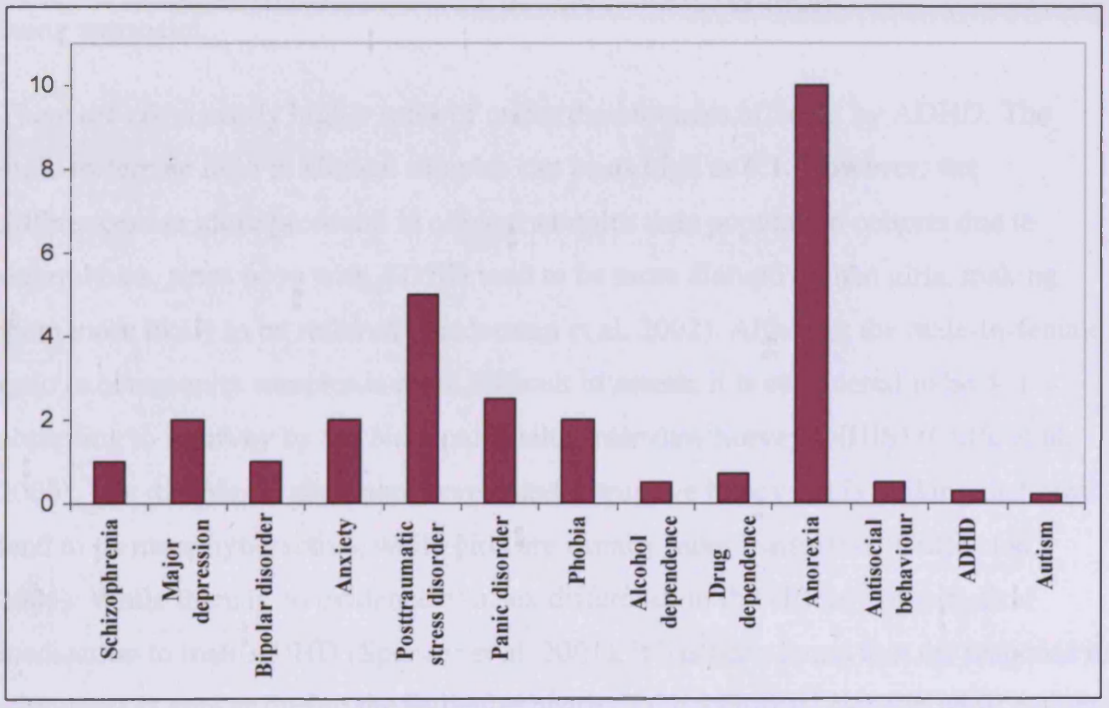

Figure 1-1. Female/male ratios in selected psychiatric disorders. Ratios >1 represent disorders with a female preponderance while ratios <1 represent disorders with a male preponderance. The data are based on a review by Holden et al. (2005)

1.1.1 Sex differences in ADHD and schizophrenia

Since this thesis focuses on the male excess in ADHD and schizophrenia, sex differences concerning these two neurodevelopmental disorders will be discussed further.

ADHD is one of the most common heritable childhood psychiatric disorders characterised by overactivity, impulsiveness and lack of concentration. It affects 1.4-5% of the population and places children at risk of poor educational attainment and of being antisocial.

There are consistently higher rates of males than females affected by ADHD. The male-to-female ratio in clinical samples can be as high as 6:1. However, sex differences are more profound in clinical samples than population cohorts due to referral bias, since boys with ADHD tend to be more disruptive than girls, making them more likely to be referred (Biederman et al. 2002). Although the male-to-female ratio in community samples is more difficult to assess, it is considered to be 4: 1 according to a survey by the National Health Interview Survey (NHIS) (Cuffe et al. 2005). The disorder is also more severe and disruptive in boys. It is striking that boys tend to be more hyperactive, while girls are usually more inattentive (Cuffe et al. 2005). While there is no evidence of a sex difference in the efficacy of stimulant medication to treat ADHD (Spencer et al. 2001), it has been found that the response to stimulants is greater during the follicular phase of the menstrual cycle in adult patients (Justice and De Wit 2000).

Schizophrenia is another debilitating psychiatric disorder characterised by psychotic symptoms, altered emotional reactivity and cognitive impairment. It is a highly heritable disorder affecting approximately 1% of the population and imposing a great burden on sufferers and their relatives. According to two systematic reviews, the median male/female rate ratio is 1.4 (Aleman et al. 2003; McGrath et al. 2004). The most striking sex differences in schizophrenia are found in age of onset and morbidity. The age of onset for men is on average 21 years while the average age of onset for women is 25 years (Ober et al. 2008). Early age of onset is associated with worse outcome and more impaired cognition. Men with schizophrenia also experience more negative symptoms compared to women, they have a history of more obstetric

complications and poorer premorbid adjustment (Preston et al. 2002). A longitudinal study following patients with schizophrenia over a period of twenty years found that women consistently showed better functioning, more periods of recovery, fewer hospitalisations and less likelihood of poor outcome even though the age of onset in these female patients was the same as in men (Grossman et al. 2006; Grossman et al. 2008). It is interesting that sex differences in schizophrenia exist in relation to brain morphology. For example, it was found that increased prefrontal lobe volume was associated with more disorganised behaviour in women with schizophrenia, while the opposite was true for men (Cowell et al. 1996). Larger ventricles were found in men but not women with schizophrenia compared to healthy controls (Nopoulos et al. 1997). Hemispheric asymmetry is also reduced in men with schizophrenia compared to controls (Crow 2002).

To sum up, ADHD and schizophrenia are two highly heritable sexually dimorphic psychiatric disorders. The reasons behind the greater burden of these disorders in men compared to women deserve to be investigated, as they could provide clues to the pathophysiology of the disorders.

1.2 Sexually dimorphic effects of genes on psychiatric disorders

Sex differences in psychiatric phenotypes are well recognised and there has been interest in identifying gene variants that may be involved in them.

Of the few examples available, *COMT* (Catechol-O-Methyltransferase) stands out. The gene encoding COMT, an enzyme catalysing the degradation of dopamine, adrenaline and noradrenaline, is on chromosome 22q11.2. Interest in *COMT* comes from its involvement in dopaminergic pathways. In normal brain, COMT activity in prefrontal cortex is 17% higher in men than women (Chen et al. 2004) despite similar mRNA and protein levels in both sexes (Chen et al. 2004). A *COMT* knockout mouse model provided even more evidence for a sexually dimorphic effect of *COMT*. Dopamine levels in the frontal cortex were increased threefold in male *COMT* knockout mice compared with wild type controls. This difference was not found in female mice. However, female *COMT* knockout mice showed more anxiety compared to wild type controls (Gogos et al. 1998).

The most studied polymorphism in *COMT* is a G/A transition found in codon 158, which results in a valine-to-methionine substitution (GenBank accession no. Z26491); Met(158) is coded by the low-activity (L allele), whereas Val(158) is coded by the high-activity variant (H allele). This polymorphism is functional and is associated with three- to four-fold differences in COMT activity between homozygous subjects with the val/val genotype and homozygous subjects with the met/met genotype (Lachman et al. 1996).

The sexually dimorphic effect of this *COMT* variant has been suggested in association studies of psychiatric disorders. The low activity allele was associated with Obsessive-Compulsive Disorder (OCD) in men in a meta-analysis ($p<0.001$, OR=1.88), while there was no effect in women ($p=0.83$, OR=0.98). There was also a significant interaction of this allele with gender ($p<0.0001$) (Pooley et al. 2007). The G allele of SNP rs165599 in *COMT* was found to be associated with schizophrenia in women ($p=6.8x10^{-6}$) (Shifman et al. 2002) but this finding was not replicated (Craddock et al. 2006b). Interestingly, there are reports of *COMT* association with anxiety (Stein et al. 2005) and panic disorder (Domschke et al. 2007) in women but not men. This could suggest that *COMT* predisposes to anxiety in both sexes, but other factors lead to the manifestation as anxiety disorder in women and OCD in men (Harrison and Tunbridge 2007). In ADHD, the *COMT* Val allele showed a nearly significant association ($p=0.054$) in the Cheuk and Wong (2006) meta-analysis for males only. A statistically significant gender effect for for the COMT Val allele association in ADHD was also found by pooling the results of two studies (Biederman et al. 2008; Qian et al. 2003).

Evidence suggesting another interesting sexually dimorphic gene came from a recent genome-wide association study for schizophrenia (Shifman et al. 2008). SNP rs7341475 in an intronic region of the *reelin* gene was associated with schizophrenia in women ($p=9.8x10^{-5}$) but not in men ($p=0.47$). A significant gene-sex interaction was also found ($p=1.8x10^{-4}$). To confirm the results, four other samples were used and although the predicted direction of effects was present in all of them, the results were significant for females in only one of the samples (Shifman et al. 2008). The *reelin* gene is expressed in the brain and it is reduced in the superficial interstitial white-

matter neurons in men but not women with schizophrenia (Eastwood and Harrison 2005).

Recent advances in identifying genes for psychiatric disorders have demonstrated that each susceptibility gene has a small effect in the risk of developing the disorder. There is no reason to doubt that this is the case for sex differences. This means that identifying genes contributing to sex differences will be even harder since any existing samples will have to be stratified by sex, further reducing the power to detect effects.

1.3 Sex differences in brain structure and function

The brain is a sexually dimorphic organ and its dimorphism is evident in both structural and in neurotransmitter system differences.

In terms of brain structure, sex differences exist in every brain lobe (Figure 1-2). The ratio of grey to white matter is different between the two sexes (Cahill 2006). The hippocampus, a brain region involved in learning and memory, is the most obviously sexually dimorphic brain region. It is larger in women than in men when adjusted for total brain size (Goldstein et al. 2001). Many neurotransmitter systems, such as serotonergic, adrenergic and cholinergic pathways within the hippocampus also show sex differences (Madeira and Lieberman 1995). Oestrogens can affect hippocampal cells by influencing their dendritic structure and they also modulate memory processes (Packard et al. 1996). The role of the hippocampus in learning is different between the two sexes (Rucker et al. 2004). It is interesting that the hippocampus reacts differently to chronic stress in the two sexes. Animal models showed that chronic stress causes damage to the hippocampus in males, but not in females (McEwen 2000). This resistance of female hippocampal cells to stress deserves to be studied in relation to female susceptibility to depression and anxiety disorder.

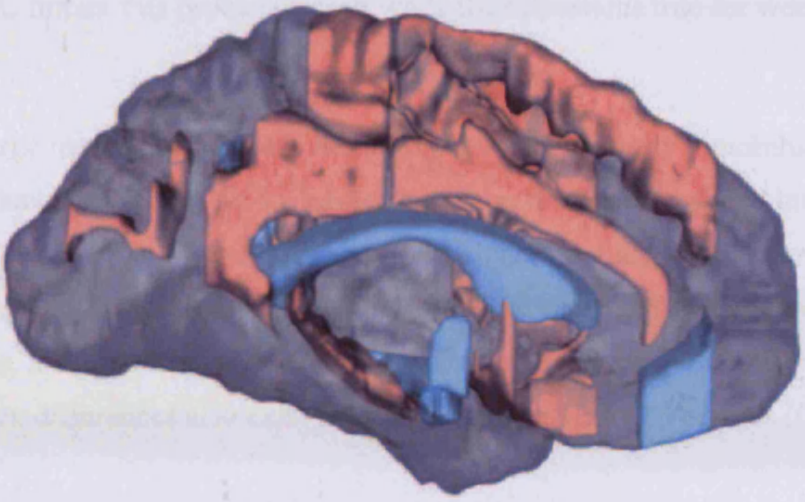

Figure 1-2. Sexually dimorphic brain structures. Data from Goldstein et al. (2001), where they measured the volume of brain regions relative to the cerebrum using MRI scans in 27 men and 21 women. Structures that are larger in women are shown in red and structures that are larger in men are shown in blue. Picture taken from Cahill (2006)

The amygdala, a brain region involved in emotional behaviour, is significantly larger in men than in women, when adjusted for total brain size (Goldstein et al. 2001). The amygdala is involved in the storage of memories of emotional events probably through interaction with stress hormones (McGaugh 2004). This function is sexually dimorphic in terms of hemispheric lateralisation. The left amygdala is involved in storage of emotional events in women while the right one performs the same function in men (Cahill et al. 2004). Hemispheric lateralisation of the amygdala according to sex has been found for reactivity to emotional facial expressions. The left amygdala was more active when women responded to happy faces, while the opposite was true for men (Killgore and Yurgelun-Todd 2004). This sexually dimorphic functional lateralisation of amygdala correlates with the pattern of amygdala dysregulation in certain diseases. In women with Turner's syndrome the left amygdala is less reactive to emotional faces (Skuse et al. 2005), while left amygdala activity is increased in women with depression (Drevets 2003).

The prefrontal cortex (PFC) is involved in working memory and decision making processes. It has also a high concentration of sex hormone, and especially oestrogen, receptors (Cahill 2006). In terms of decision making, it has been found that lesions in

the right PFC impair this process in men while the opposite is true for women (Tranel et al. 2005).

Finally, a large number of neurotransmitter systems are sexually dimorphic. Sex differences have been found in the serotonin system in both animal and human brain (Nishizawa et al. 1997). Monoamine levels are higher in many brain regions in women compared to men (Cahill 2006) and the locus coeruleus, which is very rich in monoamines, is more responsive to stress in female compared to male rats (Curtis et al. 2006). Sex differences also exist for vasopressin, GABA and opioids (Cahill 2006).

1.4 Causes of sex differences in brain and behaviour

It is generally accepted that sex differences exist not only in susceptibility to diseases but also in brain and behaviour. However, the factors that contribute to these sex differences and especially the extent to which each of them is responsible is always a debate. Hormones have been considered a primary candidate for a long time, since they are straightforward to study and can be very potent. Nevertheless, accumulating evidence has drawn attention to genetic factors contributing to sex differences and they not only include sex chromosome differences but also differences in gene expression. A microarray analysis of gene expression in mouse brain found that more than 600 genes are differentially expressed in brain according to sex (Yang et al. 2006). Adding to that are the consequences of being genetically male or female due to differential exposure and differential susceptibility to risk and environmental protective factors. Finally, it is highly unlikely that any of these factors acts in isolation; as Figure 1-3 shows, they interact making the delineation of specific factors contributing to sex differences difficult.

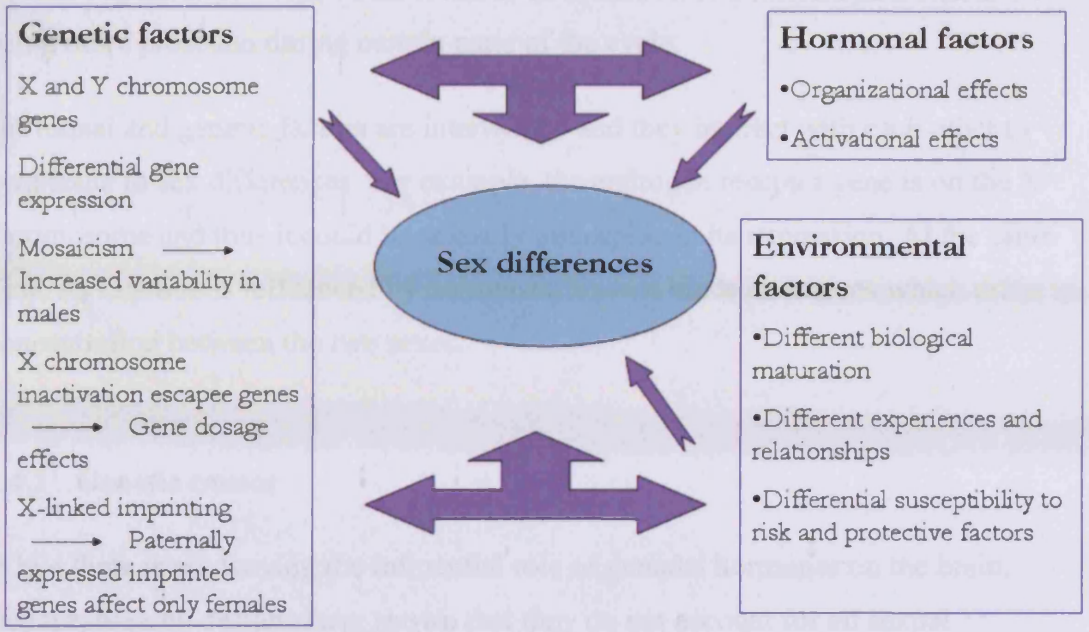

Figure 1-3. Interactions between genetic, hormonal and environmental factors contributing to sex differences

1.4.1 Hormonal effects

Hormones provide the classic example of factors contributing to sexual differentiation. This differentiation starts early in prenatal life and is determined by the presence or absence of the Y chromosome gene, Sex determining Region of chromosome Y (*SRY*). When *SRY* is present, the development of testes is initiated and testosterone is produced (Polanco and Koopman 2006). When testosterone reaches the brain, it is converted to estradiol by aromatase or to dihydrotestosterone (DHT) by 5α-reductase leading to masculinisation and defeminisation of the brain. In females the *SRY* gene is absent and so ovaries develop and produce oestrogens, but these bind to alpha-fetoprotein. The same happens to oestrogens produced by the mother, in order to protect the female foetus from the masculinising effect of oestrogens on the brain (Bakker et al. 2006). Hormones during development leave permanent "marks" on the brain, called organisational effects. Nonetheless, some of the hormonal effects can be reversed when hormones are not present and they are called activational effects. Of course, the effect of hormones does not stop in prenatal life. Levels of circulating hormones differ massively between the two sexes and especially in women they

fluctuate cyclically, which could result in sex differences in certain characteristics being more profound during certain parts of the cycle.

Hormonal and genetic factors are intertwined and they interact with each other to contribute to sex differences. For example, the androgen receptor gene is on the X chromosome and thus it could be sexually dimorphic in its expression. At the same time, its function is influenced by hormones, since it binds androgens which differ in concentration between the two sexes.

1.4.2 Genetic causes

While there is no denying the influential role of gonadal hormones on the brain, elegant research designs have shown that they do not account for all sexual dimorphism. The role of genetic factors in shaping the male and female patterns of brain function and behaviour can be as, or even more, important.

The obvious difference of the male and the female genome (with men having an X and a Y chromosome while women have two X chromosomes) would point to a potential involvement of sex chromosomes in sexual differentiation. Differential gene expression, not only from sex chromosome genes but autosomal genes as well, is important, since the phenotype is shaped by the level and pattern of gene expression. A phenomenon that has important implications in sex differences is X chromosome inactivation. The random inactivation of one of the two X chromosomes in all the cells of the female body compensates for the double dose of some X-linked genes in females compared to males. The random nature of inactivation means that females are mosaics of two populations of X chromosomes resulting in blunting of extreme phenotypes, which would be fully manifested in males. Finally, some X chromosome genes escape inactivation and thus there is a gene dosage difference between males and females.

1.4.3 The effect of sex chromosomes on brain

Despite hormones being considered the main candidate for sex differences, there is accumulating evidence that sex chromosome complement irrespective of the

hormonal milieu plays a vital role. X-linked and Y-linked genes are transcribed differently between the two sexes, since Y-linked genes are only expressed in males and at least some of the X-linked genes are transcribed more highly in females (Xu et al. 2002). Even when there is a Y-linked homologue for X chromosome genes, expression levels tend to be lower than for X chromosome genes (Figure 1-4).

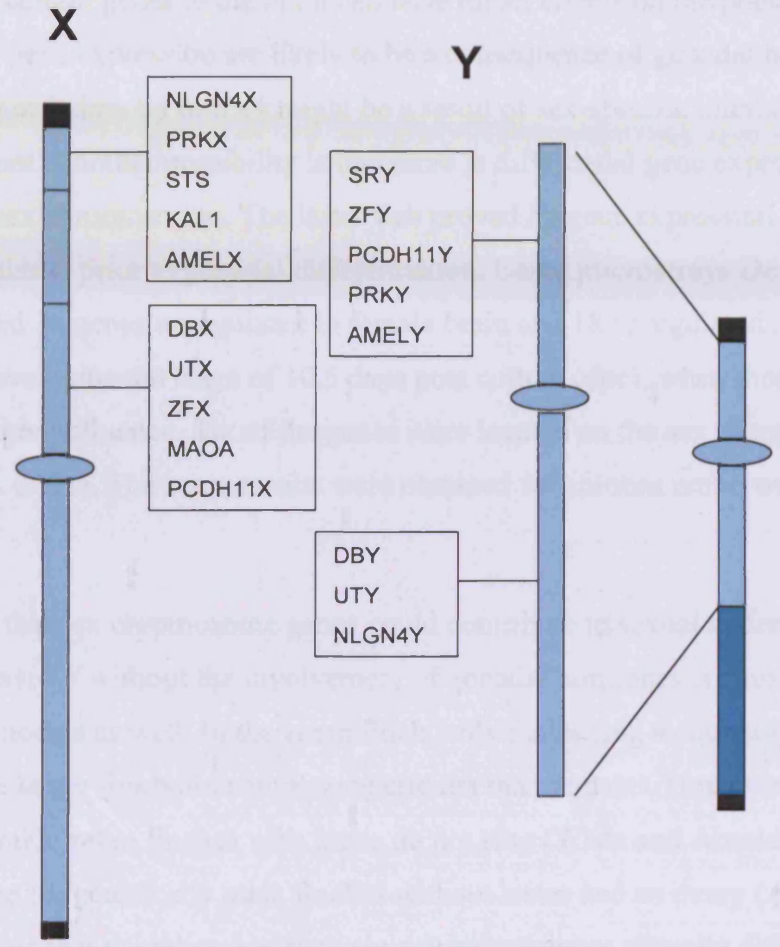

Figure 1-4. Diagram of sex chromosomes showing key genes with a homologue on both chromosomes. Pseudoautosomal regions are shown in black, and the heterochromatic region on the Y chromosome is shown in a darker colour. The euchromatic region of the Y chromosome is magnified (NLGN4: Neuroligin; PRK: Protein kinase; STS: Steroid sulfatase; KAL1: Kallmann syndrome 1 protein; AMEL: Amelogenin; DB: Developing brain homeobox; UT: Ubiquitously transcribed tetratricopeptide repeat gene; MAOA: Monoamine oxidase A; SRY: Sex-determining region on Y chromosome; ZF: Zinc-finger protein; PCDH11: Protocadherin 11

There is evidence that sex chromosomes influence the sexual dimorphism of brain and behaviour. Dissociate cultures from rodent mesencephalon exhibit sexually dimorphic characteristics before gonadal differentiation is initiated (Carruth et al. 2002). Gene expression in the brain also differs between the two sexes. In the adult mouse brain approximately 650 genes are differentially expressed in males and females. Most of

the genes that are over-expressed in females reside on the X chromosome while the over-expressed genes in male brain come from both sex chromosomes (Yang et al. 2006). Differential gene expression between the two sexes is crucial, since the level of expression of certain genes in the brain can have direct effects on the phenotype. The differences in gene expression are likely to be a consequence of gonadal hormones either directly or indirectly or they might be a result of sex-specific interactions with the environment. Another possibility is that there is differential gene expression linked to the sex chromosomes. The latter was proved by gene expression studies in mouse foetal tissue prior to gonadal differentiation. Using microarrays Dewing et al. (2003) detected 36 genes upregulated in female brain and 18 upregulated in male brain at the developmental stage of 10.5 days post coitum (dpc), when there is no gonadal hormone influence. Six of the genes were located on the sex chromosomes (Dewing et al. 2003). The same results were obtained for chicken embryos (Scholz et al. 2006).

The evidence that sex chromosome genes could contribute to sexual differentiation of brain and behaviour without the involvement of gonadal hormones is strengthened by other animal models as well. In the zebra finch, only males sing a courtship song and thus they have larger forebrain neural song circuits than females. However, genetically female zebra finches with testes do not sing (Wade and Arnold 1996). The opposite is true for genetically male finches without testes and an ovary (Arnold 1997). It appears that neural song circuits in zebra finches are sexually differentiated due to the action of factors other than gonadal hormones, most likely sex chromosomes. This notion is supported by studies on a gynandromorphic finch which is genetically male on the right side of the body and genetically female on the left side. Although both sides of the brain are exposed to the same hormonal environment, the neural song circuit of the right side is more masculine than that of the left side of the brain (Agate et al. 2003).

1.4.3.1 X chromosome

The sex chromosomes have evolved from a pair of autosomes. The emergence of a gene responsible for testes differentiation restricted the presence of the Y chromosome in males. The consequence of this was that the Y chromosome

developed a set of genes that were advantageous for males without the need for any advantage for females. The presence of genes involved in spermatogenesis on the Y chromosome supports this notion. The Y chromosome is also involved in sex differences and the evidence pointing to its involvement is discussed in Chapter 6, since it is the main topic of this thesis.

The X chromosome is a different story. It is found in only one copy in men and this means that alleles conferring an advantage to males would be expressed even if recessive. Females will be protected from the effect of these alleles by the presence of a dominant allele. Since any gene that is "bad" for females could not survive on the X chromosome, it has a higher proportion of female-biased genes (Arnold 2004).

The human X chromosome contains more than 800 genes and has an important contribution to brain function. Firstly, the number of X chromosome genes expressed in brain is disproportionately high with 40% of X chromosome genes expressed there (Ropers and Hamel 2005). The importance of X chromosome genes for brain function is also highlighted by the fact that expression of the single active X chromosome is upregulated in the brain so that it is similar to the level of expression of autosomes (Nguyen and Disteche 2006). In addition, the X chromosome is considered to be involved in cognitive ability. Compared to the autosomes it contains a significantly higher number of genes implicated in mental retardation. The fact that most of the X chromosome genes that are expressed in the brain are also conserved in distantly related organisms makes it possible that they play an important role in brain function (Zechner et al. 2001). It has been hypothesised that female sexual selection of intelligent men contributes to that (Zechner et al. 2001).

1.4.3.1.1 X chromosome inactivation

The fact that females possess two X chromosomes while males possess only one creates a need to balance expression between the two sexes. This is achieved by X chromosome inactivation, a mechanism for silencing gene transcription of one of the two X chromosomes in females. This takes place in all somatic cells in a random fashion, so that females have a mixture of two different populations of X chromosomes in their cells. A non-coding RNA gene called *XIST* (X-inactive specific transcript) is responsible for silencing X chromosomes. *XIST* is expressed only from the inactivated chromosome and it covers the chromosome in a *cis* interaction. Histone modifications and DNA methylation ensure that no transcription is possible from this chromosome and it remains inactive in all subsequent cell divisions (Xu and Disteche 2006). X chromosome inactivation can have important implications for sex differences, since females have two different populations of X chromosomes resulting in mosaicism, which will be discussed later. In addition, some genes escape inactivation and, as a consequence, females will have increased expression of them compared to males. X chromosome inactivation varies according to tissue type and developmental stage with important consequences for differences in phenotypes between men and women.

1.4.4 Gene dosage effects

Taking into account the balancing effect of X chromosome inactivation, expression should be the same in the two sexes. However, this is not true in two instances. First of all, during prenatal life after the paternal imprints are wiped and before X inactivation starts, both X chromosomes are expressed in females. It is possible that this imbalance is responsible for the higher rate of growth in female embryos (Xu and Disteche 2006).

Secondly, not all genes on the X chromosome are inactivated. It is thought that up to 15% of human X chromosome genes escape inactivation (Carrel and Willard 2005). Because of these X escapees, expression of some genes would be higher in females compared to males possibly contributing to sexual dimorphism. It is not yet clear how many genes escape inactivation in the brain, since inactivation varies according to

tissue type and developmental stage (Brown and Greally 2003). It is also possible that expression is compensated by the presence of a Y-linked homologue (see Figure 1-4 for some of the Y-linked homologues). In mouse, all X escapees have a Y-linked homologue but in human there are X escapees without a Y-linked homologue (Xu and Disteche 2006). However, this homologue is not always able to compensate for the increased expression in females. In a study of X escapees with a Y-homologous gene in mouse brain, Xu et al. (2002) found that in five cases expression of Y-homologues in males was not sufficient to reach equal levels of expression to females. In three cases genes were differentially regulated indicating that they might not be functionally equivalent (Xu et al. 2002).

The fact that there is increased expression of some X chromosome genes in females compared to males does not necessarily mean that the protein level would be higher. Indeed, there are reports that protein levels of X escapees in mouse brain did not differ between males and females, although there was a significant difference in mRNA levels (Xu et al. 2006).

A "natural experiment" of the involvement of X escapees in brain function is possible with women affected by Turner's syndrome (45,XO karyotype). The syndrome is characterised by short stature, ovarian degeneration and interestingly neuropsychological impairments with problems in visual perception and facial emotion recognition, as well as anxiety (Ross et al. 2006). Not all of these problems are alleviated with oestrogen treatment. Impairments are also found in memory, attention and social interaction, and women with Turner's syndrome are more susceptible to ADHD (Russell et al. 2006) and autism (Skuse 2000). This susceptibility to classically male-biased disorders poses the question whether haploinsufficiency of X chromosome genes that usually escape inactivation (a common characteristic between men and women with Turner's syndrome) is implicated in the increased risk for neuropsychiatric disorders (Davies and Wilkinson 2006). Brain imaging has shown that the volumes of amygdala and hippocampus are enlarged in women with Turner's syndrome compared to healthy controls (Davies and Wilkinson 2006). Good et al. (2003) performed deletion mapping analysis and considered that haploinsufficiency of genes in a small region on Xp11.3 is responsible for the enlarged brain volumes.

Animal models of Turner's syndrome have shown that 39,XO mice are more fearful than controls and they also show response accuracy and reaction time deficits (Davies et al. 2005a). To examine the involvement of X escapees in behaviour, female mice carrying an X chromosome and a Y^{*X} chromosome, consisting of an X centromere and a very small X-specific region immediately adjacent, were generated. Interestingly, the behavioural phenotype was rescued in these mice. The authors suggested that gene dosage by an X escapee gene expressed in the brain was responsible for the results (Davies et al. 2005a). Their most likely candidate was *Sts*, a PAR (pseudoautosomal region) gene encoding the enzyme steroid sulfatase. In humans, the *STS* gene is located on the X chromosome outside the PAR region and it escapes inactivation (Li et al. 1996). There is a Y-linked homologue but it constitutes a pseudogene (Yen et al. 1988). Deletion of *STS* leads to X-linked ichthyosis, a metabolic syndrome. Individuals with X-linked ichthyosis have a higher prevalence of ADHD and autism, which implies that *STS* could be implicated in attentional mechanisms in humans (Kent et al. 2008). In a recent study, ADHD was associated with polymorphisms within the *STS* gene (Brookes et al. 2008). The fact that *STS* escapes X-inactivation and that it has no functional Y homologue means that, in theory at least, it should be expressed approximately twice as highly in female brain than male brain. This potential sex difference in *STS* expression could explain why males and females are differentially vulnerable to disorders of attention and impulse control such as ADHD and pathological gambling.

1.4.5 Mosaicism

A consequence of the random nature of X chromosome inactivation is mosaicism. If an X chromosome gene is polymorphic, female organs are a mosaic of cells where two different alleles are expressed. Given the large number of X chromosome genes expressed in the brain, mosaicism could contribute to sex differences in brain function. Because males are hemizygous for all alleles on the X chromosome, they are more likely to be influenced by alleles associated with extreme phenotypes (either showing exceptionally high or low abilities). Extreme phenotypes become blunted in females. This could be contributing to the increased variability of cognitive performance in males (Skuse 2005).

1.4.6 X-linked imprinting

Imprinting is a mode of inheritance which does not involve the DNA sequence. Certain genes are epigenetically marked so that they are only expressed if they are transmitted from the mother or the father. Imprinted genes are expressed in the brain and they can be involved in mental disorders (Davies et al. 2005b).

Imprinting can result in sex differences, if it affects the X chromosome. Since only females inherit an X chromosome from the father, paternally expressed X-linked imprinted genes would only affect one sex. On the other hand, maternally expressed X-linked imprinted genes would be expressed in both sexes but their level of expression would be reduced in females due to random X chromosome inactivation. The first study to link X chromosome imprinting with mental disorders stratified women with Turner's syndrome according to whether they had inherited their X chromosome from the father or the mother and compared their cognitive profile. It found that when the X chromosome was maternally inherited (as is the case with males as well) there was impairment in social cognition and a higher susceptibility to autism (Skuse et al. 1997). These results are in agreement with the better social skills and reduced autism susceptibility in females (Skuse 1999). Mouse models provide support for the contribution of X-linked imprinted genes to sexually dimorphic neurobiological phenotypes (Davies et al. 2005a). Nevertheless, no X-linked imprinted genes have so far been discovered on the human X chromosome.

1.4.7 Environmental factors

Apart from purely biological factors, environmental factors also contribute to sexual dimorphism in behaviour. Males and females are exposed to different environmental factors and also have different susceptibility to risk and protective factors.

Firstly, biological maturation is different in the two sexes. Boys develop slower than girls and puberty occurs approximately two years later in boys compared to girls. The consequences are that boys are asked to take part in demanding tasks for which they are not developmentally ready. For girls, the early development of secondary sex

characteristics results in different social responses from other people compared to boys of the same age (Rutter et al. 2003).

Secondly, the two sexes have different experiences which are either more common or unique to one sex. Females experience pregnancy and childbirth and these experiences, coupled with the accompanying hormonal effects, are contributing to postpartum mental disorders (Rutter et al. 2003). On the other hand, males are more likely to be involved in criminal activity and suffer physical attacks, while females are more likely to suffer sexual abuse (Rutter et al. 2003). Social and cultural factors increase differential experiences. Parents, siblings and teachers respond differently according to the sex of the child. Friendship patterns differ between the two sexes.

Thirdly, it is important to realize that risk and protective factors do not affect the two sexes equally. Males are more vulnerable to physical adversities, such as infections, obstetric complications and malnutrition. The higher mortality of male infants balances the greater rate of male conceptions to create an equal sex ratio. In terms of parent-child relationships, there is a stronger association of mother-child relationship and depression in females than in males (Veijola et al. 1998) and there is evidence, although not strong, that coercive parenting can lead to more externalising behaviours in boys (McFadyen-Ketchum et al. 1996). Problems in relationships with other people can be more damaging for women, while men are more sensitive to unemployment and work-related problems (Rutter et al. 2003). Marriage is more protective against depression for men than for women (Kiecolt-Glaser and Newton 2001).

It is interesting that biological mechanisms explaining susceptibility to specific risk factors have started to emerge. For example, it was found that a polymorphism on the X-linked gene *MAOA* (Monoamine Oxidase) increases *MAOA* activity in the brain and makes boys more resilient to antisocial behaviour following maltreatment (Caspi et al. 2002). Epigenetics can also be the mediating mechanism in the differential susceptibility to environmental factors. Evidence shows that early psychosocial adversity can have an effect on DNA methylation and thus alter gene expression (Mill and Petronis 2008). It is tempting to hypothesise that this could be the mediating mechanism for the effect of environmental adversity on ADHD.

Despite the fact that sex differences in psychopathology are well-established, research on the factors that lead to this sexual dimorphism has been limited. Hormonal effects, genetic factors relating not only to the sex chromosomes, but also to phenomena such as X chromosome inactivation, and environmental factors contribute to sex differences but the extent that each factor is involved and the way it interacts with other factors is still not know. However, more research is required in this field because it can provide important information about the pathways involved in psychiatric disorders.

Chapter 2 Attention deficit hyperactivity disorder (ADHD)

2.1 ADHD symptoms and diagnosis

Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood-onset psychiatric disorders affecting up to 6% of children. The behaviour of children with ADHD is characterised by impulsiveness, lack of concentration and overactivity. This pattern of behaviour puts children at risk of developing antisocial behaviour, receiving poor education and is very disruptive for family and teachers.

2.1.1 Epidemiology

The reported prevalence of ADHD ranges from 1.5-6% depending on the population and the method of diagnosis (Cuffe et al. 2005; Polanczyk et al. 2007). The prevalence of ADHD and especially treatment in the UK tends to be lower than in the USA, although it is still higher than other European countries (Rutter et al. 2008).

There is a significantly increased prevalence in males compared to females. The sex ratio ranges from 1:4 in population samples to 1:6 in clinical samples. See Chapter 1 for more information on sex differences in ADHD.

2.1.2 Clinical presentation

The three main characteristics of children with ADHD are inattention, impulsiveness and overactivity.

Inattention involves lack of concentration and persistence, leading to inability to complete tasks. The degree and presentation of inattention varies with age and it is a symptom that puts children at risk of educational failure. It is also the most salient aspect of the disorder as children with ADHD enter adulthood (Millstein et al. 1997). Attention can be improved with immediate rewards and sufferers can show sustained attention in certain tasks, such as computer games.

Impulsiveness involves acting without thinking about the consequences. It leads to children putting the health and safety of themselves and others in danger. Severe impulsiveness can be very difficult for people to tolerate, although it is difficult (especially in adolescents) to distinguish oppositional behaviour from impulsiveness.

Overactivity is excessive movement and that can present itself as restlessness and fidgetiness. Although it is usually the most noticeable feature of children with ADHD, it does not usually persist into adolescence and adulthood.

2.1.3 ADHD diagnosis

The diagnosis of ADHD relies on reported symptoms. The use of internationally accepted diagnostic criteria has increased diagnostic reliability resulting in ADHD being one of the best-validated clinical diagnoses (Faraone et al. 2005). There are two sets of diagnostic criteria used to define ADHD: the Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (DSM-IV-TR) and the International Classification of Diseases, tenth edition (ICD-10). DSM-IV-TR was published by the American Psychiatric Association in 2000 (American Psychiatric Association 2000) and ICD was first published by the World Health Organisation in 1993 and has since been revised many times (World Health Organisation 1993). Both sets of diagnostic criteria are based on the same symptoms but they describe the disorder differently.

DSM-IV-TR uses the term ADHD and bases diagnosis on two categories of symptoms: inattention and hyperactivity-impulsivity (see Table 2-1 for a list of symptoms). It recognises three ADHD subtypes. The combined type is diagnosed when six or more symptoms from each symptom category are present, while the inattentive or overactive-impulsive type is diagnosed when six or more symptoms from only one of the two categories are present. All types require that symptoms are present for at least six months and before the age of seven and they are developmentally inappropriate. Symptoms must also be associated with impaired functioning in two or more settings.

ICD-10 uses the term hyperkinetic disorder and does not recognise subtypes. It bases diagnosis on three symptom categories: inattention, hyperactivity and impulsivity. It requires six or more symptoms of inattention, three or more symptoms of

hyperactivity and one or more symptom of impulsivity for a diagnosis to be reached. It also requires that symptoms are present before the age of seven and they are developmentally inappropriate. However, in contrast to DSM-IV-TR, symptoms (and not only impaired functioning) should be present in two different settings, usually both at home and at school. The other difference between the two sets of diagnostic criteria is that in ICD-10, hyperkinetic disorder is excluded if other disorders, such as autism and anxiety are present, while in DSM-IV-TR, ADHD is excluded only if the symptoms are better explained by a co-existing disorder.

Hyperkinetic disorder, according to ICD-10, can be considered as a subgroup of ADHD and it is usually a narrower phenotype than ADHD combined type. Children with hyperkinetic disorder are also at higher risk of neurodevelopmental abnormalities such as language problems (Taylor and Sonuga-Barke 2008) suggesting that hyperkinetic disorder indicates a more extensive brain impairment than ADHD. For research purposes DSM-IV-TR diagnoses tend to be used more often, usually requiring both parents and teachers reports. Individuals with autism, learning disabilities and depressive disorders are usually excluded. It should be noted that both sets of diagnostic criteria merely describe the disorder and there is no reason to believe that the underlying aetiology is different.

Although subdividing ADHD into subtypes can be useful for diagnostic purposes, it is not clear whether they are stable over time (Todd et al. 2008) or whether there are distinct genetic and environmental factors acting upon the different subtypes (McLoughlin et al. 2007; Thapar et al. 2006). Clinically, ADHD is usually defined categorically. For research purposes, ADHD is frequently treated as a dimension showing that genetic factors are important for normal variation in ADHD scores as well as extreme scores (Derks et al. 2008; Thapar et al. 2006). In both cases, it has been shown that ADHD is highly heritable.

Table 2-1. ADHD symptoms from the DSM-IV research diagnostic criteria

A. Inattention (six or more symptoms required)
Often fails to give close attention to details or makes careless mistakes in school work, work or other activities
Often has difficulty sustaining attention in tasks or play activities
Often does not seem to listen when spoken to directly
Often does not follow through on instructions and fails to finish schoolwork, chores or duties in workplace (not to oppositional behaviour or failure to understand)
Often has difficulty organising tasks and activities
Often avoids, dislikes or is reluctant to engage in tasks that require sustained mental effort
Often loses things necessary for tasks or activities
Is often easily distracted by extraneous stimuli
Is often forgetful in daily activities
B. Hyperactivity-impulsivity (six or more symptoms required)
Hyperactivity
Often fidgets with hands or feet or squirms in seat
Often leaves seat in classroom or in other situations in which remaining seated is expected
Often runs about or climbs excessively in situations in which it is inappropriate (subjective feelings of restlessness for adolescents or adults)
Often has difficulty playing or engaging in leisure activities quietly
If often 'on the go' or often acts as if 'driven by a motor'
Often talks excessively
Impulsivity
Often blurts out answers before questions have been completed
Often has difficulty awaiting turn
Often interrupts or intrudes on others (e.g. butts into conversations or games)
Symptoms must have persisted for at least six months, must be developmentally inappropriate and present before the age of 7. Symptoms are of a degree that is maladaptive, associated with impairment in functioning that is present in two or more settings

2.1.4 Developmental course

Half of children with ADHD will still meet full diagnostic criteria in adolescence (Klein and Mannuzza 1991). The levels of overactivity and impulsiveness tend to decrease, while inattentiveness is more persistent. During adolescence the problems of antisocial behaviour and substance abuse also emerge. Longitudinal studies have shown that there is persistence of symptoms into adulthood (Taylor and Sonuga-Barke 2008). Comorbid antisocial behaviour also tends to persist. Twin studies have shown that there are genetic influences not only on ADHD but also on the continuity of the disorder (Larsson et al. 2006) and there is the suggestion that persistent ADHD may index a stronger family history (Faraone et al. 2005).

2.1.5 Comorbidity

ADHD is often comorbid with other disorders especially of a neurodevelopmental nature but not limited to that. Conduct disorder and anxiety can also co-exist with ADHD. It is striking that two thirds of children with ADHD have a diagnosis of another psychiatric disorder as well.

Although the symptoms of ADHD and autism differ, the disorders often co-exist. ADHD can also co-exist with Tourette's syndrome. One common characteristic of neurodevelopmental disorders is that they tend to affect more commonly males (Costa e Silva 2008; Cuffe et al. 2005) and they are of childhood-onset. This is especially true for ADHD and autism. For this reason, it has been suggested that there might be a set of common risk factors affecting brain development during critical stages. Twin studies indeed do find that ADHD and autism symptoms are influenced by a shared inherited liability (Ronald et al. 2008). Further genetic or environmental factors are then responsible for the manifestation of symptoms leading to diagnosis of specific disorders. It is striking that children with autism are excluded from having a diagnosis of hyperkinetic disorder according to ICD-10 criteria, although they could benefit from treatment for their ADHD symptoms. They are also excluded from research designs despite the fact that they could represent one extreme of the distribution of neurodevelopmental disorders.

Specific learning disabilities such as dyslexia are also very common in ADHD with up to 40 % of children with ADHD presenting with them (Willcutt et al. 2000). The same is true for intellectual disability. Children with ADHD and intellectual disability (IQ<70) are excluded from most research studies, thus there is not enough knowledge about the aetiology of their disorder. Problems in keeping up with learning at school can trigger attention difficulties or it could be that lack of attention causes children to have problems learning. The other possibility is that they represent a different group which should be studied separately.

The conditions that show the greatest comorbidity with ADHD are Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD). Their prevalence in children with ADHD ranges from 20% to 40% (Goldman et al. 1998; Thapar et al. 2001) and they represent a risk factor for poor outcome for children with ADHD. It has been considered that the group of children with ADHD and childhood-onset CD is distinct from the groups of children with ADHD or CD only. In this group of children, ADHD is more severe and shows greater persistence in adult life (Moffitt 1993). Family history is also stronger (Faraone et al. 2005) and the presence of CD indicates higher genetic loading in ADHD (Thapar et al. 2001). Early onset CD also represents a risk factor for poor outcome for children with ADHD (Taylor et al. 1996). Specific genetic factors have been found to operate only in children with ADHD and early-onset CD (Caspi et al. 2008; Thapar et al. 2005), as discussed later.

Anxiety and mood disorders are also found to accompany ADHD at a rate of about 20% (Cuffe et al. 2001; Pliszka 2000). In addition, children with ADHD are more prone to accidents and injuries and are more likely to have worse academic performance (Cuffe et al. 2005).

During adolescence and adulthood, ADHD co-exists with a number of psychiatric disorders. Substance abuse is more common in adolescents with ADHD than in unaffected individuals (Taylor and Sonuga-Barke 2008), although this could be a result of the comorbid CD. It has been suggested that bipolar disorder is more common in adults with ADHD. However, the issue is whether symptoms of hyperactivity and irritability that characterise ADHD are misinterpreted as manic episodes that indicate bipolar disorder (Taylor and Sonuga-Barke 2008). Adults with ADHD have a higher incidence of motor accidents; they are more likely to change

jobs, smoke and use substances or alcohol (Cuffe et al. 2005). It is interesting that a survey in prison populations showed that 45% of incarcerated young adults had a history of or had current ADHD (Rosler et al. 2004).

2.2 Brain structure and function in ADHD

Studies of the structure and the function of the brain in ADHD have provided evidence of abnormalities in different brain regions as well as neurocognitive deficits and dysregulation in neurotransmitter pathways. It is not easy to say whether these deficits occur before the onset of disorder, since ADHD has a very early age of onset but they provide evidence of the neurodevelopmental nature of the disorder.

The brain volume of children with ADHD is reduced by 5% with both white and grey matter affected (Seidman et al. 2005). Based on similarities between people with ADHD and those with disorders caused by frontal lobe lesions (Taylor and Sonuga-Barke 2008), research has focused on the prefrontal cortex and striatum. Functional Magnetic Resonance Imaging (fMRI) studies comparing patients with ADHD and controls have found reduced activation in the prefrontal cortex (Durston et al. 2004) and the neostriatum (Vaidya et al. 2005) in patients. Reduced glucose metabolism has been shown by Positron Emission Tomography (PET) (Schweitzer et al. 2003). Structural differences have also been found in the corpus callosum (Seidman et al. 2005), the cerebellum (Castellanos et al. 2002) and the amygdala (Plessen et al. 2006). fMRI studies have contributed to evidence for altered function in the cerebellum and parietal and temporal lobes (Vaidya et al. 2005).

Individuals affected by ADHD show neurocognitive deficits. It is considered that executive function is their main weakness. Executive functions are neurocognitive processes aiming to maintain relevant information so that it can be easily accessed and used during problem solving. They include working memory, response inhibition, vigilance, planning and set-shifting amongst others. It has been proposed that executive control weaknesses are the primary deficit necessary to cause ADHD (Castellanos and Tannock 2002; Schachar et al. 2000). In a meta-analysis of 83 studies on executive functions and ADHD, it was found that children with ADHD performed significantly worse on all executive function tasks compared to controls

(controlling for intelligence and comorbidity) (Willcutt et al. 2000). Strongest effects were obtained for response inhibition, vigilance and spatial working memory but effect sizes were moderate. However, a large study assessing executive functions showed that only half of the children with ADHD exhibit an impairment in any task of executive function (Nigg et al. 2005). Furthermore, the variance of ADHD symptoms cannot be explained by executive function impairment, since ADHD symptoms are only weakly correlated with executive function task performance (Willcutt et al. 2001). In summary, executive function impairments are significantly associated with ADHD and this is true for both clinical and population-based samples (Willcutt et al. 2005). However, it is highly unlikely that they are the primary deficit causing ADHD.

There is evidence that dopaminergic pathways are implicated in the pathophysiology of ADHD. The most robust evidence comes from studies of medication used to treat ADHD. Stimulant medication, such as methylphenidate and amphetamine, reduces ADHD symptoms by increasing the availability of extracellular dopamine and noradrenaline (Volkow and Swanson 2003). Neuroimaging studies have also provided evidence that dopaminergic regions are implicated in ADHD (Spencer et al. 2005). It is very interesting that animal studies, such as work on the *DAT* knockout mouse which manifests ADHD-like behaviour, have also implicated dopaminergic systems (Gainetdinov 2007). For this reason, a number of genes involved in the dopaminergic pathway have been studied using association study designs, as will be discussed later.

ADHD is a complex disorder and it involves impairments in different brain substrates but it can be very difficult to delineate whether the impairments are causal or are the result of shared risk factors. However, the pattern of impairment in different areas coupled with the early age of onset points to a neurodevelopmental origin. Brain abnormalities can start early in life, probably during prenatal life, in individuals already susceptible to ADHD triggered by an environmental insult.

2.3 Evidence of genetic component to ADHD

Heritability of ADHD is high in the range of 75-91%. This has been proven using family, twin and adoption studies.

2.3.1 Family studies

Family studies examine the rate of a disorder in relatives of individuals with the particular disorder compared to relatives of healthy or individuals with another disorder. The design of these studies utilises parents of children with a disorder or offspring from parents with a disorder. A number of family studies of ADHD have reported increased risk of ADHD among parents and siblings of children with ADHD. The relative risk in first degree relatives is considered to be between 4 and 5.4 (Faraone et al. 2005). It is very interesting that half siblings of children with ADHD have a lower risk of developing the disorder than full siblings, further highlighting the important role of the genetic background in ADHD (Goodman and Stevenson 1989).

2.3.2 Adoption studies

Adoption studies are suitable for disentangling genetic and environmental factors contributing to ADHD. In adoption studies the risk of developing the disorder is compared in biologically related relatives and non-biologically related (adoptive family) relatives. The fact that all adoption studies have shown that the risk is greater in biologically related relatives provides strong evidence for a genetic contribution to ADHD (Cantwell 1975; Sprich et al. 2000).

2.3.3 Twin studies

Twin studies examine the heritability in genetically identical or monozygotic (MZ) and non-identical or dizygotic (DZ) twins. The extent of disease concordance in the two twin types is used to calculate heritability. Their advantage lies in the fact that they have generally used population-based and not clinical samples. They have all shown a high degree of heritability and indicated that 60-90% of variance in ADHD

can be explained by genetic factors (Faraone et al. 2005; Thapar et al. 1999). It is striking that the heritability is not 100%, suggesting contribution from non-shared environmental factors, epigenetic effects, stochastic effects or measurement inaccuracies.

2.4 Identifying complex disorder susceptibility genes

Before discussing the molecular genetic studies of ADHD, the methods used in molecular genetic studies of complex disorders should be introduced.

Psychiatric disorders are considered complex disorders, which means that, although they are genetic disorders following Mendel's laws, they are the result of multiple gene co-action, gene–gene and gene–environment interactions. As a result, identifying susceptibility genes for psychiatric disorders is challenging, but it has the potential to provide an understanding of the largely unknown pathophysiology of these disorders. This process has been based on the DNA variation that normally exists in different individuals without necessarily being pathogenic. This variation provides polymorphic genetic markers. The most widely used genetic markers are Single Nucleotide Polymorphisms (SNPs). The use of SNPs has revolutionised the field because there are large numbers of them covering a substantial part of the genome, and their genotyping is cheaper and more robust than with other markers. Advances in technology have made mass genotyping of large numbers of SNPs for large samples possible and more economical. Information (which is still growing) about a large number of SNPs can be found in a public database called HapMap (International HapMap Consortium http://www.hapmap.org/). There are three methods for identifying susceptibility genes for psychiatric disorders (Thapar and Stergiakouli 2008):

- genome-wide linkage studies
- candidate gene association studies
- Genome-Wide Association Studies (GWAS)

2.4.1 Genome-wide linkage studies

2.4.1.1 Linkage analysis principle

Linkage studies are based on the phenomenon of genetic linkage which is the result of recombination. Recombination occurs during meiosis when homologous pairs of chromosomes form chiasmata or crossovers and exchange genetic material. Thus, recombination produces new chromosomes where maternal and paternal loci are reshuffled to produce new combinations which are different to parental alleles (Strachan and Reed 2003). The importance of recombination is attributed to this reshuffling, which increases diversity in the human species.

If two loci are on different chromosomes or further apart on the same chromosome, the probability that their alleles will be inherited together is 0.5, according to Mendel's law of independent assortment. However, when two loci are near to each other, they are more likely to be inherited together because a crossover would be unlikely to form between them. This departure from the law of independent assortment is called genetic linkage (Strachan and Reed 2003).

The principle of linkage studies is that, if an allele of a marker is linked to a susceptibility locus, affected relatives will share this allele more often than expected. So when a genetic marker is co-inherited with a disease phenotype within many independent families or over many generations in an extended pedigree, there is an indication that susceptibility loci could be on the same chromosomal region. Genome-wide linkage studies aim to identify broad regions in the genome that could harbour susceptibility genes. Linkage analysis in large multiplex families is known as parametric analysis because it requires the specification of a genetic model. For this reason, it is challenging to apply to complex disorders, where the mode of inheritance is unknown. Non-parametric analysis, such as the Affected Sibling Pair (ASP) design, is more suitable. In this approach, instead of comparing affected and unaffected family members, allele sharing is only examined in affected siblings (Sham and McGuffin 2002).

2.4.1.2 Limitations of linkage studies

Linkage studies have been very successful in identifying genes that cause Mendelian traits. They have also been applied in complex disorders including psychiatric disorders. However, fine mapping of their results had limited success in identifying susceptibility genes for psychiatric disorders. One of the reasons is that linkage studies are better at picking up regions that harbour genes of larger effect sizes and not genes of modest or small effects, which are the rule in psychiatric genetics (Plomin et al. 2008). In addition, the set of markers used for linkage studies might be not dense enough to extract complete inheritance information. Finally, larger sample sizes would certainly be required to increase success, although for most complex disorders it is extremely difficult to collect thousands of affected sibling pairs (Hirschhorn and Daly 2005).

2.4.2 Candidate gene association studies

Candidate gene association studies are based on selecting specific genes based on evidence from other research findings and then comparing the frequencies of the marker alleles or genotypes in two different groups: one group of affected individuals (cases) and one group of matched healthy individuals (controls). There are two types of candidate gene studies:

• Positional – where the genes are selected following significant findings from linkage studies or by studying families who present with a disorder and a specific chromosomal rearrangement.

• Functional – when selection is based on assumptions about the pathophysiology of the disorder derived from animal, pharmacological, or imaging studies.

Before discussing the different types of association study designs, it is important to introduce the concepts of linkage disequilibrium (LD) and marker selection.

2.4.2.1 Linkage Disequilibrium (LD)

When a significant association is found between a SNP and a disease, it could be because this SNP is the causal variant (direct association) or it could be because it is correlated with the causal variant (indirect association). This non-random correlation of two loci that are inherited together for many generations is called Linkage Disequilibrium (LD) (Cardon and Bell 2001).

The degree of LD between two loci can be evaluated by using one of the two measures, D' or r^2. For both measures a value of 0 indicates no correlation between two loci and a value of 1 indicates that loci are co-inherited and are in complete LD. They are both based on the pairwise-disequilibrium coefficient D, which is a measure of covariance between two loci. It should be noted that the limitation of D' is that when a difference in allele frequency between two markers exists, a $D'=1$ can occur even though the two loci are not always co-inherited. Thus, $r^2=1$ is a better indicator of a perfect correlation between two markers (Mueller 2004).

2.4.2.2 Marker selection

When selecting markers for an association study of a particular gene or chromosomal region, knowledge of the LD structure of this gene or region is important. When the LD structure is known, it is possible to select "tag markers" which account for all the known variation at a particular chromosomal region. The HapMap (www.hapmap.org/) is a public database created by the International Haplotype Map Consortium (International HapMap Consortium 2003). It contains information, including LD, on 3.1 million SNPs (phase II) which have been genotyped in 270 individuals from four different populations.

Selection of tag SNPs depends on the Minor Allele Frequency (MAF) and degree of LD between the markers. It is common to use a pairwise tagging method, where markers are selected based on LD between each pair, and setting MAF at ≥ 0.05 and $r^2>0.8$. The tagging approach is advantageous for association studies, since it reduces cost. There is no need to genotype all polymorphisms in the region, so fewer markers are genotyped without losing information. In addition, false negative rates are

reduced, since all known genetic variation is assayed. The main limitation of the tagging approach is that rare variants are not covered.

2.4.3 Types of association study designs

There are two types of association study designs: family based and case-control association studies.

2.4.3.1 Case-control association studies

Case-control association studies involve comparing the frequency of alleles in a sample of affected individuals (cases) with that of a sample of unaffected individuals (controls). The main advantage of this design is that it is easier to collect larger samples because there is no requirement for family members and thus power is increased. However, population stratification can be an issue. In this case, differences in allele frequency between cases and controls appear but they are due to the populations not being ethnically matched rather than the alleles being causal (Cardon and Palmer 2003).

2.4.3.2 Family-based association studies

Family-based association studies use family members of affected individuals as controls, thus overcoming the problem of population stratification. Alleles are associated when they are transmitted more often than expected by chance from a heterozygous parent to an affected offspring. Although with family-based association studies population stratification is not an issue, the requirement of parents reduces the number of individuals available for ascertainment, especially for disorders with late onset. In addition, because only heterozygous parents provide informative transmissions, some genotyping information is discarded (Cardon and Bell 2001).

2.4.4 Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) are now the most promising method for systematically searching for susceptibility genes. They are basically association studies examining most of the genome for susceptibility variants. In genome-wide association studies usually several thousand or millions of SNPs are tested across the genome without an *a priori* hypothesis. For this reason they are suited to psychiatric disorders where the pathophysiology is not yet understood. Another advantage of genome-wide association studies is that the genotyping arrays can be used to examine Copy Number Variants (CNVs) at the same time. CNVs are DNA segments that differ in number when comparing genomes from two different individuals.

What made genome-wide association studies possible were the large numbers of polymorphisms in the human genome that have been discovered in recent years, the understanding of LD patterns across the genome through the HapMap project and advances in technology that considerably decreased the cost of genotyping. Although genome-wide association studies are designed to cover the entire genome, genotyping arrays used for genome-wide association studies have not usually included Y chromosome SNPs (see following chapters for more information). Genome-wide association studies have been published for bipolar disorder (Wellcome Trust Case Control Consortium 2007), schizophrenia (O'Donovan et al. 2008; Shifman et al. 2008; Sullivan et al. 2008), autism (Weiss 2009) and ADHD (Neale et al. 2008).

2.4.4.1 Allelic architecture of common diseases

Although for many common disorders, including psychiatric disorders, the allelic frequency and the effect sizes of susceptibility variants are unknown, two polarised hypotheses exist. The common disease/common variant hypothesis supports the view that common diseases are caused by several susceptibility variants of high frequency and small to moderate effect sizes. These variants are also present in unaffected individuals (Reich and Lander 2001). The alternative hypothesis suggests that common diseases are highly heterogeneous in their causes with distinct susceptibility variants in different individuals. These variants are rare and they are usually of large effect sizes (Liu et al. 2005). However, the most likely scenario is that for most

common diseases both hypotheses are valid and common as well as rare variants need to be investigated to increase understanding of the pathophysiology.

2.4.4.2 Power of association studies

The power of an association study to detect a susceptibility variant depends mainly on the effect size and the frequency of the risk allele in the population. Power increases as effect sizes and frequency increase. This means that genome-wide association studies are more powerful when picking common alleles and /or alleles of large effect sizes (Wang et al. 2005). Since the effect sizes and frequencies of alleles cannot be manipulated, researches are left with the option to increase sample sizes in order to have adequate power. Figure 2-1 shows that genome-wide association studies have little power to detect rare variants when their effect sizes are small. For example to have 80% power to detect a variant with a MAF of 0.1 and an OR of 1.3 the sample size would have to increase to more than 10,000 cases and 10,000 controls which is probably not feasible for some disorders. Even having 80% power to detect a variant with MAF of 0.1 and a modest effect size of 1.5, would require a sample of just fewer than 1,000 cases and 1,000 controls.

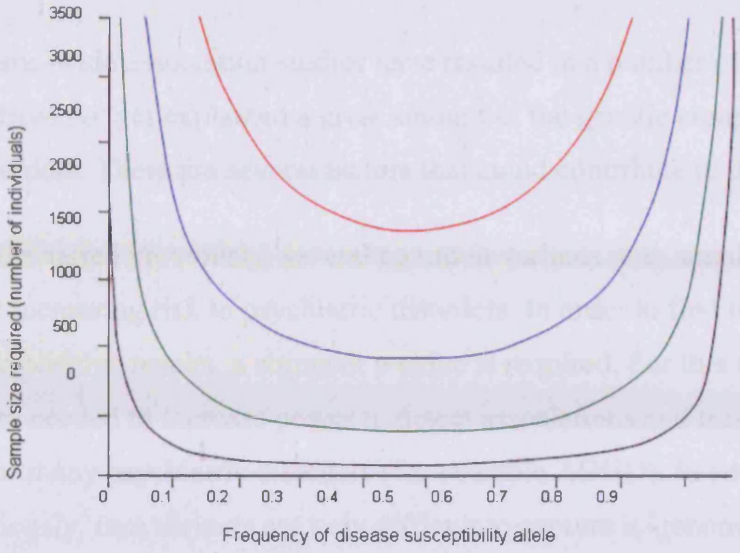

Figure 2-1. Effects of OR and allele frequency on sample size requirements. The total number of individuals (equal numbers of cases and controls) that are required for an association study to have 80% power to detect disease variants with allelic odds ratios of 1.2 (red), 1.3 (blue), 1.5 (green) and 2 (black) are shown. Significance level is set at $p<10^{-6}$. A multiplicative model and perfect LD between test markers and disease variants are assumed. Adapted from Wang et al. (2005)

In addition, genome-wide association studies are usually performed by genotyping hundred of thousands or even millions of SNPs at the same time. This means that an enormous amount of tests are performed in a single experiment and it requires a p value $<5x10^{-8}$ in order to exclude false positives results (Pe'er et al. 2008). In order to achieve this level of significance, even larger samples might be required than are presently available. For most research centres these sample sizes cannot be achieved without forming collaborations. The Wellcome Trust Case Control Consortium (WTCCC) is an example of this type of collaboration examining ~2,000 individuals for each of seven major complex disorders (bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 and type 2 diabetes) and ~3,000 shared controls (Wellcome Trust Case Control Consortium 2007). In the case of type 2 diabetes, a pooled sample of ~60,000 subjects was required to definitively implicate a large set of genes (McCarthy and Zeggini 2009).

2.4.4.3 Limitations of genome-wide association studies

Although genome-wide association studies have resulted in a number of exciting findings, they have not yet explained a great amount of the genetic component of psychiatric disorders. There are several factors that could contribute to that.

First of all, as discussed previously, several common variants with small effect sizes are likely to be increasing risk to psychiatric disorders. In order to find them without increasing false positive results, a stringent p value is required. For this reason, large sample sizes are needed to increase power to detect associations and this has not been achieved yet for many psychiatric disorders (for example ADHD). In addition, as discussed previously, rare variants are very difficult to capture in genome-wide association studies.

In most cases, genome-wide association studies assume an additive model for risk, so that the presence of two risk factors in an individual increases risk by the sum of these two factors. However, this might not be the case in some instances. Epistatic interactions, like gene-gene and gene-environment interactions, are difficult to find in genome-wide association study designs. Other factors, such as epigenetic inheritance and CNVs that are not detected by current methods, could also contribute to the "missing heritability" (Hardy and Singleton 2009).

Finally, one of the main issues faced by genome-wide association studies of psychiatric disorders is that, unlike other complex disorders, they are highly clinically heterogeneous. Diagnosis of psychiatric disorders depends on a set of criteria, which although reliable, might not always lead to a cohort of patients with the same underlying causes (Craddock et al. 2008). For these reasons, ascertainment of patients in genome-wide association studies of psychiatric disorders deserves special mention, as it is essential that comorbidities and other factors contributing to heterogeneity are recognised.

As the cost of conducting genome-wide association studies decreases, studies involving larger sample sizes are being reported. Interesting findings pointing to pathways that have not been considered before have already appeared for psychiatric disorders. However, it is important to remember that replication in independent

samples is required before a variant is considered a true susceptibility variant and is further investigated.

2.5 Molecular genetics of ADHD

2.5.1 Genome-wide linkage studies

Genome-wide linkage studies are based on families with multiple affected individuals, mainly affected sibling pairs, and aim to identify broad regions in the genome that could harbour susceptibility genes.

Genome-wide linkage scans have been based either on the affected sibling pair (ASP) or the extended pedigree approach. At the time of writing, results of seven linkage studies of ADHD have been published (Table 2-2); five using ASPs and two using multiplex families. As Table 2-2 suggests, there is some overlap in significant linkage peaks. Evidence for at least nominally significant linkage has been found more than once for chromosomal regions 5p13, 16p13 and 17p11.

Table 2-2. Summary of ADHD genome-wide linkage studies published so far (regions that show overlap are in bold)

Study	Size and design	Sample origin	Chromosomal regions	Evidence of linkage (LOD)
Fisher et al. (2002)	126 ASPs	USA	No evidence of linkage	-
Smalley et al. (2002)	203 ASPs	USA	**16p13**	4.2
Ogdie et al. (2003)	270 ASPs	USA	**16p13**	4
Ogdie et al. (2004)	308 ASPS	USA	5p13 6q12 **16p13** **17p11**	2.55 3.3 3.73 3.63
Ogdie et al. (2006)	424 ASPs	USA and Dutch	**5p13**	3.67
Bakker et al. (2003)	164 ASPs	Dutch	15q15 7p13 **5p13**	3.54 3.04 1.43
Arcos-Burgos et al. (2004)	16 extended pedigrees	Colombian	**17p11**	MLS=1.4 2
Hebebrand et al. (2006)	155 ASPs	German	**5p** (at 17cM) **17p**	2.59 3.37 (for inattention only) Nominally significant
Faraone et al. (2007)	601 ASPs	USA	No evidence of linkage	-
Romanos et al. (2008)	8 extended pedigrees	German	5q13 14q12 2q35	4.16 4.5 3.4
Asherson et al. (2008)	142 ASPs	IMAGE sample	16q23 9q22	3.1 2.13
Zhou et al. (2008)	Meta-analysis of 7 linkage scans	-	16 (between 64 and 83 Mb)	p^{SR}=0.000 34, p^{OR}=0.04

ASPs: Affected Sibling Pairs; LOD: Logarithm of the odds; MLS: Maximum Multipoint LOD; IMAGE: International Multi-centre ADHD Gene; SR: Summed Rank; OR: Odds Ratio

A study of 126 ASPs from the USA provided weak evidence for linkage at 16p13 (Fisher et al. 2002), which has also been implicated in autism. Stronger evidence for 16p13 [LOD (Logarithm Of the Odds) score 4.2] came from an expanded sample of 203 ASPs (Smalley et al. 2002). The same region also came up in a study of 308 American ASPs with an MLS (Maximum Multipoint LOD) of 3.73 (Ogdie et al. 2004).

The chromosomal region at 5p13 also came up in the latter study (Ogdie et al. 2004). Modest evidence for linkage for the same region (MLS=1.43) appeared in a study of 164 Dutch ASPs (Bakker et al. 2003). Pooled analysis of the American and Dutch samples (Ogdie et al. 2006) yielded significant evidence of linkage (MLS=3.67) for 5p13, although the signal appeared to be coming mainly from the American families suggesting sample heterogeneity had an impact on the results. Finally, Hebebrand et al. (2006) studied 155 ASPs from Germany and found weak evidence of linkage (MLS=2.59) on chromosome 5p at 17cM.

Another interesting region is 17p11, as there is evidence of linkage in both the study of 308 ASPs from the USA (Ogdie et al. 2004) with an MLS of 3.63, as well as from the study of German ASPs (Hebebrand et al. 2006). A study with a different design using 16 multigenerational and extended pedigrees from Colombia also reported weak evidence of linkage at 17p11 (Arcos-Burgos et al. 2004).

A meta-analysis of seven independent linkage scans yielded a significant association for chromosome 16 between 64 and 83 Mb [p^{SR}=0.00034, p^{OR}=0.04) (Zhou et al. 2008). Interestingly, the same region came to light from a genome-wide association study of quantitative traits of ADHD (Lasky-Su et al. 2008). This region harbours the *CDH13* (Cadherin 13) gene which has been implicated in substance use disorders (Uhl et al. 2008).

In summary, genome-wide linkage studies have yielded some interesting results for chromosomal regions that need to be further investigated. It seems difficult to achieve replication of genome-wide significant results, suggesting that there are no susceptibility genes of large effect for ADHD. For this reason, an association approach is likely to be more suitable, since it can identify genes of smaller effect.

2.5.2 Candidate gene association studies

Candidate gene association studies have not been very successful in complex disorders, possibly because the assumptions about the pathogenesis were not correct. One notable exception is ADHD, where selection of candidate genes has been informed by animal, imaging, and pharmacological studies.

Due to the large number of such studies, I will only focus on genes where significant findings have stood up to meta-analyses or pooled analyses or where there have been replications (Table 2-3 summarises all the meta-analyses for the genes discussed). Many of these functional candidate genes have been selected because they code for proteins or enzymes involved in the dopamine pathway. Interest in the dopaminergic system in ADHD has come from pharmacological, animal and imaging studies. First, stimulant medication, such as methylphenidate, reduces ADHD symptoms and inhibits the reuptake of dopamine thus increasing its extracellular concentration (DiMaio et al. 2003). In addition, imaging studies of patients with ADHD provide evidence of changes in brain regions where dopaminergic systems are more active (Spencer et al. 2005). Finally, animal studies, such as work on the *DAT* knockout mouse which manifests ADHD-like behaviour, have also implicated dopaminergic systems (Gainetdinov 2007). All this evidence has led to the investigation of *DRD4, DRD5, DAT1* and *COMT* which will be discussed. Other genes such as those encoding dopamine beta hydroxylase (*DBH*), monoamine oxidase A (*MAOA*), the dopamine D2 (*DRD2*) and dopamine D3 (*DRD3*) receptors have also been investigated but the results are not conclusive. Nearly all the published molecular genetic studies of ADHD have been based on clinical cases where ADHD has been defined according to DSM-IV diagnostic criteria. A few studies have examined ADHD defined dimensionally but the association findings here have generally been weaker (Mill et al. 2005).

Table 2-3. Meta-analysis and pooled studies for candidate gene association studies of ADHD

Gene	Reference	Type of studies included	OR	95% CI	p value
***DRD4* 7-repeat allele**	Faraone et al. (2001)	Case-control	1.9	1.4-2.2	<0.001
		Family-based	1.4	1.1-1.6	0.02
	Faraone et al. (2005)	Case-control	1.45	1.27-1.65	-
		Family-based	1.16	1.03-1.31	
	Li et al. (2006)	Case-control and family-based	1.34	1.23-1.45	2×10^{-12}
	Gizer et al. (2009)	Case-control and family-based	1.27	1.16-1.39	<0.00001
***DRD4* 5-repeat allele**	Li et al. (2006)	Case-control and family-based	1.68	1.17-2.41	0.005
***DRD5* 148-bp microsatellite repeat**	Maher et al. (2002)	Family-based	1.57	1.25-1.96	0.00008
	Lowe et al. (2004)	Family-based	1.24	1.1-1.4	0.00005
	Li et al. (2006)	Case-control and family-based	1.34	1.21-1.50	8×10^{-8}
	Gizer et al. (2009)	Case-control and family-based	1.22	1.1-1.36	0.000095
***DAT1* 480-bp VNTR**	Maher et al. (2002)	Family-based	1.27	0.99-1.62	0.063
	Faraone et al. (2005)	Family-based	1.13	1.03-1.24	-
	Purper-Ouakil et al. (2005)	-	1.19	0.99-1.41	0.21
	Li et al. (2006)	Case-control and family-based	1.04	0.98-1.11	0.20
	Yang et al. (2007)	Family-based	1.7	1.05-1.30	0.004
		Case-control	0.95	0.8-1.12	0.54
		Haplotype-based	1.5	0.97-2.33	0.07
	Gizer et al. (2009)	Case-control and family-based	1.1	1.03-1.17	0.002
***COMT* $Val^{158}Met$**	Faraone et al. (2005)	Case-control and family-based	1.0	-	-
	Cheuk and Wong	Case-control and family-based	0.99	0.88-1.12	0.87
	Gizer et al. (2009)	Case-control and family-based	0.99	0.92-1.06	0.62
***SNAP25* T1065G**	Faraone et al. (2005)	Case-control and family-based	1.19	1.03-1.38	-
	Gizer et al. (2009)	Case-control and family-based	1.15	1.01-1.31	0.03

OR: Odds ratio; CI: Confidence Interval; DRD4: Dopamine D4 receptor gene; DRD5 Dopamine D5 receptor gene; DAT1: Dopamine transporter gene; COMT: Catechol-O-Methyltransferase; SNAP25: Synaptosomal-associated protein of 25kD

2.5.2.1 Dopamine D4 receptor gene (DRD4)

The *DRD4* gene is on chromosome 11p15.5. The D4 receptor binds both dopamine and noradrenaline. Most studies have focused on a variable number tandem repeat (VNTR) polymorphism in exon III of the gene. The number of repeats ranges from 2-11 with different populations having different alleles. This polymorphism is supposed to be functional, since the 7-repeat allele reduces the ability of the receptor to bind dopamine according to *in vitro* studies.

The first meta-analysis of the *DRD4* gene in ADHD found significant association between the 7-repeat allele and ADHD in both case-control studies [OR (odds ratio) =1.9, 95% CI (confidence interval) 1.4-2.2] and family-based studies (OR=1.4, 95% CI 1.1-1.6) (Faraone et al. 2001). The same group conducted a pooled analysis and found that the association with ADHD was still significant in both case-control studies (OR=1.45, 95% CI 1.27-1.65) and family-based studies (OR=1.16, 95% CI 1.03-1.31) (Faraone et al. 2005). Li et al. (2006) included 33 association studies in their meta-analysis and also obtained strong evidence that the 7-repeat allele (P=2 x 10^{-12}, OR=1.34, 95% CI 1.23-1.45) is associated with ADHD. They suggested that the 5-repeat allele (P=0.005, OR=1.68, 95% CI 1.17-2.41) also confers increased risk for ADHD and concluded that the 4-repeat allele has a protective role (P=0.004, OR=0.90, 95% CI 0.84-0.97) (Li et al. 2006) but that evidence is not clear-cut. More recently, the previous meta-analytic results on the 7-repeat allele were replicated in the meta-analysis by Gizer et al. (2009). They found a significant, moderate association between ADHD and the 7-repeat allele (P<0.00001, OR-1.27, 95% CI 1.16-1.39), although they also reported evidence of substantial heterogeneity of effect sizes across the studies (Gizer et al. 2009).

The *DRD4* 7-repeat allele has also been reported to influence cognitive performance but findings here are mixed with some studies showing that those with the 7-repeat allele perform worse on measures of accuracy and cognitive ability and other groups finding better performance in those with the risk allele, so no conclusions can yet be drawn (Thapar et al. 2007d). There have also been longitudinal studies suggesting that those with ADHD who possess the *DRD4* 7 repeat allele show a poorer outcome (El-

Faddagh et al. 2004; Langley et al. 2008; Mill et al. 2005). However again, findings have not been entirely consistent (Shaw et al. 2007).

Another polymorphism on *DRD4* that has been investigated is a 120 bp insertion/deletion in the promoter region, however there was no association of this polymorphism with ADHD in the meta-analysis by Gizer et al. (2009).

2.5.2.2 Dopamine D5 receptor gene (DRD5)

Another dopamine receptor gene, *DRD5*, on chromosome 4p15.1-15.3 also appears to be important. The associated polymorphism is a microsatellite (a dinucleotide repeat with variable number of copies) mapping 18.5 kb away from the 5′ end of the gene (Daly et al. 1999). The first meta-analysis of *DRD5* included 5 studies and showed a significant association with the 148-bp allele (Maher et al. 2002). Significant association (OR=1.24, 95% CI 1.1-1.4) with the same allele was also reported by a joint analysis of 14 independent studies resulting in approximately 2000 ADHD cases (Lowe et al. 2004). Association was stronger in those with ADHD inattentive type. Li et al. (2006) included nine association studies of *DRD5* and also concluded that the 148-bp allele of *DRD5* confers risk for ADHD, while a significant association (p=0.000095, OR=1.22, 95% CI 1.10-1.36) with moderate heterogeneity in effect sizes was found in a subsequent meta-analysis (Gizer et al. 2009). Li et al. (2006) also suggested that the 136-bp allele of *DRD5* has a protective role. However, a study of 329 male twins examining a trait measure of ADHD found association with the same variant but in the opposite direction, i.e. the 148-bp allele had a protective role (Mill et al. 2005). Association with ADHD has also been shown for other polymorphisms in *DRD5* including two microsatellites at the 5′ end of the gene and a SNP in the 3′ UTR (Untranslated Region) of *DRD5* but these have been less widely studied.

2.5.2.3 Dopamine transporter gene (SLC6A3 or DAT1)

The dopamine transporter gene on chromosome 5p13.3 was initially considered the most likely candidate gene for ADHD. The major reason is that it is responsible for the reuptake of dopamine in the presynaptic cleft and is the target of stimulant medication (DiMaio et al. 2003). Another reason for considering *DAT1* an important

candidate gene for ADHD is because *DAT* knockout mice exhibit hyperactivity and deficits in inhibitory behaviour. Treating these mice with stimulants reduces symptoms (Gainetdinov 2007). Despite these findings, association studies have not been able to produce clear evidence about the involvement of *DAT1* in ADHD.

The best studied polymorphism is a VNTR in the 3′ UTR of the gene. The first meta-analysis of nine *DAT1* studies in ADHD reported an OR of 1.27 with a trend for association with the 480-bp allele (p=0.063) (Maher et al. 2002). However, there was evidence of heterogeneity across samples. In a meta-analysis by Faraone et al. (2005) the association of the 480-bp allele was significant, although the OR=1.13 (95% CI 1.03-1.24) was small. Another meta-analysis of twelve family-based studies the same year failed to find any association of the *DAT1* with ADHD (Purper-Ouakil et al. 2005). No association of the 480-bp allele of *DAT1* with ADHD was also found in the meta-analysis by Li et al. (2006). Another meta-analysis showed a small but significant association (p=0.004, OR=1.17, 95% CI 1.05-1.30) for family TDT (Transmission Disequilibrium Test) studies but not for haplotype-based studies or case-control studies (Yang et al. 2007). However, the number of haplotype-based and case-control studies was small (Yang et al. 2007). Finally, a significant association between the 480-bp allele and ADHD (P=0.002, OR=1.1, 95% CI 1.03-1.17) was found in an updated meta-analysis by Gizer et al. (2009). The same meta-analysis found significant associations with other polymorphisms in the *DAT1* gene. It is also interesting that they report evidence of substantial heterogeneity.

Apart from the 3′ UTR VNTR, association has been also reported for a SNP (rs40184) in the IMAGE sample (Brookes et al. 2006a). Haplotypes that include the 3′ UTR VNTR and other microsatellite repeats have also been found to be associated with increased risk for ADHD (Asherson et al. 2007; Brookes et al. 2006b).

The evidence overall from an abundance of meta-analytic studies is inconsistent. It is worth noting that most of the meta-analyses have found evidence of sample heterogeneity. One possibility is that the effect size of this variant is very small. The polymorphism in question may not be directly responsible for increasing risk for ADHD but it could be in linkage disequilibrium with another functional polymorphism. After all, it is in the 3′ UTR of the gene, which as the name suggests, is transcribed to mRNA but not translated into protein; thus its function is

questionable. Another source of heterogeneity could be the presence of multiple polymorphisms within the DAT1 gene increasing risk to ADHD. Heterogeneity can also be the result of gene-environment interaction (Thapar et al. 2007c). Evidence of interaction between the haplotype containing the 3′ UTR VNTR and exposure to maternal alcohol consumption during pregnancy has been found (Brookes et al. 2006b) but this requires replication. Two studies suggested that those with the 480-bp (10 repeat) allele, who were exposed to maternal smoking in pregnancy, showed higher levels of ADHD symptoms (Becker et al. 2008; Kahn et al. 2003). However, another study found evidence of gene-environment interaction for the 9 repeat allele in relation to exposure to maternal smoking in pregnancy (Neuman et al. 2007) and a different group failed to replicate these findings (Langley et al. 2008).

2.5.2.4 Catechol-O-Methyltransferase (COMT)

The gene encoding COMT, an enzyme catalysing the degradation of dopamine, adrenaline and noradrenaline, is on chromosome 22q11.2. Interest in *COMT* comes from its involvement in dopaminergic pathways. The most studied polymorphism in *COMT* is a G/A transition found in codon 158, which results in a valine-to-methionine transition. This polymorphism is functional with the val/val genotype increasing enzyme activity. No individual study or two pooled analyses (Cheuk and Wong 2006; Faraone et al. 2005) have found evidence of association between this variant and ADHD. A recent meta-analysis also reports no evidence of association (Gizer et al. 2009).

Interestingly, the *COMT* Val allele yielded an almost significant result in the male group of the Cheuk and Wong (2006) meta-analysis. Thus, it could potentially be involved in male susceptibility to ADHD (Cheuk and Wong 2006). In addition, there is evidence that the *COMT* val/val genotype is associated with conduct disorder symptoms in patients with ADHD (Thapar et al. 2005). Since this first report, the same genotype was subsequently found to be associated with antisocial behaviour in those with ADHD (but not in those without ADHD) in two independent populations from a UK and a New Zealand birth cohort (Caspi et al. 2008). A pooled analysis of four studies also showed significant association with the val/val genotype (Caspi et al.

2008). These results suggest that some gene variants operate by modifying the ADHD phenotype rather than by increasing risk.

2.5.2.5 Synaptosomal-associated protein of 25kD (SNAP25)

SNAP25 is a neuron specific protein involved in the regulation of neurotransmitter release. It emerged as a candidate gene for ADHD when it was discovered that the coloboma mouse mutant, which lacks a chromosomal region containing *SNAP25* as well as other genes, exhibits hyperactivity that could be reduced by the use of stimulant medication D-amphetamine. One meta-analysis including this gene reported a significant association with the T1065G SNP at the 3′ end of the gene (Faraone et al. 2005). A polymorphism in the 3′ UTR (rs3746544) of the gene was also significantly associated with ADHD in a meta-analysis across seven studies (P=0.03, OR=1.15, 95% CI 1.01-1.31) (Gizer et al. 2009). Nominally significant association with *SNAP25* was also reported using the IMAGE sample, although the group tested different markers (Brookes et al. 2006a). A study of 12 SNPs in *SNAP25* using two independent samples also yielded evidence of association but in only one of the samples (Feng et al. 2005). Finally, Kim et al. (2007) reported a modestly significant association with two SNPs that have not been studied before in a TDT analysis. Interestingly, in this study, a stronger association with *SNAP25* was found in patients with ADHD and comorbid depression possibly highlighting the importance of taking into account psychiatric comorbidity in association studies (Kim et al. 2007).

To summarise, the most robust evidence is for an association between the *DRD4* VNTR, the *DRD5* microsatellite marker in ADHD and the *DAT1* 480-bp VNTR. It is still unclear whether the *DRD4* VNTR is causal. The *DRD5* 148-bp microsatellite is 18.5kb away from the gene but that does not mean it does not influence *DRD5* in a yet unknown fashion. As for *DAT1*, there is significant evidence of heterogeneity across studies, although it initially appeared as the strongest candidate gene. One explanation could be that the *DAT1* 480-bp VNTR is not responsible for the association but is tagging the functional polymorphism which is yet to be found. The evidence that *COMT* val/val genotype has a modifying effect on antisocial behaviour in ADHD is strong. Further studies on *SNAP25* are needed before conclusions can be drawn.

2.5.3 Genome-wide association studies (GWAS)

Genome-wide association studies of ADHD and related phenotypes have been performed, although they are still in early stages compared to other complex disorders. The first study used a sample of 959 family trios collected as part of the IMAGE study (see Franke et al. 2009 for a review on genome-wide association studies on ADHD). This study included 600,000 genotyped SNPs (Neale et al. 2008). Although, none of the SNPs reached genome-wide significance (the highest p values were 7.45E-06), these results do not exclude the effect of genes that increase risk for ADHD. In the list with the top-25 SNPs, there are some interesting candidate genes for ADHD, including CNR1 (cannabinoid receptor 1) (Neale et al. 2008). This gene falls into a suggestive linkage region in a recent meta-analysis of linkage studies (Zhou et al. 2008) and it has been also associated with alcohol and drug abuse (Zuo et al. 2007). The most likely reason that there are no genome-wide significant results is that risk variants for ADHD have very small effect sizes and genome-wide scans will require 10,000-20,000 cases to have a good chance of finding them. In addition, rare variants with large effect sizes would be difficult to detect unless sample sizes increase. Sample homogeneity and the ascertainment of the disorder can also be crucial factors (Neale et al. 2008).

Results from the same genome-wide scan have been used to test for association with quantitative phenotypes generated from the ADHD symptoms (Lasky-Su et al. 2008). The authors tested three quantitative phenotypes (number of hyperactive-impulsive symptoms, number of inattentive symptoms and total number of symptoms) under three different genetic models, which increased the multiple testing burden. There were two genome-wide significant results; one of them is an intronic SNP in *CDH13* (Cadherin 13) associated with the total number of ADHD symptoms (Lasky-Su et al. 2008) and the other one is an intronic SNP in *GFOD1* (glucose-fructose oxidoreductase-domain containing 1) associated with the number of inattentive symptoms. *CDH13* has been previously involved in substance abuse (Uhl et al. 2008) and is in the list with the top-25 SNPs from the previous genome-wide association study (Neale et al. 2008), although the two studies are not independent.

Based on a list of candidate genes for ADHD compiled *a priori*, SNPs within *SLC9A9* (solute carrier family 9), as well as other genes previously associated with ADHD, were associated with the disorder, although they failed to reach genome-wide significance (Lasky-Su et al. 2008). Apart from being the top hit, *SLC9A9* also contains the largest number of associations in terms of SNPs and phenotypes. However, it is also the largest gene in the study.

A pooled genome-wide association study not associated with IMAGE analysed 343 adult patients with persistent ADHD and 304 controls from Germany (Lesch et al. 2008). The authors used additional criteria to report their findings, including, to name just a few, SNPs localised within genes, presence in suggestive linkage regions, expression in the brain. Although there were no genome-wide significant findings, a large proportion of the top-ranked SNPs are located in genes expressed in the brain and, some of them, have also shown association with substance abuse disorders, which are found very frequently in adult ADHD patients (Lesch et al. 2008).

In terms of overlap of genome-wide association studies in ADHD, *CDH13* on chromosome 16q24.2-24.3 stands out, since it is reported in two previous genome-wide association studies of ADHD (Lasky-Su et al. 2008; Neale et al. 2008). This gene is further supported by the fact that it falls in the only significant linkage region from the meta-analysis by Zhou et al. (2008). Although *CDH13* is not a classic candidate gene for ADHD, it codes for a cell-cell adhesion protein and is also a regulator of neural cell growth. Other interesting findings in the top-ranks of these genome-wide association studies are *TLL* (Tolloid-like) genes (Lasky-Su et al. 2008; Lesch et al. 2008; Neale et al. 2008). They code for metalloproteases that cleave collagen and they are expressed in the brain. These findings highlight the fact that genome-wide association studies can point to pathways of the disorder that would not be explored otherwise.

The IMAGE dataset was used to perform a genome-wide association study on ADHD with comorbid CD (Anney et al. 2008), which is hypothesised to have a different aetiology to ADHD without CD (Thapar et al. 2005). There were no reported genome-wide significant associations, although fifty-four markers were nominally associated. Some of them are located near genes that could be interesting for ADHD.

However, the authors point out that their analysis was exploratory and any associations would require replication in an independent sample (Anney et al. 2008).

An interesting approach to genome-wide association studies has been taken by researchers investigating gene-environment interactions in such a setting. The association of maternal warmth and criticism with ADHD severity and comorbid CD were tested in the IMAGE dataset (CD was evaluated both as a categorical and a quantitative trait) (Sonuga-Barke et al. 2008). Although no genome-wide significant results were produced, nominal effects were found both with and without main effects. Some of the observed interactions implicated the *SLC1A1* (neuronal and epithelial glutamate transporter) and *NRG3* (Neuregulin-3) genes, which have been implicated in psychiatric disorders previously (Sonuga-Barke et al. 2008).

The most recent genome-wide association study on ADHD has used samples from IMAGE II, which is a collaborative sample set and initial analysis suggests that there are no genome-wide significant associations (Neale 2009 personal communication). Unpublished data on a meta-analysis of genome-wide association studies of ADHD have again failed to produce genome-wide significant results (Neale 2009 personal communication). This again supports the notion that common gene variants of very small effects operate on ADHD and very large samples will be needed to have a chance of finding them. Rare variants of large effect size could also play a role but genome-wide association studies would require extremely large sample sizes to find them.

2.5.4 Copy number variants (CNVs)

Copy Number Variants (CNVs) are DNA segments, of at least 1 kb in size, that vary in number when genomes of different individuals are compared. They can be copy number gains (called insertions or duplications), when there is a relative gain of a DNA segment compared to the control genome, or they can be copy number losses (called deletions), when there is a relative loss of a DNA segment. The presence of a CNV in an individual does not necessarily mean that there would be an effect on the phenotype, since CNVs are part of the normal variation of the human genome (Scherer et al. 2007). However, it has emerged that CNVs can increase risk to

disorders, especially those of a neurodevelopmental nature such as autism (Sebat et al. 2007) and mental retardation; thus the advances in technology that have made CNV detection more cost and time effective are very welcome.

There is only one study of CNVs in ADHD and it did not find increased numbers of deletions or duplications in ADHD compared to controls (Elia et al. 2009). However, the authors report that the gene set associated with inherited rare CNVs in cases was significantly enriched for genes studied in relation to autism, schizophrenia and Tourette's syndrome (Elia et al. 2009). Another unpublished study on CNVs in ADHD coming from the Department of Psychological Medicine and Neurology in Cardiff University reports an increase of large rare CNVs in cases compared to controls (Williams 2009 Personal Communication). This excess of large rare CNVs is more profound in cases with ADHD and intellectual disability. In addition, the authors report an excess of chromosome 16p13.11 duplications in cases and enrichment for CNVs previously reported in both autism and schizophrenia (Williams 2009 Personal Communication).

Although CNV studies in ADHD are still in their infancy, they can point to new pathways of risk factors for ADHD and they highlight the overlap in risk factors operating in different neurodevelopmental disorders. They also suggest that rare variants can play an important role in the pathophysiology of ADHD.

2.6 The contribution of gene-environment interaction in ADHD

It is well-recognised that psychiatric disorders, as other complex disorders, are influenced by both genetic and environmental factors (Rutter 2007). Gene-environment interplay encompasses both gene-environment correlation and gene-environment interaction. Gene-environment correlation applies when exposure to a certain environmental factor depends on the genetic make-up of the individual. This means that genes increase risk of developing ADHD by influencing exposure to environmental factors. For example, maternal smoking and stress in pregnancy have both been linked to ADHD. However there is evidence to suggest that these

environmental risks are influenced by maternally-provided inherited factors and index liability to ADHD (Thapar and Rutter 2009).

Gene-environment interaction takes place when a genetic factor exerts its effect only when the individual is exposed to a certain environmental factor (Thapar et al. 2007b). Genetic studies, such as case-control association studies with information about exposure to environmental factors and epidemiological studies can both assess the effect of gene-environment interaction. Twin or adoption studies on this topic have not been published for ADHD, although this could indicate a publication bias against negative results (Thapar et al. 2007b).

Several environmental factors have been discussed in relation to ADHD, including maternal smoking during pregnancy, low birth weight, alcohol consumption during pregnancy, maternal stress during pregnancy, poor maternal diet and toxins during the prenatal or neonatal period (Linnet et al. 2003). It is notable that the common characteristic of these environmental factors is that they require exposure during a critical period. For this reason it has been suggested that they might result in epigenetic processes being triggered during a developmental window (Mill and Petronis 2008).

Investigating gene-environment interaction in ADHD is still at an early stage and there is lack of robust replication studies. For this reason the discussion will only focus on the most promising interactions.

Maternal smoking has been considered the environmental factor that is most robustly-associated with ADHD. A meta-analysis has reported a pooled odds ratio of 2.38 (95% CI 1.61-3.52) (Langley et al. 2005). Animal studies have also suggested that prenatal exposure to nicotine is linked to low birth weight, increased locomotor activity and cognitive impairment (Ernst et al. 2001). A combination of maternal smoking during pregnancy and the *DAT*1 148-bp allele has been associated with increased ADHD symptoms (Kahn et al. 2003). An interaction between maternal smoking and the *DAT*1 allele has been found by another study as well, although the association was with a different allele (Neuman et al. 2007). The same study also reported an interaction between smoking and the *DRD4* 7-repeat allele to increase

ADHD symptoms (Neuman et al. 2007). In contrast, there were two reports of no association between smoking and *DAT1* (Brookes et al. 2006b; Langley et al. 2008).

Low birth weight is considered an important environmental risk factor for ADHD with an odds ratio of 2.64 from a meta-analysis (Bhutta et al. 2002). Low birth weight and possession of the val/val *COMT* genotype have been shown to increase susceptibility to ADHD (Thapar et al. 2005). There was also a modifying effect on the disorder by increasing conduct disorder symptoms (Thapar et al. 2005).

Prenatal exposure to alcohol was investigated in ADHD and it was found that it increased ADHD symptoms when a specific haplotype of SNP alleles in *DAT1* was present (Brookes et al. 2006b).

Two markers in the same gene, *DAT1*, have been reported to interact with psychosocial adversity (defined by measures of family and parental adversity) to increase ADHD symptoms (Laucht et al. 2007).

Finally, the interaction between the *DRD4* 7-repeat allele and season of birth in relation to ADHD have been investigated by two groups (Brookes et al. 2006b; Seeger et al. 2004) but the results were in opposite directions.

To sum up, so far evidence in support of gene-environment for ADHD is not yet convincing. Most of the environmental factors that are considered to increase susceptibility to ADHD exert their effect during the prenatal period. For this reason, it is difficult to disentangle whether the factor responsible is the prenatal environment or the effect is mediated by shared genetic factors between the mother and the offspring.

A study with a novel design has tried to solve this problem by using data collected by families with children born through *in vitro* fertilisation (IVF). By comparing children born by mothers genetically related to them (homologous IVF) to children born by mothers not related to their offspring (oocyte or embryo donation and gestational surrogacy), it is possible to conclude on the nature of the effect an environmental factor has. When the effect of an environmental factor is present both in the related and unrelated group, it means that it is a true environmental effect but when it is only present in the related group, it is a result of shared genetic factors (Thapar et al. 2007a). This new experimental design has been able to elucidate the nature of putative

environmental effects thought to be involved in psychopathology. It was found that some of them were not due to real environmental effects but they resulted from shared genetic factors between the mother and the offspring. One such factor was maternal smoking during pregnancy and ADHD, which was due to genetic factors conferring susceptibility to the mother to smoking during pregnancy and susceptibility to the offspring to developing ADHD (Thapar et al. 2009). The same was also true for the effect of maternal smoking during pregnancy and antisocial behaviour (Rice et al. 2009a; Rice et al. 2009b). These studies highlight the complexity of studying the involvement of environmental factors in psychiatric disorders and the caution that should be exercised when interpreting such associations.

To conclude, ADHD is a childhood-onset neurodevelopmental disorder with high heritability. However, the genetic component of this disorder requires more research to be elucidated. Case-control association studies point to the involvement of dopaminergic genes while genome-wide association studies of larger sample sizes and careful ascertainment are required. Apart from common variants with small effect sizes, rare variants are likely to be involved in the pathophysiology of the disorder.

Chapter 3 Schizophrenia

3.1 Schizophrenia symptoms and diagnosis

Schizophrenia is a severe psychiatric disorder accounting for a significant proportion of psychiatric morbidity and requiring long-term medical help and social support. Patients with schizophrenia are often unable to work and sustain interpersonal relationships resulting in enormous economic and psychological burden for the patient, the family and the society.

3.1.1 Epidemiology

The lifetime risk for developing schizophrenia is ~1% and it is very stable across different countries and cultures. There is an increased rate of men affected by schizophrenia compared to women; the male/female ratio is ~1.4, although sex differences in the age of onset and the severity of the disorder are profound. Increased rates of schizophrenia have been found for migrants and those living in an urban versus rural environment (McGrath et al. 2004). Environmental factors that act early in life have also been studied in schizophrenia. These include obstetric complications, prenatal maternal psychological stress and maternal influenza exposure during pregnancy. There is also a link with season of birth, which could be mediated by the increased risk of influenza exposure during pregnancy in winter (Kinney et al. 1994). One of the strongest associated environmental factors is maternal nutritional deficiency. Data from two natural experiments, where women were exposed to severe malnutrition due to famine during pregnancy, showed that the relative risk for schizophrenia and schizoid personality disorder in their offspring was double compared to matched controls (St Clair et al. 2005; Susser et al. 1996).

3.1.2 Clinical presentation

The symptoms characterising schizophrenia can be categorised into two groups. Positive symptoms include hallucinations, delusions, thought disorder, as well as bizarre and disorganised behaviour. Negative symptoms, such as loss of motivation

and drive, flattened mood, social isolation, poverty of speech and poor personal hygiene are also present. Cognitive impairment (usually mild) is common with the most affected domains being attention, working memory and executive functions. IQ is generally a standard deviation lower than expected. Cognitive deficits are often present at the onset or even before the onset of the disorder and they tend to increase, even when there are no other symptoms. It has been suggested that they are caused by abnormal frontotemporal interactions (Pearlson 2000). Depression is also common, usually following an acute episode.

3.1.3 Schizophrenia diagnosis

Since there are no biological markers to diagnose schizophrenia, the diagnosis is based on self-reported symptoms and behavioural observations. There are two sets of diagnostic criteria used to identify schizophrenia: the Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (DSM-IV-TR) (American Psychiatric Association 2000) and the International Classification of Diseases, tenth edition (ICD-10) (World Health Organisation 1993). They both require that positive symptoms are present for at least one month causing dysfunction in everyday life. Drug-induced psychoses and organic causes should be excluded.

Although specific criteria should be met for a diagnosis of schizophrenia, there is still great heterogeneity and different subtypes are recognised:

- **Paranoid type:** the most common subtype where delusions and hallucinations are evident
- **Catatonic type**: Rarer, characterised by immobility, rigidity or purposeless movement.
- **Disorganised (or hebephrenic) type**: Characterised by unpredictable behaviour, usually the prognosis is poor.
- **Undifferentiated type**: Psychotic symptoms are present but the criteria for paranoid, disorganised or catatonic types have not been met.

- **Residual (chronic) type**: there is a history of psychotic episodes but at the moment only negative symptoms and/or cognitive impairment is present

3.1.4 Prognosis

Cohort studies have found that ~15% of patients recover completely, ~50% have relapses but minimum impairment between the relapses, 25% have chronic illness and significant impairment and 10% commit suicide. The prognosis is worse in patients who discontinue medication, are socially isolated, have an earlier age of onset or abuse drugs. Men have a poorer prognosis than women and more negative symptoms (Katona et al. 2008). It is interesting that cognitive deficits increase with age in patients with schizophrenia, with 50% of older patients with schizophrenia meeting criteria for dementia but not Alzheimer's disease (Pearlson 2000).

3.1.5 Management of patients with schizophrenia

Management of patients with schizophrenia requires administration of antipsychotic medication. There are two types of antipsychotics: typical and atypical; they are both effective in treating positive symptoms and preventing relapses. Adherence to medication is a problem and discontinuing medication increases the chance of a relapse. However, antipsychotics have side-effects which can be serious. Psychological and social interventions are also beneficial.

3.1.6 Phenotypic complexity

Although the current diagnostic criteria provide diagnostic reliability, they are not valid disease entities. The complexity of the phenotype suggests that schizophrenia is actually a collection of conditions with different symptoms and outcomes. More importantly the underlying causes might be different with important consequences for treatment.

Genetic studies are especially susceptible to this lack of biological validity. Searching for genes conferring susceptibility to complex disorders is hard enough without the

heterogeneity of schizophrenia. If the current diagnostic criteria encompass a variety of conditions with different aetiology, then the situation resembles the hypothetical situation of considering fever a disorder on its own and looking for a cause.

The other side of this problem is that the phenotype extends beyond the diagnosis of schizophrenia to include schizoaffective disorder and schizotypal personality disorder, as shown by family studies (Owen et al. 2002). Then, the search for susceptibility factors would be confounded by the fact that individuals with these disorders would not be included in the "affected" group.

Another layer of complexity is added by the overlap of schizophrenia and bipolar disorder. Current diagnostic criteria have been based on the Kraepelinian dichotomy between schizophrenia and manic depressive disorders. However, family and twin studies have shown higher than expected rates of bipolar disorder in relatives of patients with schizophrenia and the opposite. A recent meta-analysis provides further evidence of familial coaggregation of schizophrenia and bipolar disorder (Van Snellenberg and de Candia 2009). Schizoaffective disorder occurs more often in families of patients of both disorders. In addition, the largest family study of schizophrenia and bipolar disorder involving 2 million Swedish families showed that relatives of patients with schizophrenia and bipolar disorder have increased risk of both disorders and they attributed this to partly shared common genetic causes (Lichtenstein et al. 2009). Similar to family studies, molecular genetic studies have found evidence of shared common genetic causes between schizophrenia and bipolar disorder. Overlapping regions have been reported by linkage scans of both disorders, while association studies of *DISC1* (Disrupted-in-Schizophrenia 1) and *NRG1* (Neuregulin1) suggest common susceptibility factors (see Craddock et al. 2006a for a review). Most interestingly the *ZNF804A* (zinc finger protein 804A) gene, which is the best-supported finding from a genome-wide association study on schizophrenia, shows evidence of association with bipolar disorder (O'Donovan et al. 2008). Additionally, the best-supported gene from a genome-wide association study on bipolar disorder, *CACNA1C* (calcium channel, voltage-dependent, L type, alpha 1C subunit), is suggested to be associated with schizophrenia as well (Ferreira et al. 2008; Green et al. 2009).

The conclusion that can be made here is that schizophrenia is a very heterogeneous disorder and there is a possibility that schizophrenia and bipolar disorder have more similarities than differences. Until aetiology is better understood, categorising patients according to diagnostic criteria will be limiting for both clinicians and researchers. The way molecular genetic studies can help advance the field of phenotype refinement and diagnostic validity is by "thinking outside the box" in terms of choosing cohorts with patients and controls. Instead of strictly relying on the diagnostic criteria, patients from a broad phenotype or across disorders can be chosen based on common characteristics (for example, performance on a cognitive task). This way the sample size will still remain large enough and common risk factors that are more likely to be closer to being causal can be identified.

3.2 Schizophrenia as a neurodevelopmental disorder

The fact that schizophrenia is currently described and diagnosed based only on symptoms does not mean that there is no underlying neuropathology, although this might be heterogeneous. Current views suggest that schizophrenia is a neurodevelopmental disorder with the cause of the abnormality occurring early in life, most likely during prenatal life. Cell migration abnormalities can start during foetal life in individuals already susceptible to schizophrenia due to genetic factors and they can be triggered by an environmental insult. Certain environmental factors have been found to be associated with schizophrenia, although the mechanism through which they trigger these brain abnormalities is not known. In addition, although brain abnormalities exist in patients with schizophrenia, it is very difficult to establish whether they are causal or they are caused by the disease or its treatment.

One of the neuropathological findings in patients with schizophrenia is that cortical thickness and brain weight is reduced by 5% (Pearlson 2000). Another consistent finding is that patients with schizophrenia have increased lateral ventricle size. This has been found in patients at the onset of the disorder making it less likely to be a disease consequence (Degreef et al. 1992). In a systematic review it was calculated that ventricular size was increased by 40% in patients with schizophrenia compared to controls (Lawrie and Abukmeil 1998). Family studies strengthen these results by showing that obligate carriers of susceptibility factors, such as children with two

affected parents also have larger ventricles (Harrison 1999). This finding, despite providing a glimpse into the pathology of this extremely complicated disorder, is not specific enough to be used as a biological marker because it can be found in healthy individuals as well.

Neuroimaging techniques have shown alterations in all brain regions in different types of patients with schizophrenia with or without medication (Pearlson and Marsh 1999). The prefrontal cortex is characterised by altered neuronal sizes and abnormal synaptic connections (Pearlson 2000). Functional imaging suggests abnormal activity in the fronto-hippocampo-parietal circuit when carrying out a set-shifting task. Deficit symptoms are more correlated with reduced activity in frontal and parietal cortex (Pearlson 2000).

Brain abnormalities have also been found microscopically with the most interesting being a decrease in neuronal size accompanied by increased neuronal density affecting the hippocampus and the dorsolateral prefrontal cortex (Harrison 1999).

Reported neurochemical abnormalities involve the dopamine, serotonin, GABA and glutamate systems. The dopamine hypothesis for schizophrenia suggests that the disorder is caused by increased dopaminergic transmission and it is based on neuropharmacological findings. Dopamine agonists, such as amphetamines, can provoke or exacerbate positive symptoms while the efficacy of conventional antipsychotic drugs is proportionate to blockage of dopamine receptors. It has also been suggested that the neurocognitive deficits in schizophrenia are linked to reduced dopaminergic activity in the prefrontal cortex (Pearlson 2000). The only safe conclusion that can be made from the literature is that neurotransmitter systems are definitely dysregulated in schizophrenia. However, treating this dysregulation treats the symptoms but not the cause of the disorder.

Neurocognitive deficits have been suggested to be the best candidate markers of schizophrenia. A large longitudinal study found that cognition in people who developed schizophrenia in adulthood was affected from as early as seven years of age. Reading and IQ were particularly impaired (Crow et al. 1996).

The neurodevelopmental hypothesis suggests that environmental influences can act early in life to modify risk to the disorder. Some of the environmental factors that

have been studied in schizophrenia include obstetric complications, prenatal maternal psychological stress, maternal influenza exposure during pregnancy and maternal nutritional deficiency during pregnancy (Kinney et al. 1994; St Clair et al. 2005; Susser et al. 1996).

To sum up, schizophrenia is a disorder that occurs usually in early adulthood but its origin is neurodevelopmental with dysregulation starting in prenatal life. The following evidence strengthens this notion:

- Neuropathological evidence points to structural brain abnormalities
- Neuroimaging evidence suggests that anomalies exist in different brain areas in patients of all stages
- Neurocognitive deficits are present before the onset of the disorder from a very young age
- Environmental factors acting during prenatal life, such as pregnancy complications, have been linked to schizophrenia

3.3 Evidence of a genetic component to schizophrenia

Genetic epidemiological studies of schizophrenia have provided compelling evidence of a genetic component to schizophrenia.

3.3.1 Family studies

Family studies examining the risk of developing schizophrenia in relatives of patients with the disorder compared to the general population suggest that genetic factors increase risk. Using 40 reliable family studies conducted in Western Europe between 1920 and 1987, Gottesman (1991) showed that the risk of developing schizophrenia increases from 1% in the general population to 6% in second degree relatives, 9% in siblings, 13% in children, 17% in dizygotic (DZ) twins and 48% in monozygotic (MZ) twins (Figure 3-1). In addition, the risk increases with the number of affected

relatives. Children with parents that both suffer from schizophrenia have 46% risk of developing the disorder (Gottesman 1991).

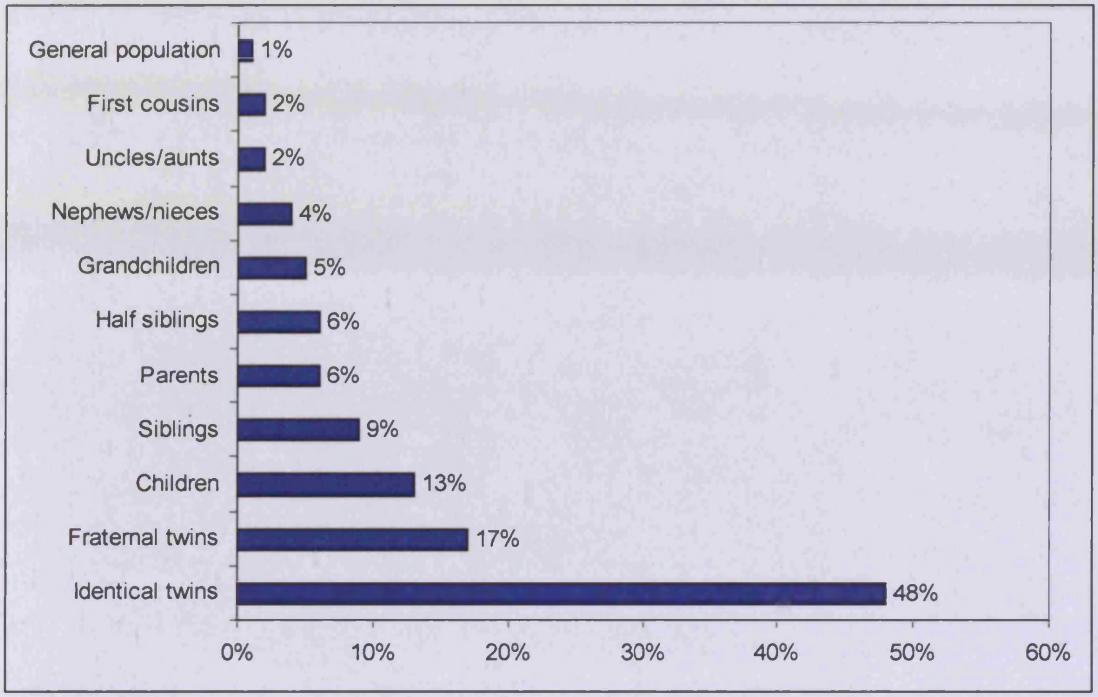

Figure 3-1. Risk of developing schizophrenia in relatives of patients with schizophrenia. The data are based on a review of studies compiled by Gottesman (1991)

3.3.2 Adoption studies

By examining the risk of developing the disorder in biologically related relatives and non-biologically related (adoptive family) relatives, it is possible to disentangle the contribution of genetic and environmental factors, which is not possible in family studies. Adoption studies of schizophrenia agree that there is an increase in the risk in biological relatives of patients with schizophrenia compared to non-biological relatives. The Danish adoption study, for example, found that the rate of schizophrenia in first-degree biological relatives of adopted patients with schizophrenia was 7.9% compared with 0.9% for biological relatives of adopted control individuals (Kendler et al. 1994).

3.3.3 Twin studies

The level of disease concordance in pairs of MZ twins compared to DZ twins has been used to calculate the heritability of schizophrenia. A review based on five systematically ascertained twin studies found a rate of concordance in MZ twins of 41-65% and 0-28% in DZ twins (Figure 3-1), thus calculating a heritability of 80-85% (Cardno and Gottesman 2000).

To sum up, family, adoption and twin studies have confirmed a major role of genetic factors in the pathophysiology of schizophrenia. The risk of developing schizophrenia increases with the degree of relatedness (and thus the number of genetic factors shared). However, the rate of concordance in MZ twins, who share all of their genes, never reaches 100% providing strong evidence for the involvement of environmental factors in the aetiology of schizophrenia.

3.3.4 Mode of transmission

The rate of concordance of schizophrenia in MZ twins is approximately 48%, while the degree of genetic similarity is clearly more than that. This means that the same genetic factors do not always lead to the phenotype being expressed. Mechanisms involving interaction with environmental or stochastic factors are very likely to operate. The pattern of disease recurrence in relatives of patients with schizophrenia excludes the possibility of a single-gene Mendelian disorder (Owen et al. 2002). The mode of transmission of schizophrenia is most likely complex reflecting a polygenic model with a threshold effect. According to this model, each risk or protective allele exerts a small effect. These effects are combined in an additive manner along with environmental factors to lead to disease manifestation when a critical threshold is exceeded (Owen et al. 2002). It is likely that these risk factors are relatively common alleles of small effect and some rare alleles with relatively large effects (Owen et al. 2009). Risch (1990) studied the relative risk of schizophrenia in different families. The relative risk (λ_s for siblings) is a ratio of the probability of a relative of a patient with schizophrenia being affected to the probability of being affected in the general population. He concluded that the existence of one locus of $\lambda_s>3$ is unlikely and that a model of two or more loci of $\lambda_s<2$ are more likely (Risch 1990). However, the current

model of common variants of small effects does not exclude the existence of rare variants with larger effects acting on a small number of affected individuals. In addition, epistatic effects are likely to be involved in schizophrenia, including gene-gene and gene-environment interactions. Epistatic interactions mean that the susceptibility conferred by two factors (genetic or environmental) is greater than their individual susceptibilities (Frankel and Schork 1996). In general, although it has been demonstrated that schizophrenia is heritable, the number of susceptibility loci, the risk conferred by them and the degree of interaction between them remain unknown.

3.4 Molecular genetics of schizophrenia

3.4.1 Genome-wide linkage studies

Genome-wide linkage studies of schizophrenia started by using a small number of families with multiple affected individuals. Because of the lack of highly penetrant alleles, larger numbers of pedigrees have been used. However, it has not been possible to consistently replicate linkage findings (Owen et al. 2005). The large number of linkage scans made it possible for meta-analyses to be performed (Table 3-1).

Table 3-1. Summary of meta-analyses of schizophrenia genome-wide linkage scans (regions that show overlap are in bold)

Meta-analysis	No of studies included	No of pedigrees	Chromosomal regions	Evidence of linkage
Badner and Gershon (2002)	18	681	**8p**	$P=2*10^{-4}$
			13q	$P=7*10^{-6}$
			22q	$P=9*10^{-6}$
Lewis et al. (2003)	20	1,208	**2q**	P=0.000417
			5q, 3p, 11q, 6p, 1q, 22q, **8p**, 20q, 14p	P<0.05
Ng et al. (2008)	32	3,255	5q	P=0.0046
			2q	P=0.00035

The first meta-analysis supported evidence of linkage on 8p, 13q and 22q (Badner and Gershon 2002). The second meta-analysis used more stringent criteria for the studies to be included (Lewis et al. 2003). The strongest signal from this meta-analysis was on 2q, while support was obtained for 5q, 3p, 11q, 6p, 1q, 22q, 8p, 20q and 14p. They also found that the number of loci meeting the aggregate criteria for significance was greater than the number of loci expected by chance ($p<0.001$) (Lewis et al. 2003). Two regions were common in both meta-analyses: 8p and 22q. The large number of non-overlapping regions can be attributed to the difference in the criteria for inclusion in the meta-analysis.

The most recent meta-analysis included 32 independent genome-wide linkage scans, comprising 7,476 genotyped cases. More nominally significant linkages than would be expected by chance were reported. Genome-wide evidence of linkage was detected on chromosome 2q, while suggestive evidence of linkage was observed on chromosome 5q (Ng et al. 2008). The region on chromosome 2q is of particular interest because it contains *ZNF804A*, the gene containing a SNP that has been significantly associated with schizophrenia in a recent genome-wide association study (O'Donovan et al. 2008). When the authors included only samples of European ancestry, suggestive evidence of linkage was detected on chromosome 8p (Ng 2008). Suggestive evidence of linkage was also found for this region in a recent genome-wide linkage scan involving eight clinical samples and 1,615 affected genotyped individuals (Holmans et al. 2009).

To sum up, the modest linkage signals show that it is unlikely that there are genes of large effects on the risk of developing schizophrenia. In the era of genome-wide association studies the usefulness of linkage scans is questioned. Although genome-wide association studies have better power to detect common SNPs and they can also detect some CNVs, linkage scans can find evidence of the aggregate effect of multiple rare and common variants as well as CNVs. Thus, following linkage regions up by high-throughput resequencing could be another way of finding genes that contribute to the risk of developing schizophrenia.

3.4.2 Candidate gene studies

Candidate gene studies focus on selecting specific genes based on assumptions about the pathophysiology of the disorder derived from animal, pharmacological, or imaging studies (functional studies). The other approach is to choose genes following interesting areas that have come up from linkage scans (positional). There have been a large number of association studies based on hypotheses about the pathophysiology of schizophrenia but they have not been very successful, probably due to the fact that very little is known about the pathophysiology of schizophrenia. Positional candidate gene studies have been more successful in schizophrenia. Over 1,000 genetic association studies have been published about schizophrenia, thus I will be focusing on the studies that have provided the most convincing results.

3.4.2.1 Dystrobrevin-binding protein 1 (DTNBP1)

Dystrobrevin-binding protein 1 (*DTNBP1*) or dysbindin is located on chromosome 6 and was first implicated in schizophrenia after Straub et al. (2002) followed up a linkage region on chromosome 6p22.3. The same group followed their finding by genotyping *DTNBP1* in 203 German families (Schwab et al. 2003). Two large studies, one including 900 cases from United Kingdom and Ireland (Williams et al. 2004) and one including 488 families from Bulgaria (Kirov et al. 2004), also confirmed the association. The majority of published studies on *DTNBP1* have found positive associations, thus rendering *DTNBP1* one of the strongest candidates for involvement in schizophrenia (Kirov et al. 2005). However, there is a great deal of disagreement over the specific SNP alleles or haplotypes that confer risk or protection to schizophrenia. One of the reasons for the discrepancies could be that there are multiple risk or protective alleles in this gene (*DTNBP1* is a large gene with over 300 SNPs) (Williams et al. 2004). Interestingly, *DTNBP1 cis*-acting variants affect expression in the human brain (Bray et al. 2003; Bray et al. 2005). In addition, reduced mRNA and protein has been found in postmortem brain samples from patients with schizophrenia (Talbot et al. 2004; Weickert et al. 2004).

The mechanism through which *DTNBP1* confers risk or protection against schizophrenia is not clear because its function is not entirely known. *DTNBP1* binds

to both α- and β-dystrobrevin which form the dystrophin glycoprotein complex. Its expression in the brain is confined to neurons where it is found both pre- and postsynaptically (Williams et al. 2005). The presynaptic DTNBP1 is reduced in glutamergic neurons of the hippocampus of patients with schizophrenia (Talbot et al. 2004). This finding coupled with evidence that glutamate is reduced in cultured neurons with reduced DTNBP1 expression suggests that *DTNBP1* could confer risk to schizophrenia by influencing the glutamate system.

3.4.2.2 Neuregulin1 (NRG1)

Neuregulin 1 (*NRG1*) was first considered to be involved in schizophrenia after fine-mapping across the 8p locus, to which linkage was observed in Icelandic families (Stefansson et al. 2002). The haplotype associated with schizophrenia in this first sample was also found to be associated in samples of different origin. There have been negative associations as well that could reflect the difference in the origin of the samples (Owen et al. 2005). The specific susceptibility variant has not been discovered yet, although there have been resequencing efforts (Stefansson et al. 2002). *NRG1* codes for many different mRNA species resulting in 15 proteins with diverse functions in the brain. The possible mechanism of *NRG1* involvement in schizophrenia is by altered mRNA expression supported by the finding that there are altered ratios of NRG1 mRNA species in the brain of people with schizophrenia (Hashimoto et al. 2004) and findings from a mutant mouse model heterozygous for *NRG1* (Stefansson et al. 2002).

3.4.2.3 SzGene: a database of association studies of schizophrenia

The literature on association studies of schizophrenia includes more than 1,000 studies with very inconsistent results. The recent creation of an online database of all published genetic association studies of schizophrenia is an effort to systematically record this literature. This database, called SzGene, is regularly updated, free to access and where possible it provides meta-analyses (Schizophrenia Research Forum 2005) (www.schizophreniaforum.org/res/sczgene/default.asp). Using the data available on April 2007, Allen et al. (2008) performed 118 meta-analyses resulting in

24 variants in 16 different genes being nominally associated. The strongest associations were with *DRD2* (dopamine D2 receptor) gene and *TPH1* (tryptophan hydroxylase 1). The authors, additionally, applied a set of guidelines developed by the Human Genome Epidemiology Network to assess the validity of these associations. Strongest credibility was found for SNPs in *DRD1*(dopamine D1 receptor), *DTNBP1*, *TPH1* and *MTHFR* (methylenetetrahydrofolate reductase) (Allen et al. 2008). The association of polymorphisms in two dopamine receptors is interesting given the hypothesis for excess dopamine in the brain of patients with schizophrenia and the correlation of efficacy of anti-psychotic drugs with the ability to block the DRD2 receptor (Williams et al. 2009a). In addition, the polymorphism in the *DRD2* receptor is functional resulting in altered mRNA stability (Duan et al. 2003). The *DTNBP1*gene is a strong candidate for involvement in schizophrenia, although there could be multiple risk and protective alleles or great heterogeneity in this gene. *TPH1* encodes for tryptophan hydroxylase, the rate limiting enzyme in the synthesis of serotonin. However, *TPH2* (tryptophan hydroxylase 2) and not *TPH1* is expressed in the brain (Walther et al. 2003).

3.4.3 Genome-wide association studies (GWAS)

Genome-wide association studies are considered very promising in the field of genetics of complex disorders, since they have the power to detect moderate to small effect sizes (provided that studies are large enough) while at the same time making no assumptions about the pathophysiology of the disorder. Markers are spread across the genome in order to achieve maximum coverage. However, the Y chromosome has been an exception up until recently (see following chapters for more information).

Genome-wide association studies on schizophrenia have been published (Table 3-2) using pooled DNA samples (Kirov et al. 2008b; Mah et al. 2006; Shifman et al. 2008) and individual genotyping (Lencz et al. 2007; O'Donovan et al. 2008; Shi et al. 2009; Stefansson et al. 2009; Sullivan et al. 2008; The International Schizophrenia Consortium 2009).

Table 3-2. Summary of genome-wide association studies for schizophrenia published so far

Genome-wide association study	Study design	Cases	Controls	Evidence of association	P value	OR
Mah et al. (2006)	DNA pooling	320	325	*PLXNA2* (chr 1)	0.006	1.49
Lencz et al. (2007)	Individual genotyping	178	144	*CSF2RA* (PAR1)	$3.7x10^{-7}$	3.23 (2.04-5.15)
Kirov et al. (2008b)	DNA pooling	574	605	*CCDC60* (chr 12)	$2x10^{-6}$	
O'Donovan et al.(2008)	Individual genotyping	479	2,937	*ZNF804A* (chr 11)	$1.61x10^{-7}$	1.09
Shifman et al. (2008)	DNA pooling	2,274	4,401	*Reelin* (chr 7)	$2.9x10^{-5}$	1.58 (1.31-1.89)
Sullivan et al. (2008)	Individual genotyping	738	733	No genome-wide significant results	-	-
Need et al. (2009)	Individual genotyping	1,460	12,995	No genome-wide significant results	-	-
The International Schizophrenia Consortium (2009)	Individual genotyping	3,322	3,587	MHC genes (6p22.1) *TCF4*	$9.5x10^{-9}$ $1x10^{-7}$	1.22
Shi et al. (2009)	Individual genotyping	2,681	2,653	-	-	-
Shi et al. (2009)	Meta-analysis	8,008	19,077	MHC genes (6p22.1, 6p21.32)	$9.54x10^{-9}$	1.14
Stefansson et al. (2009)	Individual genotyping	2,663	13,498	-	-	-
Stefansson et al. (2009)	Meta-analysis	12,945	34,591	MHC, *PRSS16* MHC, *NOTCH4* *NRGN*	$1.4x10^{-12}$ $1.4x10^{-12}$ $2x10^{-10}$	1.16

PLXNA2: plexin A2; CSF2RA: colony stimulating factor, receptor 2α; CCDC60: coiled-coil domain containing 60; ZNF804A: zinc finger protein 804A; MHC: Major Histocompatibility Complex; TCF4: Transcription Factor 4; PRSS16: Thymus-specific serine protease; NOTCH4: Neurogenic Locus notch Homolog protein 4; NRGN: Neurogranin; OR: Odds Ratio

In DNA pooling, samples from cases are mixed and samples from controls are mixed separately. The analysis involves looking for differences in allele frequencies. This method has the advantage of saving time and money but reduces power because it is prone to measurement error. It is very interesting that one of the pooling genome-wide association studies of schizophrenia achieved a sex-specific association. The association was female-specific for a SNP in the reelin (*RELN*) gene ($p=2.9x10^{-5}$) using an Ashkenazi Jewish population (Shifman et al. 2008). A significant gene-sex interaction was also found ($p=1.8x10^{-4}$). The same group attempted to replicate their initial association in four different populations. Although the effects were in the same direction in all the populations, they were significant in one of them achieving an overall relative risk for women carrying the common allele of 1.58 ($p = 8.8 \times 10^{-7}$; $p = 1.6 \times 10^{-5}$ for gene-sex interaction). This finding apart from its importance in studying sex differences in schizophrenia supports the idea of a neurodevelopmental nature to the disorder, since reelin is involved in corticogenesis and neuronal positioning (Shifman et al. 2008).

In terms of individual genotyping, the first genome-wide association study on schizophrenia was performed on 178 cases and 144 controls. The strongest result was for a SNP near the *CSF2RA* (colony stimulating factor, receptor 2α) gene in the pseudoautosomal region ($p=3.7x10^{-7}$) (Lencz et al. 2007). Sequencing in a smaller independent sample revealed association with both common haplotypes and rare variants in *CSF2RA* and the neighbouring *IL3RA* (interleukin 3 receptor alpha). Although the association is surprising in such a small sample, the link between schizophrenia and the immune system is one that definitely deserves further studying (Lencz et al. 2007) and appears in other genome-wide association studies, as will be discussed later.

The second genome-wide association study on schizophrenia involved genotyping 738 cases and 733 controls but there were no SNPs achieving genome-wide significance (Sullivan et al. 2008). Follow-up studies might yield something interesting but the heterogeneity of the sample in terms of origin can be a problem. A genome-wide association study of 1,460 cases and 12,995 controls also failed to report any genome-wide significant findings (Need et al. 2009).

Another genome-wide association study was designed to follow up all results from the initial sample of 479 cases and 2,937 controls surpassing a threshold of $p<10^{-5}$ in a sample of 6,829 cases. The strongest result was for a SNP in *ZNF804A* (zinc finger protein 804A) ($p=1x10^{-7}$), although it just fell short of genome-wide significance (O'Donovan et al. 2008). However, genome-wide significance was achieved when the samples included patients with bipolar disorder ($p=9x10^{-9}$, OR=1.09). *ZNF804A* is a transcription factor with unknown function. Its discovery through a genome-wide association study shows that this study design without any *a priori* hypothesis can point to genes that could not have been found otherwise. The other important point stemming from this study was that the observed pattern of replication was highly improbable given the null hypothesis ($p=9x10^{-8}$) (O'Donovan et al. 2008). This means that there is at least one common variant increasing risk to schizophrenia and bipolar disorder, thus the hypothesis that common variants would be eliminated through the reduced fecundity of people with schizophrenia is not true. The most likely scenario would involve a large number of common variants with very small effects (<1.1) as well as rare variants of larger effects (Williams et al. 2009b).

More recently genome-wide association studies on schizophrenia have involved larger sample sizes and collaborative efforts to achieve them. The International Schizophrenia Consortium has published a genome-wide association study in 3,322 cases and 3,587 controls with even larger numbers for the replication sample (The International Schizophrenia Consortium 2009). The most interesting finding involved a region on chromosome 6p22.1 spanning the major histocompatibility complex (MHC) ($p=9.5x10^{-9}$). The authors also evaluated the role of common variants in schizophrenia and provided evidence for a substantial polygenic component to the risk of schizophrenia and bipolar disorder. Thousands of common variants are likely to contribute to risk and collectively they explain at least one-third of the variability (The International Schizophrenia Consortium 2009).

No genome-wide significant results were reported in the study using the Molecular Genetics of Schizophrenia (MGS) case-control sample (2,681 cases and 2,653 controls) (Shi et al. 2009). However, in a meta-analysis, which extended the original sample to 8,008 cases, SNPs on chromosomal region 6p22.1 were significantly associated with schizophrenia ($p=9.54x10^{-9}$). The region includes a histone gene

cluster and MHC genes (Shi et al. 2009). Another study using the SGENE Consortium sample (2,663 cases and 13,498 controls) failed to report genome-wide significant results (Stefansson et al. 2009). However, in the meta-analysis, which included the International Schizophrenia Consortium (ISC), the SGENE Consortium and the MGS samples, five SNPs spanning the MHC region on chromosome 6p21.3-22.1 reached genome-wide significance levels (strongest finding: $p=1.4x10^{-12}$, OR=1.16) as well as a SNP on chromosome 11q24.2 upstream of the neurogranin gene and a SNP in transcription factor 4 (TCF4) on chromosome 18q21.2 (Stefansson et al. 2009).

Despite the initial encouraging results from genome-wide association studies on schizophrenia it is clear that a large number of common variants remain to be found. As is the case for genome-wide association studies from other complex diseases, large scale endeavours are necessary. However, a potential problem with schizophrenia and other psychiatric disorders is that increasing the sample size could lead to an increase in heterogeneity, thus masking the already small effects.

3.4.4 Copy number variants (CNVs)

Copy number variants (CNVs) were first linked with schizophrenia due to the increased rates of schizophrenia in individuals with velo-cardio-facial syndrome, which is caused by microdeletions of chromosome 22q11 (Murphy and Owen 2001). The largest study of CNVs in schizophrenia involving 3,391 cases and 3,181 controls found that the total burden of CNVs is increased 1.15-fold in patients with schizophrenia compared to controls (The International Schizophrenia Consortium 2008). The increase was greater for rare and larger CNVs. Other studies report different effects but it is expected, since they differ in design and sample composition. However, it is not yet clear whether the increase in burden comes from a small number of highly penetrant CNVs or a large number of variants with small effects.

In terms of the rate of *de novo* CNVs in schizophrenia, studies have not been in agreement. Two studies reported a very low level of *de novo* CNVs in schizophrenia (Kirov et al. 2008a; Kirov et al. 2008b) and childhood-onset schizophrenia (Walsh et

al. 2008), while in another study the rate of *de novo* CNV events was 10% in cases and 2% in controls (Xu et al. 2008).

CNVs in specific loci have been associated with schizophrenia, in some cases in more than one study. The most interesting is a rare transmitted deletion that affects neurexin 1 (*NRXN1*) found by Kirov (2008a) in a patient with schizophrenia and an affected sibling. CNVs in *NRXN1* were also reported in two other studies of schizophrenia (Rujescu et al. 2009; Walsh et al. 2008), as well as autism (Autism Genome Project Consortium 2007). CNVs have also been found in genes interacting with *NRXN1*, such as a *de novo* duplication including the gene coding for amyloid beta A4 precursor binding protein, which binds *NRXN1* (Kirov et al. 2008a) and a rare CNV in neuroxophilin 2 (Kirov et al. 2009). In addition, there is a biological reason why disruptions in the *NRXN1* gene caused by CNVs could lead to increased risk to schizophrenia, since it is involved in synaptic development and function.

Other interesting CNVs associated with schizophrenia are two rare deletions, one at 1q21.1 and the other at 15q13.3 both with very strong effects (OR estimated to be around 10 across studies) (Kirov et al. 2009). These deletions have been reported by two large studies of CNVs in schizophrenia, one involving 1,443 cases (Stefansson et al. 2008) and the other 3,391 cases (The International Schizophrenia Consortium 2008). These regions include a large number of genes and it is not clear at this point, which of them are responsible for increasing risk to schizophrenia. The same studies have reported another deletion at 15q11.2 with a smaller effect size (OR=3) (Kirov et al. 2009; Stefansson et al. 2008). Microduplications with large effect sizes (OR=14.5) at 16p11.2 (McCarthy et al. 2009) and 16p13.1 (OR=3) have also been reported (Ingason et al. 2009).

The involvement of structural variants with low frequency in schizophrenia provides further support to the notion that schizophrenia is a neurodevelopmental disorder, especially since the same loci have been implicated in other neurodevelopmental disorders, such as autism and mental retardation (Mefford et al. 2008). However, these results show that the same structural variant can cause a variety of phenotypes, including healthy individuals and individuals with serious neurodevelopmental disorders. This means that none of the CNVs are sufficient to cause a disease and

points to the involvement of other factors, such as the rest of the genetic background and environmental factors, interacting with CNVs to lead to a specific phenotype.

To conclude, schizophrenia is a psychiatric disorder with a neurodevelopmental nature and increased phenotypic heterogeneity. The pathophysiology of the disorder is still not clear. The genetic component of schizophrenia includes common variants of small effect sizes and rare variants of large effect sizes, although more research is required to elucidate the role of risk factors and their interactions.

Chapter 4 The Y chromosome

4.1 The evolution of sex chromosomes

In mammals sex is determined by two sex chromosomes, XX gives rise to females while XY with the sex-determining gene, SRY, results in males. Although pairs of autosomes are highly homologous in mammals, sex chromosomes are heteromorphic, which means that they share a very small amount of homology. However, they can pair during meiosis in male cells due to two homologous regions at both ends of the sex chromosomes, called pseudoautosomal regions (PARs). Homologous regions occur in other places on the two sex chromosomes, which suggests that sex chromosomes have evolved from an ancestral pair of autosomes. This happened when an allelic variation on a gene on one of the ancient autosomes resulted in the gene taking on control of sex determination. The autosome with this specific variation evolved into the male-specific chromosome, the Y chromosome, while the other autosome developed as the X chromosome (for more details see Graves 2006).

After the development of the sex determining gene, the two sex chromosomes started becoming increasingly dissimilar and recombination was repressed to preserve a male specific chromosome. This absence of recombination allowed the accumulation of mutations on the male specific chromosome and loss of homology between the two sex chromosomes. Lack of recombination resulted in degeneration of the Y chromosome. This explains the smaller number of genes on the Y chromosome and the fact that most of them are involved in male functions (Graves 2006).

In birds and reptiles, sex determination is either genetic or environmental. Distinct sex chromosomes have independently evolved not only in mammals but also birds (where females are the heterogametic sex, ZW, and males are the homogametic sex, ZZ) and certain species of fish and reptiles (although for most of them sex determination depends on environmental factors).

The sex determining gene is *SRY* (Sex-determining region of chromosome Y) in mammals and it was thought to have evolved 350 million years ago since the divergence of synapsid reptiles. This a branch of reptiles from which mammals evolved and which left no other descendants. It was thought that the SRY gene must

have developed after the divergence of mammals from reptiles, as it is only found in mammals and no reptile species. Birds and snakes, for example, have a ZZ male and ZW female. The sex determining gene in birds is called *DMRT1* (Doublesex and mab-3 related transcription factor 1) (Delbridge and Graves 1999).

However, the analysis of the platypus genome showed that the SRY must have evolved earlier, ~ 165 million years ago (Veyrunes et al. 2008). The platypus belongs to the monotremes, which diverged from therian mammals (marsupials and placental) ~ 165 million years ago and therefore falls evolutionary-wise between birds/reptiles and mammals. The platypus sex chromosomes have strong homology with bird, but not to therian sex chromosomes and orthologes of conserved genes on the X chromosome map on platypus chromosome 6. This means that this is probably the ancestral autosome from which the sex chromosomes evolved (Veyrunes et al. 2008).

The *SRY* gene evolved from *SOX3* (SRY-related HMG-box 3), which is found on the X chromosome in mammals. Both *SRY* and *SOX3* belong to a family of HMG (high-mobility group) domain transcription factors. The *SRY* sequence is poorly conserved outside the HMG box, which suggests that this is the most important region for sex determination. *SOX3* is highly conserved in mammals and is expressed in the developing brain as well as other tissues of mice and humans. Thus, *SRY* evolved from a gene with a possible role in brain function.

To conclude, the sex chromosomes evolved from a pair of autosomes when a mutation in a gene appeared that enabled this gene to be responsible for male sex determination.

4.2 The Y chromosome structure

Although genome-wide linkage studies and genome-wide association studies are designed to achieve the maximum possible coverage across most of the genome, up until recently they did not include any marker on the Y chromosome. This means that the Y chromosome has been neglected in studies of complex disorders including psychiatric disorders and the association studies that have been performed of the Y chromosome in relation to psychiatric phenotypes have mostly used small and heterogeneous populations.

The presence or absence of the Y chromosome defines the sex of an individual. It is 60 Mb in length but it has been considered a genetic wasteland for many years because of the paucity of phenotypes showing solely father-to-son transmission and the small number of identified Y chromosome genes. The discovery of the sex determining gene, SRY (Sex Determining Region on chromosome Y) (Gubbay et al. 1990; Sinclair et al. 1990), led to the notion that the Y chromosome was essential for sex differentiation but had lost all other important functions during evolution. Nowadays with the improvement of genomic technologies, it has emerged that the Y chromosome harbours a number of genes and informative polymorphisms.

The part of the Y chromosome found only in men is called the non-recombining region (NRY) or male-specific region (MSY). This region covers the vast majority of the Y chromosome (95%) and is passed unchanged from father to son apart from mutations because there is no recombination with the X chromosome.

4.2.1 Pseudoautosomal regions (PARs)

The pseudoautosomal regions (PARs) are located at both ends of the Y chromosome and they recombine respectively with PAR1 and PAR2 on the X chromosome. These regions cover only 2.7 Mb for PAR1 and 0.33 Mb for PAR2. There are 30 genes reported in the pseudoautosomal regions (Ross et al. 2005) and the *SHOX* (short stature homeobox) gene has been associated with short stature and bone malformations (Blaschke and Rappold 2006). Early linkage studies of the PAR regions in schizophrenia have failed to report significant results (Asherson et al. 1992). In recent years, the PAR regions have been overlooked in linkage and association studies and have not been included in the early versions of SNP arrays used for genome-wide association studies. Due to the high level of recombination in these regions, the numbers of markers required to cover them should be larger than for autosomal regions of the same size (Flaquer et al. 2008). Nevertheless, the cost would not increase considerably compared to the overall cost of genome-wide association studies and for this reason the latest Illumina Human1M-Duo chip (http://www.illumina.com/pages.ilmn?ID=40) covers the PAR regions adequately.

4.2.2 Male-specific region (MSY)

Since the MSY is the factor on the Y chromosome that differentiates the sexes, it should be discussed in more detail. The MSY comprises both heterochromatic sequences and euchromatic sequences. Heterochromatic DNA remains highly condensed throughout the cell cycle and genes on it are silenced through histone methylation. Euchromatic DNA, however, is lightly packed and is responsible for all gene transcription from the Y chromosome. The latter cover 8 Mb on Yp and 14.5 Mb on Yq. Currently, 156 transcription units have been found on the human euchromatic MSY. The number of protein-coding genes is 78 and they code for 27 proteins or protein families. It is interesting that only 18 genes are found in a single copy. The remaining genes have several copies with at least 98% sequence identity (Skaletsky et al. 2003). Genes in the MSY fall into two categories: genes that are expressed throughout the body (including the brain) and genes that are expressed mainly or exclusively in testes.

In terms of euchromatic sequences in the human MSY, there are three different types (Figure 4-1) (Skaletsky et al. 2003):

- X-transposed sequences have 99% homology to X chromosome, although they do not recombine and they are characterised by low gene and high repeat density. They are thought to be the remnants of the ancient autosomes from which the Y chromosome, as well as the X, have evolved.

- X-degenerate sequences include single-copy genes and pseudogenes with an X chromosome homologue which are ubiquitously expressed. It is striking that no ubiquitously expressed gene has been found in the other two types of sequences. Only one gene that is expressed predominantly in testes belongs in this category, the sex determining gene *SRY*.

- Ampliconic sequences (named like that due to the large number of repeats and palindromes) exhibit a high degree of similarity (99.9%). They include the largest number of both coding and non-coding genes, although their expression is, most of the time, limited to testes. They are characterised by eight massive palindromes covering 25% of the euchromatic sequences. Also present on ampliconic sequences are inverted repeats and long tandem arrays.

These regions are usually large allowing for gene conversion, a phenomenon of non-reciprocal recombination between Y chromosome sequences. This has resulted in the large number of repeats on the Y chromosome. It is argued that gene conversion is as frequent in MSY as recombination is in autosomes with multiple events taking place per generation.

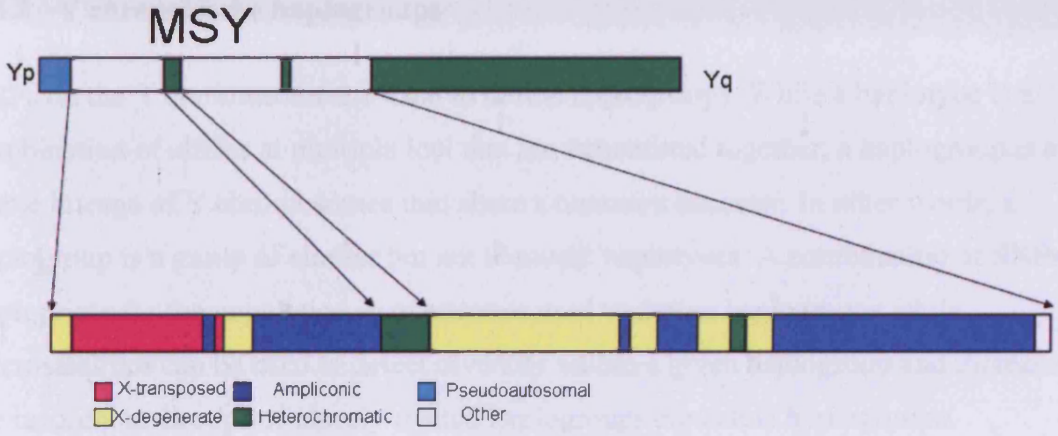

Figure 4-1. The male-specific region of Y chromosome. The diagram shows the different types of euchromatic sequences regions in the MSY. Heterochromatic sequences are also shown.

4.3 Markers and haplogroups to study the Y chromosome

The unique property of the absence of recombination in most of the Y chromosome means that traditional approaches, such as genetic mapping, are not useful. Thus, the availability of the complete euchromatic Y chromosome sequence is having an impact on the way it is studied.

4.3.1 Y chromosome markers

Studies of the variation on the Y chromosome have revealed a large number of polymorphisms which can be used as molecular markers of Y chromosome diversity. The lack of recombination, which means that polymorphisms remain unchanged apart from mutations for a large number of generations, provides the Y chromosome with

the best haplotypic resolution in the human genome. There are two categories of molecular markers for the Y chromosome:

- Biallelic markers, which are mainly SNPs and an Alu element insertion, with low mutation rate
- Multiallelic markers, which are microsatellites and minisatellites, with higher mutation rates

4.3.2 Y chromosome haplogroups

SNPs on the Y chromosome are used to define haplogroups. While a haplotype is a combination of alleles at multiple loci that are transmitted together, a haplogroup is a stable lineage of Y chromosomes that share a common ancestor. In other words, a haplogroup is a group of similar but not identical haplotypes. A combination of SNPs appropriate for the population in question is used to define haplogroups while microsatellites can be used to detect diversity within a given haplogroup and increase the resolution. Groups of closely related haplogroups are called hypergroups.

The number of SNPs on the human Y chromosome was small until Underhill et al. (1997) used DHPLC (Denaturing High-Performance Liquid Chromatography) to discover 19 novel SNPs. Employment of the same method has led to the discovery of 205 SNPs that can be typed using PCR methods (Shen et al. 2000; Underhill et al. 2000). Nowadays, genetic databases, like UCSC (WJ et al. 2002, http://genome.ucsc.edu/) and Ensembl (Hubbard et al. 2009, http://www.ensembl.org/index.html), report a large number of Y chromosome polymorphisms but unfortunately most of them are not verified and they have to be considered with caution. Moreover, the nomenclature of Y chromosome SNPs has been inconsistent with researchers following up to seven different nomenclature systems making any comparison between studies very difficult (The Y Chromosome Consortium 2002).

4.3.3 The Y Chromosome Consortium (YCC)

These issues were addressed by the Y Chromosome Consortium (YCC), a collaborative effort to study variation in the human Y chromosome. This consortium was initiated in 1991 by Michael Hammer and Nathan Ellis with the following goals:

- 'to establish a repository of lymphoblastoid cell lines derived from a sample of males representing worldwide populations'
- 'to provide DNA isolated from these cell lines to investigators searching for polymorphisms on the non-recombining region of the Y chromosome'
- 'to establish a common database containing the results of typing DNA from these cell lines' (The Y Chromosome Consortium 2002)

The YCC has produced a nomenclature system incorporating all verified Y chromosome markers with the flexibility to include newly discovered ones. In addition, by genotyping 245 Y chromosome SNPs in a common, globally representative sample of the male population they were able to construct a single most parsimonious phylogenetic tree including 153 Y chromosome haplogroups (Figure 4-2) (The Y Chromosome Consortium 2002). This tree should be considered as a set of nested lineages sharing a common ancestor, which is the derived state of one or more SNPs. The state of each SNP (derived or ancestral) is determined by sequencing homologous regions of closely related non-human primates. The root of the tree is indicated with an arrow and each major haplogroup is named with a capital letter (A-R). Haplogroup Y is the one that all men belong to, since it includes all known haplogroups (Figure 4-2). Although the assignment of the letters to haplogroups can be considered arbitrary, haplogroups on the lower part of the tree are thought to have appeared more recently in evolution compared to haplogroups on the upper branches of the tree. This means that haplogroups that are more close to each other alphabetically are also more closely related in terms of their evolutionary origin.

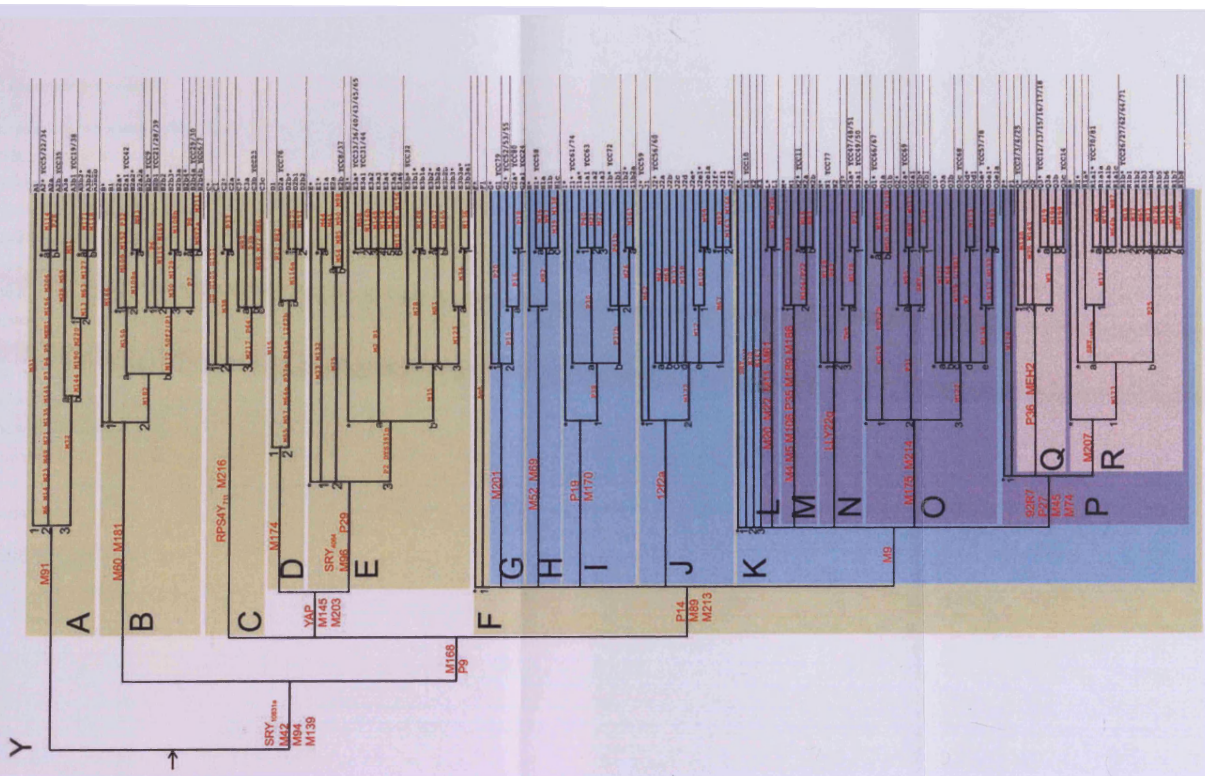

Figure 4-2. The single most parsimonious Y chromosome tree including 245 markers and 153 haplogroups. The arrow indicates the root of the tree. Capital letters are used to name major clades. The most inclusive is Y, where all haplogroups belong. Different colours are used for the different clades. Haplogroup names and YCC sample numbers are given on the right. Markers needed in order to capture a haplogroup are given along the branches. The length of each branch is arbitrary. The * symbol indicates a paragroup (see text for explanation) (The Y Chromosome Consortium 2002)

4.3.4 YCC nomenclature systems

The YCC has proposed two nomenclature systems for the MSY haplogroups. An example will be used to illustrate the way the two nomenclature systems work. Let us consider haplogroup H1b according to the first nomenclature system. The capital letter H is the name of the major haplogroup. The numbers and letters that follow H are used to specify the subclades within this major haplogroup. H1b will refer to subclade H1 and specifically the 'b' part of subclade H1 (Figure 4-3). Nested subclades within a major clade or haplogroup are named using this alphanumerical alternating system until reaching 'the tip of the branch'. Notice that in some subclades the name includes an asterisk, for example H1* (Figure 4-3). This is called a paragroup and indicates that this Y chromosome belongs to haplogroup H1 but not in any of the known subclades of H1. The most likely explanation is usually that there are more subclades within this haplogroup but the SNPs that define them have not been discovered yet.

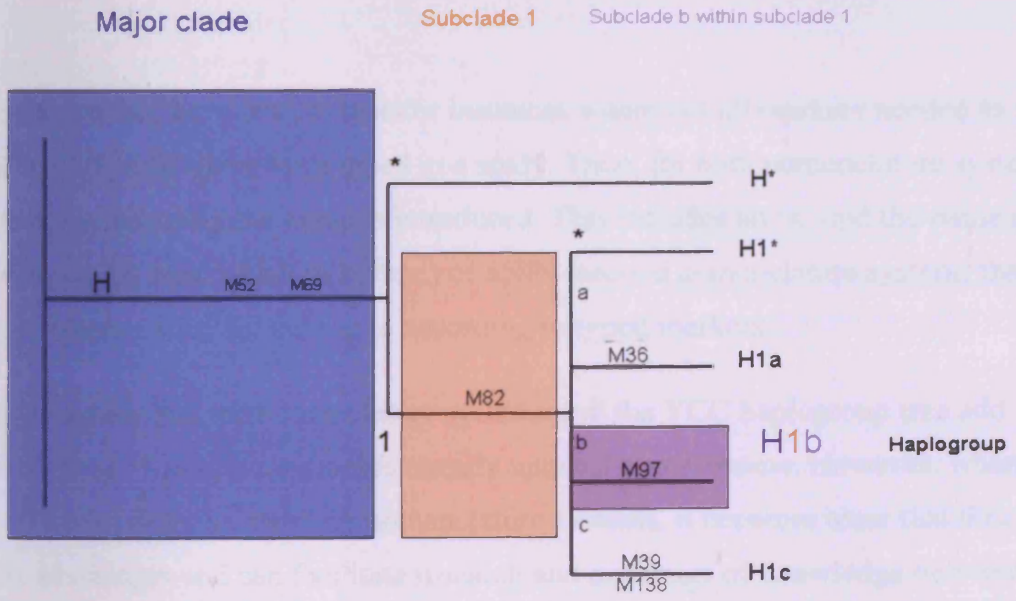

Figure 4-3. Haplogroup H1b named according to 1st nomenclature system. Different colours are used to indicate the levels of clades within the haplogroup

The second nomenclature system again uses the same capital letters for naming major haplogroups. However, instead of naming the subclades alphanumerically it uses the name of the terminal SNP that defines the subclade. Thus, the example haplogroup H1b will be named H-M39 (Figure 4-4) (note that the use of the dash indicates that the name is produced using the second nomenclature system). The same rules apply for paragroups as well.

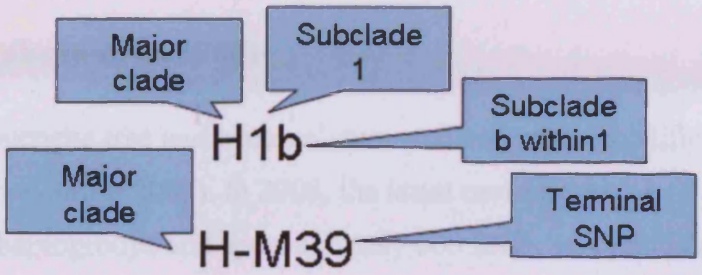

Figure 4-4. The same haplogroup named with the two different nomenclature systems. The figure shows what each letter and number represents

Another rule has been put in place for instances where not all markers needed to reach 'the tip of the tree' have been typed in a study. Then, for both nomenclature systems a parenthesis following the name is introduced. This includes an 'x' and the name of subclades (first nomenclature system) or SNPs (second nomenclature system) that the Y chromosome does not belong to according to typed markers.

It might appear that the nomenclature systems and the YCC haplogroup tree add another layer of complexity to this already unusual chromosome. However, when they are compared with the previous nomenclature systems, it becomes clear that they have major advantages and can facilitate research and exchange of knowledge between researchers. First of all, haplogroups are named systematically and there is no need to change their names even if new mutations are discovered, although this may lead to clades being split or joined. The hierarchical nature of the nomenclature system provides inherent information about the evolutionary relationship between different haplogroups and clades within the same haplogroup. Paragroups are clearly indicated, so that we know that no safe conclusions can be drawn about their evolutionary

relationship with other clades within the major haplogroup they belong to. At present the YCC nomenclature committee oversees the expansion of the haplogroup tree, when new SNPs are discovered, and is responsible for naming the new haplogroups and clades (http://ycc.biosci.arizona.edu.). The same website provides a table which can be used as a cross-reference to previous nomenclature systems, although direct comparisons are not possible in some cases due to the inconsistencies displayed by the nomenclature systems.

4.3.5 YCC haplogroup tree 2008

The YCC phylogenetic tree and nomenclature received minor modifications in 2003 (Jobling and Tyler-Smith 2003). In 2008, the latest revised version of the tree containing 311 haplogroups and approximately 600 SNPs was published (Karafet et al. 2008). The SNPs genotyped came mainly from Michael Hammer and Peter Underhill's group. SNPs in public databases were also utilised. The SNPs were genotyped and further confirmed by sequencing in 33 males from three different populations. The most important changes in the haplogroup tree are the following (Karafet et al. 2008):

- P143 SNP unites haplogroups C and FT
- Haplogroups I and J are joined to the IJ clade
- Haplogroups N and O are joined to the NO clade
- P256 merges haplogroup M with 2 K haplogroups into M hypergroup
- 2 K clades are named as haplogroups S and T

Since the publication of the YCC 2008 haplogroup tree there have been yet more Y chromosome SNPs discovered (Repping et al. 2006). As the authors of the YCC 2008 haplogroup tree point out, the quality and reliability of Y chromosome studies would benefit from a set of publicly available positive controls. Since the YCC cell lines typed in 2002 do not exist any longer, it would be worth typing the known Y chromosome SNPs in CEPH individuals (Karafet et al. 2008).

4.4 Geographical differentiation of Y chromosome haplogroups

The geographical differentiation of Y chromosome haplogroups is higher than for any other part of the genome. It can differ even between populations of the same country. The factors contributing to this differentiation are:

- Lack of recombination
- Small effective population size
- Social factors
- Selection

4.4.1 Lack of recombination and its importance

As mentioned before, the Y chromosome consists of the MSY region and two pseudoautosomal regions (PAR1 and PAR2). The PAR regions recombine with the relevant regions on X chromosome during meiosis. However, the rate of recombination in these regions in men is much higher than in women and they exhibit the highest recombination frequencies of the genome (Flaquer et al. 2008). Consequently, the PAR regions have very low levels of linkage disequilibrium.

The opposite is true for the MSY. This region does not recombine and it is passed unchanged from father to son apart from mutations. The sequence of the MSY can only change by mutations, thus providing a stable record of the Y chromosomes of the male ancestors of an individual.

There are rare occasions when the MSY can recombine. These include individuals that carry two Y chromosomes. Nevertheless, they are either infertile (Chevret et al. 1997) or they have eliminated one Y chromosome in their germline (Solari and Rey Valzacchi 1997), so the product of this recombination never makes it to the next generation. Another instance when recombination in the Y chromosome can occur is between the normal Y chromosome and segments of the Y chromosome carried on another chromosome as asymptotic translocations or between the Y chromosome and highly homologous X chromosome sequences (Cooke and Noel 1979). However,

these events are extremely rare and have yet to be found in humans (Pecon-Slattery et al. 2000). Although recombination does not exist in the MSY, there is evidence of frequent gene conversion or intrallelic recombination. Gene conversion is the process of nonreciprocal transfer of sequence information between a pair of nonallelic DNA sequences. It has been calculated that on average 600 nucleotides in the MSY per generation undergo gene conversion (Rozen et al. 2003).

4.4.2 Effective population size

A factor that influences the geographical distribution of the Y chromosome is its effective population size. Only one copy of the Y chromosome is carried by men. This results in the effective population size of Y chromosomes being one quarter of all the other autosomes (carried in two copies by both sexes) and one third of the X chromosome (carried in two copies by women and in one copy by men). The significantly smaller effective population size means that mutations have fewer 'chances' to change the sequence of the Y chromosome. This is observed as lower sequence diversity on the Y chromosome than anywhere else in the genome (Thomson et al. 2000).

Furthermore, the smaller effective population size renders the Y chromosome more susceptible to genetic drift. This phenomenon is the accumulation of random changes in the abundance of genetic variants. Since the number of Y chromosomes is smaller, random changes can result in significant differentiation between Y chromosomes of different populations.

4.4.3 Social factors

Social factors also influence the geographical distribution of the Y chromosome. Men tend to stay closer to their birthplace compared to women who usually follow them after marriage. Patrilocality results in further increases of the differentiation between Y chromosomes of different populations. It is striking that the opposite is true for societies that practice matrilocality, like Thailand. In these cases, mitochondrial DNA is the one that displays significant differentiation between populations (Oota et al. 2001).

4.4.4 Selection

In the absence of recombination, an advantageous mutation would affect the entire chromosome and increase the frequency of a specific haplogroup more rapidly than expected. The opposite would be true for non-advantageous mutations.

All these factors lead to a high degree of geographical differentiation, more than anywhere else in the human genome. The consequences are very interesting. First of all, the study of Y chromosome haplogroups is a powerful tool for population geneticists seeking to detect population migration patterns that shaped the human race. However, there are serious threats to studying the Y chromosome posed by this differentiation. Because of the low frequency of some Y chromosome variants, it is almost certain that there are polymorphisms that have not been discovered yet. If a set of Y chromosome variants that have been discovered in a certain population is used in a genetically distinct population, the erroneous conclusion could be drawn that the variation in the second population is much less than in the first one. Of course, this is not true. It is just a result of the fact that other markers are more suitable for characterising this population. Unfortunately, the geographical differentiation of the Y chromosome is so high that even populations within the same country can be very different in terms of their Y chromosomes. Researchers trying to associate Y chromosome variation with particular phenotypes should also be cautious about the geographical differentiation of the Y chromosome. It is vital that a well-matched control with the same geographical origin is used to avoid associations that are merely the result of population stratification. Examples of such studies will be discussed later.

4.4.5 Geographical distribution of Y chromosome haplogroups in Europe and United Kingdom

The Y chromosome distribution in Europe has not been studied extensively. Early studies were limited by the availability of markers and difficulty in comparing between the studies due to the problematic nomenclature systems.

Y chromosome markers are the ideal tool for deciphering the population history of Europe. The largest study of European Y chromosomes has involved genotyping 11 Y chromosome SNPs in 47 European populations. These define 10 haplogroups. Spatial autocorrelation analysis showed that the pattern of geographical distribution in Europe is highly non-random and that two haplogroups represent 45% of all the Y chromosomes in the European continent (Rosser et al. 2000). Although this study was published before the YCC haplogroup tree and nomenclature, these two haplogroups roughly correspond to paragroup P*, which includes a part of haplogroup R, and haplogroup J. The presence of two haplogroups in such high numbers in Europe provides evidence for population migration events. The same study has shown a significant correlation between Y chromosome diversity and geography but not language. However, the effects of genetic drift and natural selection acting on the Y chromosome cannot be excluded (Rosser et al. 2000). This study included Scottish, Cornish and East Anglian populations and showed that haplogroups P*, R1a and BR* made up the majority of the Y chromosomes in the UK. However, the number of Y chromosome markers used was not enough to provide a clear picture of the Y chromosome haplogroup distribution in the United Kingdom (Rosser et al. 2000).

A study focusing only on the British Isles used a combination of Y chromosome SNPs and microsatellites covering approximately the same haplogroups as before (Wilson et al. 2001). Populations from Wales, Ireland and Orkney were chosen and compared. They were also compared with a Scandinavian population, representing the Viking contribution, a Basque population, which is considered to be descended from the Palaeolithic inhabitants of Europe and an Asian population from the Near East. This study found that the Irish and the Welsh populations were not significantly differentiated. The Scottish populations showed an intermediate pattern between Celtic and Scandinavian populations. In addition, the similarity between Celtic and

Basque populations was striking. According to the authors, this supports the notion of a Palaeolithic component in the British Isles and it indicates that there has been no effect on the paternal genetic component during the Neolithic and Iron Age transitions. On the other hand, the Scandinavian invasion has left its signature on the Y chromosomes of the inhabitants of the British Isles (Wilson et al. 2001).

Evidence for substantial Anglo-Saxon male migration into Central England but not North Wales was provided by a study that genotyped Y chromosome SNPs and microsatellites in samples collected along an east-west transect from East Anglia to North Wales (Weale et al. 2002).

A very detailed approach was followed by Capelli et al. (2003) with typing of 11 Y chromosome SNPs and six microsatellites in 1,772 men from 25 small urban locations of the British Isles. The most frequent haplogroups in their group were I(xI1b2), R(xR1a1) and R1a1. These authors argued that the Anglo-Saxon contribution to the male genetic heritage of the British Isles is small and more obvious in the Central-Eastern part of England. They also showed evidence of heterogeneity in all populations (Capelli et al. 2003).

Although some of these studies have been very detailed and their contribution was valuable in terms of population genetics, they have failed to provide a consensus of Y chromosome markers and the haplogroups they define for use in medical genetic studies. The key issue in medical genetic studies is that sample collection depends highly on availability of affected individuals. For this reason, the Y chromosome markers need to be representative of the British population as whole. A high degree of haplogroup breakdown would be undesirable because it would be useful for identifying distinct populations within the United Kingdom but not for comparing individuals that do not necessarily come from the same region of the country.

4.5 Y chromosome associations with phenotypes

The fundamental property of absence of recombination in the MSY is important for studies seeking Y-linked phenotypes. The MSY essentially behaves as a chromosomal block in complete linkage disequilibrium. With that in mind, any functional

polymorphism predisposing to a disease or affecting any phenotype would be correlated with the SNPs that define the haplogroups. In order to perform a Y chromosome association study, it is important to choose a well-matched control sample with the same geographical origin as the group of patients. The next thing needed is a set of SNPs defining the most frequent haplogroups in the population in question. By comparing the frequencies of haplogroups in the group of patients and healthy controls, it can become apparent whether individuals belonging to certain Y chromosome haplogroups are more susceptible than others to develop a disease or have a specific phenotypic manifestation. There can be two possible explanations for this association. One of them is that the functional polymorphism responsible for this association has appeared before the SNPs that define the associated haplogroup. Thus, all Y chromosomes belonging to these haplogroup would carry the causal variant. However, it is most likely that the functional polymorphism has appeared against a specific haplogroup background and for this reason, only certain individuals within this haplogroup will have it (Krausz et al. 2004).

Unfortunately, some of the associations with Y chromosome haplogroups that have been reported are down to differences in geography or language. It should also be noted that due to high geographical differentiation of Y chromosome haplogroups the absence or presence of association in one population cannot be generalised in other populations without appropriate replication. A summary of the association studies performed on the Y chromosome so far can be seen at Table 4-1.

Table 4-1. Y chromosome association studies

Phenotype	Population	No of Y chr SNPs	Use of other Y chr markers	Result	Reference
XX maleness with infertility	European	2	1 microsatellite	Association with haplogroup Y^{*}(xP)	(Jobling et al. 1998)
Reduced sperm count\azoospermia	Japanese	3	-	Association with haplogroup D	(Kuroki et al. 1999)
Reduced sperm count\azoospermia	Japanese	10	5 microsatellites	No association	(Carvalho et al. 2003)
Infertility	Italian	8	-	No association	(Previdere et al. 1999)
Reduced sperm count	Danish	3	-	Association with haplogroup K(xP)	(Krausz et al. 2001)
Testicular cancer	English	7	6	No association	(Quintana-Murci et al. 2003)
Prostate cancer	Japanese	0	1 microsatellite	Association with one microsatellite allele	(Ewis et al. 2002)
Prostate cancer	4 ethnic groups	118	-	Association with a haplogroup found only in Japanese populations	(Paracchini et al. 2003)
Hypertension	Australian	-	HindIII	Association with HindIII(-)	(Ellis et al. 2000)
Hypertension	Polish Scottish	-	HindIII	Association with HindIII(+)	(Charchar et al. 2002)
Hypertension	European	5	2 microsatellites	No association	(Rodriguez et al. 2005)
Autism	Norwegian Swedish French	10	-	No association	(Jamain et al. 2002)
Alcohol dependence	Finnish	1	7	Association with 3 Y chr clades	(Kittles et al. 1999)
Aggressive behaviour	Pakistani	49	-	Association of haplogroups R2 and R1a1 with increased aggression	(Shah et al. 2009)

4.5.1 Reproductive medicine

The Y chromosome has been an obvious candidate for reproductive phenotypes in males, since it harbours the sex-determining gene, SRY, and genes affecting spermatogenesis.

Sex-reversed XX infertile males can appear because of ectopic recombination between genes *PRKX* (Protein Kinase X-linked) and *PRKY* (Protein Kinase Y-linked) resulting in the translocation of extra Y-chromosomal material, including *SRY*, to the X chromosome. Men with this translocation lack the Y chromosome genes responsible for the formation of active sperm and they are infertile. By studying the translocated part of the Y chromosome and paternal Y chromosomes, a frequency of 90% of haplogroup Y*(xP) has been found in individuals with this condition, while it only represents one third of the population they come from. This should be considered as a true association because cases and controls were well-matched and also because there is a plausible mechanism explaining this association. The control population not belonging to the risk haplogroup carries an inversion protecting them against this particular translocation (Jobling et al. 1998)

A Japanese population has been used in order to test for association of Y chromosome haplogroups with reduced sperm counts and azoospermia. A haplogroup specific to East Asian populations has been found to be associated with reduced sperm count and increased risk of azoospermia (Kuroki et al. 1999). However, another study of men of Japanese origin failed to replicate the association (Carvalho et al. 2003). It is important to note that the associated haplogroup varies geographically even within Japan adding another layer of complexity.

A study of Italian men with idiopathic infertility proves just how important it is to take into account the geographical differentiation of the Y chromosome. While an association was found between a certain Y chromosome haplogroup and idiopathic infertility, the control sample was later subdivided to better reflect the cases who all came from Central Italy and the association disappeared (Previdere et al. 1999).

Reduced sperm count was also associated with a Y chromosome haplogroup in a group of Danish men. The rarity of this haplogroup in the normal Danish population and the reduced reproductive fitness poses the possibility of negative selection acting

on this haplogroup. The authors hypothesise that structural rearrangements in this haplogroup affect the normal expression of a gene involved in spermatogenesis (Krausz et al. 2001).

4.5.2 Male cancers

Deletions and structural changes on the Y chromosome have been associated with various types of cancers suggesting that the Y chromosome harbours tumour suppressor genes or oncogenes (Center et al. 1993; Hunter et al. 1993; Sauter et al. 1995). It has also been suggested that poor spermatogenic function is present in patients with testicular cancer a long time before diagnosis suggesting that there is a general testicular dysfunction. Thus, it is likely that a Y chromosome variant is involved (Boisen et al. 2001).

The role of Y chromosome haplogroups in testicular cancer has only been investigated in an English population without producing significant results. However, the geographical differentiation of the Y chromosome and the multifactorial nature of cancer do not exclude its involvement in another population (Quintana-Murci et al. 2003).

Prostate cancer was investigated in a Japanese population using one microsatellite marker and a significant association was reported with one of the alleles (Ewis et al. 2002). Although the results are not very convincing, since they are based on one microsatellite with high mutation rate, they are supported by a multi-ethnic cohort of four populations. An increased frequency of patients suffering from prostate cancer in one haplogroup was found only in the Japanese population (Paracchini et al. 2003).

4.5.3 Hypertension and lipid levels

The Y chromosome has been studied in relation to phenotypes showing higher incidence in males, since it harbours a chromosomal region, MSY, unique to men.

Hypertension is such a phenotype with males having significantly increased blood pressure compared with pre-menopausal women. Consequently, males are at increased risk of cardiovascular disease.

The role of the Y chromosome in increased blood pressure is supported by both animal and human studies. The male Spontaneously Hypertensive Rat (SHR) and the male Stroke-Prone Spontaneously Hypertensive rat (SHRSP) show increased blood pressure compared to females (Ely et al. 1997). The Y chromosome of these strains has been found to be responsible for the elevated blood pressure after the breeding of strains of rats with a hypertensive Y chromosome on a normotensive background. However, blood pressure was not as high as for the SHR strain indicating that the Y chromosome and autosomal genes interact to produce this phenotype (Negrin et al. 2001). It is intriguing that the SHR rat is also one of the best animal models of ADHD (Russell 2007).

In humans, paternal hypertension has been found to increase blood pressure in male offspring (Uehara et al. 1998). The first study of the Y chromosome and hypertension was performed in a sample of Australian men and used the HindIII centromeric alphoid satellite polymorphism. This polymorphism is present in two allelic forms HindIII(+) and HindIII(-). An association of HindIII(-) and increased blood pressure was found (Ellis et al. 2000). The same polymorphism was genotyped in two more samples, one of Polish and one of Scottish ancestry but the opposite allele HindIII(+) was associated with increased blood pressure (Charchar et al. 2002). This presence of association in three different populations points to an involvement of the Y chromosome in regulation of blood pressure, although not necessarily of this polymorphism. The use of haplogroup analysis could be more fruitful in this case. This approach has been followed in the most comprehensive study of the Y chromosome and blood pressure. By genotyping five SNPs and two microsatellites in 2,743 Caucasian men, Rodriguez et al. (2005) failed to replicate the association of Y chromosome haplogroups and elevated blood pressure.

Similar to animal studies, there is evidence from humans that the Y chromosome interacts with autosomal loci. An epistatic interaction between the HindIII polymorphism and the aldosterone synthase gene (*CYP11B2*) to increase blood pressure was demonstrated in the Scottish sample, although it was not replicated in a study of three white populations (Charchar et al. 2002).

Another interesting phenotype of the SHR rat is the high levels of triglycerides and high-density lipoprotein (HDL) compared with normotensive rats (Kren et al. 2001). In humans, lipid levels exhibit ethnic differences (Cappuccio 1997), which cannot totally be attributed to diet, and they indicate a possible involvement of certain Y chromosome haplogroups. The HindIII(-) allele has been implicated in increased lipid levels in a Polish sample and the association was independent of testosterone levels and alcohol consumption (Charchar et al. 2004). It is also interesting that in men of African origin a Y chromosome haplogroup is linked to a favourable lipoprotein pattern, thus reducing their susceptibility to cardiovascular disease (Russo et al. 2008).

4.5.4 Psychiatric phenotypes

Psychiatric disorders, such as autism, ADHD, schizophrenia and alcohol dependence as well as behavioural traits, such as aggression, have higher incidence in males.

For this reason, the role of Y chromosome haplogroups has been investigated in some of these phenotypes.

An association study genotyping 10 Y chromosome SNPs in individuals with autism did not result in a positive association. However, it should be pointed out that the sample used was not homogeneous (Jamain et al. 2002). In contrast, Kittles et al. (1999) used only one Y chromosome SNP and seven microsatellite markers to genotype a sample of Finnish men with alcohol dependence. They found that three Y chromosome clades were more frequent amongst the individuals with alcohol dependence than in controls. They also tested for an association with antisocial personality disorder comorbid with alcohol dependence but they did not find any difference between cases and controls. Nevertheless, their use of microsatellites, which have a high mutation and recurrence rate, is not ideal and it is striking that their

results indicate three different Y chromosome mutations predisposing to alcohol dependence in Finland (Kittles et al. 1999).

Finally, the role of the Y chromosome in aggressive behaviour has been studied in a sample of Pakistani men. There was no difference in the mean scores of aggression across the five different haplogroups captured in the study but effect-size comparisons showed an association of two haplogroups with lower scores of aggression (Shah et al. 2009).

4.6 Y chromosome and brain function

It is well established that sex differences exist in cognition and behaviour in both humans and animals. What we do not know yet is how much of this differentiation is a result of differential hormonal environment and how much is a direct effect of sex chromosome genes. The Y chromosome, specifically, contains genes that are only expressed by males making it a suspect for contribution to sex differences.

4.6.1 Animal studies

In mammals it is common for male embryos to develop faster than female embryos before the differentiation of the gonads (Erickson 1997). A microarray analysis of foetal mouse brain has shown that more than fifty genes were differentially expressed in foetal brains of male and female mice. It is remarkable that this analysis was performed at 10.5 days post coitum (dpc) before gonadal hormones have started to be produced (Dewing et al. 2003).

Sex chromosome genes are expressed in the brain. A large number of X chromosome genes have been implicated in mental retardation (Skuse 2005). At least ten Y chromosome genes are also expressed in the brain (Xu et al. 2002) with genes like *SRY*, *ZFY* and protocadherin expressed in brain of both human and mouse.

Several studies in mice have implicated the Y chromosome in aggression (Guillot et al. 1995; Maxson 1996; Maxson et al. 1979), the distribution of hippocampal mossy fibres (Hensbroek et al. 1995), levels of dopamine (Sluyter et al. 1995) and serotonin

in the brain (Tordjman et al. 1995) but with uncertainty about the possible involvement of androgens.

A method that can be used in order to test how much difference can be attributed to sex chromosomes is to use animals of the same sex but with different genetic background. Male mice differing only in their Y chromosome have different levels of aggression (Guillot et al. 1995). However, this effect could be mediated by increased levels of testosterone. Comparisons between mice with and without the SHR (Spontaneously Hyperactive Rat) Y chromosome against the same genetic background showed that the SHR Y chromosome was associated with increased aggression due to increased levels of testosterone and decreased levels of serotonin in the brains of mice with the SHR Y chromosome (Toot et al. 2004).

4.6.2 Four core genotypes model

An interesting animal model has been used to dissociate the effect of the Y chromosome from hormonal effects. The four core genotypes mice model utilises the *Sry* gene and its unique property of initialising male sex differentiation. The *SRY* is deleted from the Y chromosome (Y^-) resulting in XY^- female mice. When an *Sry* transgene is inserted in an autosome, mice develop to males irrespective of their sex chromosome complement. With appropriate crossings it is possible to have four genotypes: XX females, XY^- females, XX*Sry* males and XY^-*Sry* males (De Vries et al. 2002). By comparing XX and XY female mice with XY and XX male mice it became possible to evaluate the effect of sex chromosomes independent of the hormonal milieu. A large number of sexually dimorphic neural and behavioural phenotypes have been assessed to date. Some of the phenotypes correlated with the presence of testes or ovaries indicating a strong hormonal effect. However, sex chromosome complement had an effect on aggression and parental behaviour that was not related to the hormonal milieu (Gatewood et al. 2006). In addition, both male and female mice with a Y chromosome were more masculine than those without it in terms of their density of vasopressin-immunoreactive fibres in the lateral septum (De Vries et al. 2002).

4.6.3 Y chromosome abnormalities in humans

Case or small-scale studies have suggested a role of Y chromosome genes in some neuropsychiatric disorders. For example, individuals with 47,XYY syndrome appear to be at increased risk of developing antisocial behaviour (Gotz et al. 1999), as well as being at risk of developing bipolar disorder and schizophrenia (Mors et al. 2001), which could be due to MSY gene over-expression. A potential role of Y-linked genes in the pathogenesis of schizophrenia is also supported by the observation of an isodicentric Y chromosome in a patient with schizophrenia (Yoshitsugu et al. 2003). A case study has implicated Y-linked genes in ADHD susceptibility; in this case, the affected boy possessed a rare deletion of Yq with duplication of Yp (Mulligan et al. 2008). Interestingly, the duplicated region included the *SRY* gene, suggesting that overdosage of this specific gene may be responsible (either directly or indirectly) for the observed behavioural phenotype. One important caveat with studies such as these is that the co-occurrence of a particular cytogenetic mutation and a particular neuropsychiatric manifestation does not necessarily imply that the two are linked but they can constitute an important starting point for studying certain regions in detail.

4.7 *SRY* (Sex-determining region of chromosome Y)

SRY is probably one of the most interesting Y chromosome genes as it is the gene responsible for male sex differentiation. It was first discovered in 1990 by Sinclair in XX males by analysis of fragments of the Y chromosome that had translocated in the X chromosome of these individuals. It was then proved that *SRY* is the TDF (Testis Determining Factor) gene on the Y chromosome that researchers were looking for (Sinclair et al. 1990). Its presence or absence determines the sex of an individual irrespective of Y chromosome complement.

The human *SRY* is a single-copy gene in the MSY region of the Y chromosome and belongs to the X-degenerate class of sequences (Skaletsky et al. 2003). In humans it is an intronless gene which spans 897 bp. It codes for a protein of 204 aminoacids and molecular mass of 24 kD. The protein can be divided into three domains. The HMG (High-Mobility Group) domain is in the middle of the gene and is conserved in rodents and sheep among other species. It is responsible for binding and bending

DNA. The other two domains, the C-terminal and N-terminal domains are not conserved. The mouse SRY protein has a long glutamine-rich domain, which is absent from the human SRY (Harley et al. 2003).

In mice and humans, *Sry* is a single copy gene; in contrast, in rats and certain other rodent species multiple copies of the *Sry* gene have been reported on the Y chromosome (Turner et al. 2007).

4.7.1 *SRY* transcription regulation

The promoter of *SRY* contains a number of regulatory elements, although their position is not conserved across species. In addition, multiple transcription initiation sites have been detected in human and mice (Clepet et al. 1993). Transcription factor binding sites have been found within the promoter of the *SRY* gene. Among them there is a binding site for SF1 (Steroidogenic Factor 1), a protein that could be responsible for the activation of *SRY*. SF1 is a nuclear hormone receptor. When knocked down in mice, it causes the arrest of the developing gonad (Luo et al. 1995). Furthermore, an *SF1* mutation has been found in a sex-reversed individual (Luo et al. 1995). However, the most likely scenario is that *SRY* transcription regulation requires coordinated action of transcription factors like *SF1*, *SP1* (Specificity Protein 1) and *WT1* (Wilms Tumour 1). Other genes that are potential candidates for activating *SRY* are *IR* (Insulin Receptor), *IRR* (Insulin-Related Receptor), *IGF1R* (Insulin-like Growth Factor 1 Receptor) and *FOG2* (GATA4-Friend of gata2) (Sekido and Lovell-Badge 2009) (Figure 4-5) . *DAX1* (dosage-sensitive sex-reversal adrenal hyperplasia congenital X chromosome 1) is hypothesised to be an antagonist of *SF1* by recruiting a nuclear corepressor (Harley et al. 2003).

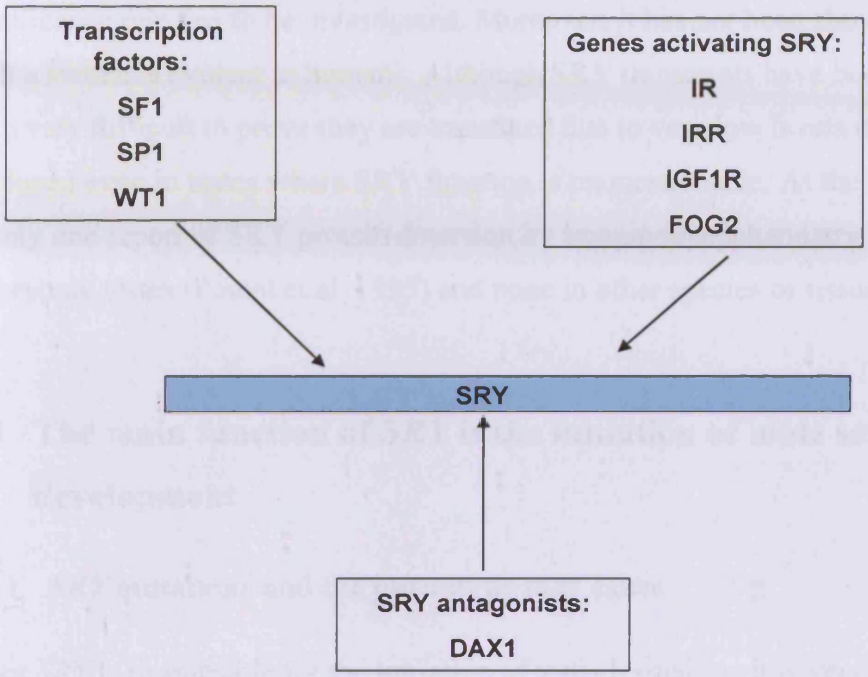

Figure 4-5. Diagram summarising the genes involved in SRY transcription regulation (SF1: Steroidogenic Factor 1; SP1: Specificity Protein 1; WT1: Wilms Tumour 1; IR: Insulin Receptor; IRR: Insulin-Related Receptor; IGF1R: Insulin-like Growth Factor 1 Receptor; FOG2: GATA4-Friend of gata2; DAX1: Dosage-sensitive sex-reversal adrenal hyperplasia congenital X chromosome 1)

4.7.2 SRY expression

Although *SRY* is chiefly expressed in the testes, it is also expressed to some extent in other tissues including heart, liver and kidney, and certain brain regions. For this reason, besides influencing male-specific traits indirectly, it is possible that it influences neurodevelopment and brain function in a direct manner. In humans, *SRY* expression has been described in the hypothalamus, frontal and temporal cortex (Mayer et al. 1998). In adult rodents, the gene is expressed in the hypothalamus and midbrain (notably the substantia nigra and the ventral tegmental area) (Dewing et al.

2006; Lahr et al. 1995). However, the expression of SRY in brain is developmentally regulated. Circular and probably untranslatable SRY mRNA was found in mouse brain prenatally. This changed to linear and likely translatable transcripts postnatally (Mayer et al. 2000). The transition between expressed forms is presumably due to a developmental switch in promoter usage, but the evolutionary and functional significance remains to be investigated. Moreover, it has not been shown whether such a switch is evident in humans. Although SRY transcripts have been found, it has been very difficult to prove they are translated due to very low levels of SRY protein produced even in testes where SRY function is unquestionable. At the moment, there is only one report of SRY protein detection by immunohistochemistry in human embryonic testes (Poulat et al. 1995) and none in other species or tissue.

4.8 The main function of *SRY* is the initiation of male sex development

4.8.1 *SRY* mutations and the phenotype they cause

Since *SRY* is responsible for the initiation of testis formation, it is very difficult to disentangle its direct effects from the effects of testosterone. In humans, research is restricted to studying individuals with *SRY* mutations. Most *SRY* mutations have been found within the HMG box and they cause sex reversal in XY individuals. They usually present as normal females with female internal and external genitalia but they suffer from complete gonadal dysgenesis and ovarian dysfunction (Hawkins et al. 1992). Since gonadal dysgenesis is associated with high risk of gonadal tumours, their ovaries are usually preventatively removed. Their psychosexual behaviour is normal and some of them have managed to become mothers with the use of IVF treatment (McCarty et al. 2006). Most of the *SRY* mutations are *de novo* and they tend to be missense mutations clustering in the HMG box highlighting its role in the function of the gene. It is intriguing that familial *SRY* mutations have been found in fathers of XY sex-reversed individuals indicating that some *SRY* mutations are not completely dominant (Harley et al. 2003).

Although *SRY* can be found in most cases of XX males, *SRY* mutations are only responsible for 15% of cases of XY females (Harley et al. 2003). This means that the

presence of *SRY* commits sexual differentiation to the male pathway but it is not enough to complete the process. Sexual differentiation requires a coordinated network of genes, where their interactions and timing is essential. Unfortunately, we are still a long way of fully understanding this fundamental process.

4.8.2 *SRY* and male sex differentiation

Sexual differentiation can be seen just after conception in the reproductively bipotent embryo with male embryos having more cells and being more metabolically active than female embryos (Mittwoch 2000). The undifferentiated genital ridge is formed from the surfaces of the mesonephros under the influence of transcription factors, such as *WT1* and *SF1*. Since *SRY* is a transcription factor, it has been proposed that it either represses a repressor of male development or that it activates a gene responsible for male development (Sekido and Lovell-Badge 2009). The finding that *SRY* expression must reach a threshold at an appropriate time to initiate male development is in line with a role in initiation but not maintenance of male development.

There is accumulating evidence that the direct target gene of *SRY* is *SOX9* (SRY-related HMG Box 9). This belongs to a gene family along with *SRY* and more than 20 genes that have a HMG domain. It is located on chromosome 17 with mutations causing campomelic dysplasia, a form of severe bone malformation. All patients with campomelic dysplasia also suffer from sex reversal highlighting the vital role of *SOX9* in sexual development (Harley et al. 2003). *SOX9* expression is upregulated immediately after the expression of *SRY* while it is decreased in the ovary and it has been found that all *SRY*-expressing cells also express *SOX9*. Transgenic XX mice ectopically expressing *SOX9* under the Wtl promoter developed to males showing that the only function of *SRY* during sexual development is the activation of *SOX9* (Vidal et al. 2001).

SRY is downregulated after testis determination. In mouse the genital ridge forms at 10 dpc (days post coitum) and *SRY* is detected in it from 10.5 dpc but by 12.5 dpc its levels are very low. In humans, the genital ridge forms at 33 days gestation and *SRY* is detected from 41 days, peaking at 44 days gestation. Sertoli cells continue to express

SRY (although at lower levels) even after testis formation suggesting a possible role of *SRY* in spermatogenesis.

The target cells for *SOX9* are the pre-Sertoli cells within the Wolffian ducts. *SOX9* upregulation triggers them to develop to Sertoli cells, which in turn produce AMH (Anti-Mullerian hormone). This causes the regression of the Mullerian ducts and migration of the cells that will develop to Leydig cells, the main producers of testosterone. Primordial germ cells migrate into the testicular cords and enter meiotic arrest (Yao et al. 2002). In the absence of *SRY*, *DAX1* inhibits expression of *SOX9* and *SF1*. Although female sexual development has not been studied as extensively as male, *DAX1* is considered an essential gene, since its duplication has been found to result in female development even in the presence of *SRY* (Swain et al. 1998).

4.9 *SRY* involvement in brain function and behaviour

The fact that *SRY* is expressed in the brain has led to attempts to link it with specific brain functions and behaviours. It is very interesting that *SRY* expression has been found in catecholaminergic regions (Milsted et al. 2004). Catecholamines, such as dopamine have been implicated in a number of neuropsychiatric disorders with higher male prevalence, such as ADHD, Parkinson's disease and schizophrenia. They are also part of the sympathetic nervous system, thus affecting cardiovascular functions (cardiovascular disease is also more prevalent in males).

SRY has been found to be able to increase reporter gene activity under the control of the tyrosine hydroxylase (*TH*) promoter (Milsted et al. 2004). TH is the rate-limiting enzyme in dopamine synthesis and it will be discussed in a following chapter. Dewing et al. (2006) found that *Sry* is expressed in the substantia nigra of the adult mouse brain in TH-expressing neurons. Downregulation of Sry using antisense oligonucleotides caused a decrease in TH expression. This led to motor deficits in male rats providing robust evidence for a direct effect of *Sry* on the brain without involvement of gonadal hormones. The authors came up with the interesting hypothesis that *SRY* is the factor responsible for increasing the number of TH-expressing neurons in males. Thus, it is compensating for the lack of oestrogen in males, which in females acts to increase TH-expressing neurons (Dewing et al. 2006).

The presence of *SRY* exclusively in men and its involvement in dopaminergic function regulation suggests that it can be an interesting candidate for increasing male susceptibility to certain neuropsychiatric disorders involving dysregulation of dopaminergic function, such as ADHD, schizophrenia and Parkinson's disease.

4.10 Other Y chromosome genes expressed in brain

As mentioned before, Y chromosome gene expression is not restricted to the gonads. Apart from *Sry* another six genes from the MSY region are expressed in mouse brain (Xu and Disteche 2006). Three of them (*Smcy/Jarid1c*, *Uty* and *Eif2s3y*), although they have an X chromosome homologue, escape X chromosome inactivation indicating that there is a dosage difference between males and females in their expression. It is interesting that *Smcy/Jarid1c* and *Uty* are expressed in human brain as well.

ZFY is another *MSY* gene which is expressed in the hypothalamus, frontal and temporal cortex of adult human brain (Mayer et al. 1998). This gene does not have a mouse counterpart and its function is poorly understood apart from the fact that it might be involved in male sex differentiation.

Protocadherin 11Y *(PCDH11Y)* is another Y chromosome gene which has received a lot of interest regarding its role in neurodevelopment. This gene and its counterpart on the X chromosome are located within a hominid-specific region of the sex chromosomes and encode members of the protocadherin superfamily responsible for cell-cell interactions during development of the central nervous system (Blanco et al, 2000). The fact that the *PCDH11Y* gene is expressed in males in a highly regulated and spatiotemporally dynamic manner (Blanco et al. 2000) and is involved in synapse formation and neuronal path finding in the brain, processes which go awry in a number of common male-biased mental conditions (Blanco et al. 2000), make it an attractive candidate for involvement in the male excess of neuropsychiatric disorders. Finally, as mentioned previously, *PCDH11Y* is absent in non-human primates such as chimpanzees and gorillas (Wilson et al. 2006); both *PCDH11Y* and *PCDH11YX* were created by duplicative translocation that occurred close to chimpanzee–human bifurcation 6 million years ago. Moreover, the gene has shown accelerated sequence

change in the hominid lineage (Wilson et al. 2006). For this reason, it has been proposed, most vociferously by Timothy Crow and colleagues, that aberrant expression of the protein encoded by *PCDH11Y* could predispose males to disorders of human-specific functions (Crow 2002), notably schizophrenia and autism (Hill et al. 2008). However, candidate gene studies examining possible association between SNPs within *PCDH11Y* and a number of neuropsychiatric disorders have failed to find evidence for a link to date (Giouzeli et al. 2004). Alternatively, it may be that it is the expression pattern of the gene (which may be modulated epigenetically via environmental influences) which is more important in modulating its effect on male disease vulnerability than its sequence.

Another candidate gene that could potentially influence neuropsychiatric phenotypes is *NLGN4Y* (neuroligin 4 Y-linked), the Y homologue of *NLGN4X*. Neuroligin genes encode cell adhesion molecules involved in the process of synaptogenesis (Ylisaukko-oja et al. 2005). Early findings that mutations in *NLGN4X* were present in families with mental retardation and autism spectrum disorders suggested a possible causal link between the disorders and the gene mutation (Jamain et al. 2003; Laumonnier et al. 2004). The results of subsequent studies have suggested that, if mutations in *NLGN4X* are pathogenic, they are likely to be rare, and are not likely to explain the majority of cases of autism (e.g. (Gauthier et al. 2005; Talebizadeh et al. 2004; Vincent et al. 2004).

The Y chromosome, the smallest chromosome in the human genome, defines the sex of an individual. With its unique characteristics of absence of recombination and high geographical differentiation poses difficulties when studying. However, the Y Chromosome Consortium efforts have made it easier for Y chromosome to be included in association studies. Since the Y chromosome is only found in men, it could be a candidate for involvement in disorders with a male preponderance.

In summary, psychiatric disorders exhibit sex differences in prevalence, severity, manifestation and age of onset. ADHD is a highly heritable childhood-onset neurodevelopmental disorder with a male prevalence. Schizophrenia is also heritable with males having an earlier age of onset and increased disease severity. Both ADHD and schizophrenia are complex disorders with common variants of small effect likely to account for a large amount of the genetic component of the disorder. Rare variants also appear to be important. According to a study by the International Schizophrenia Consortium it is likely that thousands of common alleles of very small effect contribute to the risk of schizophrenia (The International Schizophrenia Consortium 2009). Although genome-wide association studies in schizophrenia have shown evidence of association with several common variants, a substantial genetic component remains to be discovered. Due to the small effect size of common variants, genome-wide association studies require large samples in order to discover some of the common variants that are scattered across the genome (Wang et al. 2005). In addition, CNV studies have shown that rare variants also increase risk for psychiatric disorders. Although the effect sizes of large, rare variants are considerably larger than common variants, each one is relevant for only a very small percentage of patients (Sebat et al. 2009). Rare CNVs have been shown to be implicated in both schizophrenia (Kirov et al. 2009; The International Schizophrenia Consortium 2008) and ADHD (Elia et al. 2009; Lesch et al.).

However, to date none of the studies has included rare or common variants on the Y chromosome. The Y chromosome is only found in males and it possesses a male-specific region, which does not recombine with the X chromosome. Y chromosome variation in animals has been associated with aggression (Guillot et al. 1995), parental behaviour (Gatewood et al. 2006) and alterations in the levels of dopamine and serotonin in the brain (Tordjman et al. 1995). In humans, common Y chromosome variation is captured using Y chromosome haplogroups. The Y chromosome Consortium has compiled a catalogue of Y chromosome SNPs that have been validated (The Y Chromosome Consortium 2002). Common variants on the Y chromosome have been studied in relation to infertility and male cancers as well as disorders that show a male prevalence, such as hypertension, autism, alcohol dependence and aggression (see section 4.4 for more information). Common Y

chromosome variation has not been studied extensively in psychiatric disorders, although some of these show a clear male excess. In addition, in cases where Y chromosome variation has been studied in psychiatric disorders, the populations studied were not ethnically homogeneous, which is of great importance for Y chromosome variation due to its high geographical distribution. Genome-wide association studies have only recently started to include Y chromosome markers, thus the Y chromosome has escaped being studied in this setting. Although the Y chromosome Consortium simplified studying common variation on the Y chromosome, this is not the same for the study of rare variants on the Y chromosome, which is still in its infancy.

In this project, the role of common Y chromosome variants in ADHD and schizophrenia was investigated after selecting Y chromosome haplogroups appropriate for UK populations and performing two case-control association studies, one for ADHD and one for schizophrenia using cases and controls of UK origin. Since common variants can have subtle effects on the phenotype of the disorders rather than main effects, a possible modifying role of Y chromosome variation on ADHD and schizophrenia was investigated after selecting sexually dimorphic phenotypic measures for each disorder. A possible interaction of Y chromosome with tyrosine hydroxylase was investigated because of evidence in animals that the SRY gene on the Y chromosome interacts with tyrosine hydroxylase, which is the rate-limiting enzyme in dopamine metabolism.

Chapter 5 Hypotheses and aims of the project

The aim of this project was to investigate the role of Y chromosome variation in sex differences in ADHD and schizophrenia and this was achieved by testing the following hypotheses:

1. Testing for a direct effect of Y chromosome variants on ADHD and schizophrenia. It was hypothesised that the Y chromosome variants can increase risk to ADHD and schizophrenia and this was tested by performing two case-control association studies, one for ADHD and one for schizophrenia.

2. Testing for a modifying effect of Y chromosome variants on the phenotype of ADHD and schizophrenia. It was hypothesised that Y chromosome variation can have a modifying effect on the disorders by influencing clinical presentation or cognitive ability. This hypothesis was tested by selecting sexually dimorphic clinical parameters and cognitive measures for each disorder and testing them for association with the Y chromosome variants.

3. Testing for an indirect effect of Y chromosome variants on ADHD and schizophrenia. It was hypothesised that Y chromosome variation can interact with the gene encoding Tyrosine Hydroxylase, which is the rate-limiting enzyme in dopamine metabolism, and increase risk to the disorders. Evidence for interaction of Tyrosine Hydroxylase and SRY (Sex Determining Region on chromosome Y) has been found in male rats (Dewing 2006). This hypothesis was tested by performing two case-control association studies, one for ADHD and one for schizophrenia using Tyrosine Hydroxylase tag SNPs. Then, these SNPs were tested for a sex-specific effect on the disorders and, finally, Tyrosine Hydroxylase SNPs were tested for interaction with already genotyped Y chromosome SNPs.

4. Investigating the role of Y chromosome variation on cognitive function in the general population. Due to the fact that all tests for a modifying effect would be performed within the sample of patients without the use of a control

sample, we were interested in establishing whether the Y chromosome had a disease-specific effect or its effect could be generalised in the general population as well. This hypothesis was tested by genotyping two Y chromosome SNPs in male individuals from the AVON Longitudinal Study of Parents and Children (ALSPAC). The Y chromosome variants were then tested for association with cognitive ability in all the individuals from ALSPAC and in a subset of the sample with individuals who had an ADHD diagnosis.

Chapter 6 Materials and Methods

6.1 DNA samples

The DNA samples used in this study came from a sample of 331 patients with ADHD, a sample of 487 patients with schizophrenia and 1,076 individuals from the 1958 UK Birth Cohort. The characteristics of all samples are described in Materials and Methods Sections of following Results Chapters.

6.1.1 DNA extraction and storage

High molecular weight DNA was extracted from either lymphocytes in venous whole blood or from buccal cavity epithelial cells via saline mouthwash. The DNA preparation consisted of phenol/chloroform DNA extraction followed by ethanol precipitation according to routine procedures. All DNA samples used in this study were extracted by other members of the Department of Psychological Medicine research group. All stock and diluted samples were stored at -20°C in water or TE buffer.

6.1.2 DNA quantification and assessment

Extracted DNA was first quantified using a spectrophotometer (Beckmann Instruments, Fullerton, California, USA). Each DNA sample was diluted to a 5% solution in sterile water (5μl DNA sample in 95μl of water). The absorbance (A) of UV light at 260nm and 280nm wavelengths (λ) was measured, and DNA concentrations were calculated on the assumption that an A260nm value of 1 was equivalent to 50μg of DNA. A ratio of A260nm to A280nm above a value of 1.8 indicated a suitable level of clean DNA and the absence of contaminating protein.

Samples were more accurately quantified using a Fluoroskan Ascent fluorometer (Thermo Labsystems, Altrincham, UK) and the PicoGreen doubled-stranded DNA quantification kit (Invitrogen, Oregon, USA). The PicoGreen reagent specifically

interacts with double-stranded DNA and is therefore a more accurate quantifying method than spectrometry. The samples were first diluted to less than 50ng/μl using spectrometer readings. Next the samples were diluted 1/100 with 1X TBE in a white 96 well cliniplate (Thermo Labsystems, Altrincham, UK). The PicoGreen working solution is prepared by adding 5μl of PicoGreen to 995μl of 1X TE.

In order to measure the DNA concentration, 100μl of the PicoGreen working dilution was dispensed into each diluted sample. The fluorometer measures the concentration of the sample using an UV excitation wavelength of 485nm and an emission wavelength of 538nm. A standard curve is then used to calculate the concentration of DNA for each sample. These values are adjusted so that each sample is at a concentration of 4ng/ml ± 0.5μg /ml.

6.2 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is an enzymatic *in vitro* cycling technique for the amplification of a specific DNA sequence that lies between two regions of known sequence. The enzyme used is a thermostable Taq polymerase (HotStart Taq, Qiagen, Valencia, California, USA) which performs the synthesis of a complementary strand from the DNA template in the presence of suitable buffers and deoxyribonucleotide triphosphates (dNTPs). Two oligonucleotide primers are designed to flank the specific region of DNA to be amplified and it is these primers which provide the double stranded starting point for the Taq polymerase to begin its 5' to 3' synthesis.

A PCR reaction is comprised of three basis steps, a denaturation step which produces a single stranded DNA template, a primer annealing step where the primers bind their complimentary sequence and an elongation step when the synthesis of DNA occurs. Each step is accompanied by controlled temperature changes using a MJ tetrad thermocycler (MJ Research, Rayne, UK) and there are typically 30-45 cycles per reaction. PCR primers in this study were synthesised by Sigma-Aldrich Company Ltd. (Dorset, UK) and Metabion (Martinsried, Germany).

6.3 Agarose Gel Electrophoresis

DNA fragments can be fractioned according to their size when a potential difference is applied through a porous substance such as an agarose gel, as the negative phosphate groups allow the movement of DNA towards the anode. Analysis of pre- or post-PCR samples was performed using 0.5-2% agarose gels, dependent upon fragment size and resolution required. To construct a 1% agarose gel for electrophoresis, 1g of agarose Sigma-Aldrich Company Ltd. (Dorset, UK) was dissolved in 100ml 0.5x TBE buffer (Ultra pure electrophoresis sequencing grade, National Diagnostics). The mixture was heated until the solution was clear and then cooled and 1µl Ethidium Bromide solution (Fischer Scientific, UK) (10mg/ml) added. The solution was then poured into a gel-former and appropriate well formers added before allowing the gel to cool and form a solid.

To run samples on the gel, an appropriate volume of PCR product was mixed with a loading buffer (6x loading buffer: 15% Ficoll, 0.25% Bromophenol blue, 0.25% Xylene cyanol, in water) and placed in a formed well. An aliquot of size-standard (1kb plus DNA ladder, Invitrogen, Oregon, USA) was also run alongside samples to allow size delineation. The gel and sample were then run at between 100-120V in an electrophoresis tank for an appropriate amount of time to see the DNA size expected.

Samples assayed by agarose gel electrophoresis were visualised using a UV transilluminator (UVP, Upland, California, USA) and photographs taken using an attached Kodak Electrophoresis Gel documentation and analysis system (Eastman Kodak Company, Rochester, New York, USA).

6.4 Sample Processing

In order to set up large scale PCR and post-PCR reactions, DNA and reagent mastermixes, were aliquoted into suitable (96 or 384 well) microtitre plates (ABgene, Epsom, UK) by robot liquid handling systems. DNA samples were usually stored within shallow well DNA boxes (ABgene, Epsom, UK). DNA samples and pre-PCR reagents were aliquoted using a Biomek FX Laboratory Automation Workstation (Beckman Coulter Inc, Fullerton, California, USA). Post-PCR reagents were

aliquoted using a Biomek NX Laboratory Automation Workstation (Beckman Coulter Inc, Fullerton, California, USA). Programs for the Beckman-Coulter FX and NX microdispensers were written by Sarah Dwyer.

6.5 Genotyping using iPLEX MassARRAY

Genotyping for this project was performed using the MassARRAY genotyping platform due to its ability to provide highly accurate genotyping combined with a high multiplexing level. The Sequenom MassARRAY platform combines iPLEX GOLD primer extension chemistry with MALDI-ToF (Matrix Assisted Laser Desorption Ionisation – Time of Flight) Mass Spectrometry (MS).

The initial step of MassARRAY genotyping involves the design of a multiplex assay using the Sequenom Design Assay software. For each polymorphism that an assay needs to be designed the flanking DNA sequence is obtained and other polymorphisms or characteristics of the sequence that are able to confound genotyping are highlighted to prevent an assay being designed over them. The software, then designs PCR and extension primers to the highest multiplex level. For the PCR primers to be detected by the MALDI-ToF mass spectrum a 10bp non-specific tag sequence is added to the 5' end of the PCR primers. The primer annealing temperature is close to 56°C for all primers to allow them to anneal at the same time during PCR. Regardless of the number of SNPs, the reaction conditions are universal for all stages with optimisation only required for the adjustment of the extension primer concentration.

6.5.1 Sequenom PCR

Each PCR is performed in a 384-well microtitre plate (ABgene, Epsom, UK) by adding 5μl of PCR to 3μl of dried genomic DNA (4ng/μl). The PCR primer mix composition can be seen in the following table (Table 6-1).

Table 6-1. Sequenom PCR mix composition

Reagent	Volume (µl)
Water	3.35
RCR buffer 10X (Qiagen, USA)	0.625
$MgCl_2$ (25mM) (Qiagen, USA)	0.323
dNTPs (25mM) (Amersham Biosciences, UK)	0.1
HotStart Taq (Qiagen, USA)	0.1
PCR primer mix (forward and reverse) (1pmol/µl)	0.5

The PCR program is the following:

95°C for 15mins

94°C for 20s

56°C for 30s

72°C for 1min

Repeat steps 2-6 for 44 cycles

72°C for 3mins

15°C for 10mins

After PCR, positive and negative control samples are electrophorised on a 2% gel to check for PCR efficiency and any possible contamination. If there is no sign of contamination, a mix of Shrimp Alkaline Phosphatase (SAP) is added. The consistency of the mix for one sample is shown on Table 6-2.

Table 6-2. SAP mix composition

Reagent	Volume (µl)
Water	1.53
SAP	0.3
SAP Buffer	1.53

2µl of this product is added to the previous PCR products and the reaction undergoes the following thermocycling conditions:

37°C for 30mins

85°C for 10mins

95°C for 5mins

15°C for 10mins

The next stage involves the addition of an extension primer mix (Table 6-3). All unextended extension primers are added at the same time, along with ddNTPs and their concentration needs to be optimised. The extension primers are categorised into four groups according to their mass (group 1 has the lowest mass and group 4 has the highest mass) because products with lower mass generate a lower signal to noise ratio. The groups are then diluted to the following concentrations: 0.938µM, 1.17µM, 1.425µM and 1.875µM. During the optimisation procedure concentrations are adjusted according to the peak height obtained. Low peaks may require an increase in the extension primer concentration to increase the signal to noise ratio while the opposite is usually true for very high peaks. Failed or abnormal assays are usually removed at this stage.

Table 6-3. Extension primer mix composition

Reagent	Volume (µl)
iPLEX GOLD reaction buffer	0.2
iPLEX GOLD termination mix	0.2
iPLEX GOLD enzyme	0.041
Adjusted primer mix	1.559

2µl of this product is added to the previous products and the reaction undergoes the following thermocycling conditions:

94°C for 30s

94°C for 5s

52°C for 5s

80°C for 5s

Repeat step 3 for 5 cycles

Repeat step 2 for 40 cycles

72°C for 3mins

15°C for 10mins

The next step involves removing residual salts from the solution because ions can interfere with analysis by affecting the spectra obtained from the MALDI-ToF. This is performed by adding 6mg of clean-up resin using a Sequenom dimple plate and 25µl of water to the reaction mix. The samples are, then, mixed on a rotor for 20 minutes at least. After incubation, the resin is separated from the solution by centrifuging the samples for 15 minutes at 3000rpm (rotations per minute).

The samples are then ready to be spotted onto a Sequenom MassARRAY SpectroCHIP using a nanodispencer liquid handler (Sequenom, San Diego, California, USA). Each chip can accommodate up to 384 samples and it is composed of a combustible matrix. This matrix along with the samples ionise when excited by a

laser. The resulting ions are carried into gas phase and into the mass analyser. The ions are accelerated by applying voltage and are allowed to drift inside the flight tube, so that they separate according to their mass to charge ratios. This means that lighter ions travel faster than heavy ions to the detector. The time the ions need to travel the flight tube is converted to mass. The MALDI-ToF analysis is performed using the MassARRAY RT software (SpectroAcquire, Sequenom, San Diego, California, USA). This software takes into account the intensity of the mass signal or peak height and the extension primer yield, which is the successful extension of the unextended primer compared to the residual unextended primer, as well the assay design output file. The output of the software is a genotype for each sample. Figure 6-1 shows the output from the Typer software (Sequenom), which is used to view and manually revise genotypes. The three categories of genotypes (homozygotes and two types of heterozygotes) are given in different colours. Red indicates failed samples or negative controls.

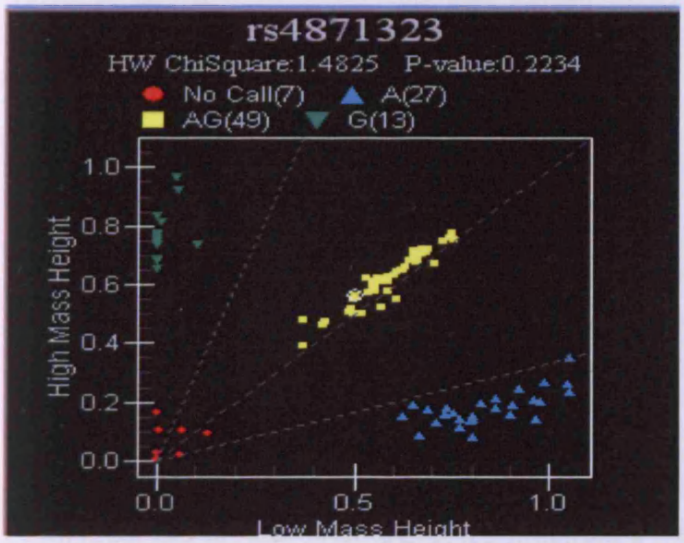

Figure 6-1. Screenshot from Typer analysis software showing two clusters of homozygotes (in green and blue), one cluster of heterozygotes (yellow) and failed samples or blanks (red).

To confirm that genotyping is accurate all assays are initially optimised in 30 CEPH (Centre d'Etude du Polymorphisme Humain) parent-offspring trios. Genotypes from these samples were compared to those on the HapMap and genotyping assays were only followed if the genotypes matched. All plates contained affected and non-affected individuals, negative controls and CEU samples as positive controls.

Genotypes were called twice with the second person blind to the genotypes called by the first one.

6.6 Bioinformatic and Statistical analysis

6.6.1 PLINK

PLINK is a freely available whole genome association analysis toolset (http://pngu.mgh.harvard.edu/~purcell/plink/), designed to perform a range of basic, large-scale analyses in a computationally efficient manner (Purcell et al. 2007). It has five main domains: data management for data recoding and formatting, summary statistics, which are useful for quality controls measures such as Hardy-Weinberg equilibrium calculations, call rates and inbreeding, population stratification detection, association analysis for both case-control and family-based samples and identity-by-descent estimation.

6.6.2 HAPLOVIEW

HAPLOVIEW is designed to simplify Linkage Disequilibrium (LD) and haplotype block analysis (http://www.broadinstitute.org/haploview/haploview). The software allows the user to import marker data such as a CEU HapMap dataset or case-control genotype data. The quality of the data can be assessed via the summary measures including Hardy-Weinberg equilibrium p values and percentage of individuals genotyped. LD measures for the SNPs imported are also available and they are used to create a graphical representation. The Tagger function uses the LD values to select "tag" markers for association studies after defining the selection criteria, such as r^2, minor allele frequency and type of analysis. HAPLOVIEW can also be used to test for haplotype association (Barrett 2009).

6.6.3 SPSS (Statistical Package for the Social Sciences)

SPSS is a statistical analysis software.

Some of its functions include:

- Descriptive statistics: Cross tabulation, Frequencies, Descriptives, Explore
- Bivariate statistics: Means, t-test, ANOVA, Correlation, Nonparametric tests
- Linear and logistic regression
- Prediction for identifying groups: Factor analysis, cluster analysis (two-step, K-means, hierarchical)

The many features of SPSS are accessible via pull-down menus or can be programmed with a command syntax language. SPSS datasets have a 2-dimensional table structure where the rows typically represent cases (such as individuals) and the columns represent measurements (such as age, sex). The graphical user interface has two views which can be toggled by clicking on one of the two tabs in the bottom left of the SPSS window. The Data View shows a spreadsheet view of the cases (rows) and variables (columns) while the Variable View shows a summary of the available variables. SPSS versions 12.0 and 16.0 has been used to perform statistical analysis for this project, including ANOVA, Mann Whitney U test, Pearson's chi-square test, independent samples t-test as well as testing whether the assumptions for each test were met.

6.6.4 The International HapMap Project

The International HapMap Project aims to develop a haplotype map (HapMap) of the human genome, which will describe the common patterns of human genetic variation (International HapMap Consortium http://www.hapmap.org/). The information produced by the project is made freely available to researchers around the world. The

International HapMap Project is a collaboration among researchers in Canada, China, Japan, Nigeria, the United Kingdom, and the United States. Several populations (30 trios of Yoruba people in Ibadan, Nigeria; 45 unrelated Japanese individuals from Tokyo; 45 unrelated Ham Chinese individuals and 30 trios from the US with ancestry from Northern and Western Europe which were collected by the Centre d'Etude du Polymorphisme Humain (CEPH)) were genotyped as part of the project. HapMap now contains information, including LD, on 3.1 million SNPs (phase II).

Chapter 7 The role of Y chromosome in ADHD

7.1 Introduction

Psychiatric disorders, as well as a large number of complex disorders, exhibit sex differences in their prevalence rates and clinical presentation. Sex differences in ADHD are well-established with higher rates of males being affected in both clinical and population samples. The disorder is also more severe and disruptive in boys. It is striking that boys tend to be more hyperactive, while girls are usually more inattentive (Cuffe et al. 2005).

Although sex differences are recognised in ADHD, research on their causes has been limited. Factors that might contribute to these differences include hormones, genetic and environmental factors. Studying the genetic factors contributing to sex differences could shed light on the causes of ADHD and other psychiatric disorders with a male preponderance. There is evidence from animal models that sex chromosome genes could contribute to sexual differentiation of brain and behaviour without the involvement of gonadal hormones (Agate et al. 2003; Dewing et al. 2003). This suggests that sex chromosome genes may play an important role in psychiatric disorders and, as discussed elsewhere, the focus of this thesis is on the Y chromosome.

Y-linked genes are only expressed in males and they have been shown to play a role in the sexual dimorphism of the brain. At least ten Y chromosome genes are expressed in the brain (Xu et al. 2002). The Y chromosome has been linked to behavioural phenotypes using the four core genotypes mouse model (Guillot et al. 1995). In addition, a Y chromosome gene, *SRY*, has been suggested to be involved in the regulation of the dopaminergic pathway, which is affected in ADHD. This regulation is mediated by an interaction of *SRY* with the rate-limiting enzyme in dopamine metabolism, tyrosine hydroxylase (*TH*) (Dewing et al. 2006). Individuals with an extra Y chromosome, 47,XYY syndrome, appear to be at increased risk of developing antisocial behaviour, which could be mediated by an effect on hyperactivity (Gotz et al. 1999). Individuals with Klinefelter's syndrome have an impairment of executive function and inhibitory deficits (Temple and Sanfilippo

2003). Finally, a case of ADHD carrying a rare deletion of the long arm and duplication of the short arm of the Y chromosome has been reported (Mulligan et al. 2008).

Compared to autosomes or the X chromosome, the Y chromosome has been widely neglected in genetic and genomic studies. The majority of the Y chromosome does not recombine with the X chromosome and it is passed unchanged from one generation to the next apart from mutations. The lack of recombination poses difficulties in performing traditional family-based association studies and mapping studies but at the same time it provides stable haplogroups. It is important to take into account that the distribution of Y chromosome haplogroups depends highly on the geographic origin of the individuals. The confusing nomenclature makes it necessary to rely on a reliable source of information about Y chromosome markers. This source is the Y Chromosome Consortium (The Y Chromosome Consortium 2002). Using data from this consortium it is possible to select the appropriate haplogroups and the markers to genotype in order to capture them for a particular population. Comparing the frequencies of Y chromosome haplogroups between cases and controls will show if specific haplogroups are associated with or modify the phenotype of a disorder.

7.2 Aims of the study

The present study investigated the role of Y chromosome variation in ADHD. Y chromosome haplogroups appropriate for UK populations were selected according to Y Chromosome Consortium data and personal communication with experts in the field.

The hypothesis that Y chromosome SNPs and haplogroups are associated with ADHD was tested by performing a case-control association study. Both allelic and haplogroup analyses were performed, although they provide similar information, in order to confirm any positive findings. Haplogroup analysis also provided evidence about the population distribution of Y chromosome haplogroups, since this is the first study to genotype Y chromosome markers in a large sample of UK males.

The hypothesis that Y chromosome variation is associated with modifying effects on ADHD by affecting clinical presentation, cognitive performance or comorbidity was tested within the group of patients with ADHD after stratifying them according to their Y chromosome SNP alleles or haplogroup.

7.3 Materials and methods

7.3.1 ADHD sample description

The ADHD sample used in this study consists of 210 boys who were consecutively referred to district child and adolescent psychiatry and paediatric clinics in South Wales, the South-West of England, Greater Manchester and Cheshire, with a suspected diagnosis of ADHD. To ensure genetic homogeneity, participants were required to be of British Caucasian origin to the index child's grandparents. Children who suffered from major medical or neurological conditions, autism, Tourette's syndrome or pervasive developmental disorders were excluded from the study.

Ethical approval was obtained from the North West England Multicentre Research Ethics Committee. Informed written consent from parents and assent from children was obtained for all participating families.

The age of the children was between six and sixteen years with a mean age at assessment of 9 years and 3 months (SD: 2 years and 6 months). The mean full scale IQ test score was 89 (SD: 11.2; range: 70-123). The presence of comorbid disorders is summarised in Table 7-1.

Table 7-1. Co-morbid disorders present in the sample of 210 boys with ADHD

Co-morbid disorder	(%)
Oppositional Defiant Disorder (ODD)	47.6
Conduct Disorder (CD)	13.8
Anxiety disorder	5.2
Depressive disorder	0.5
Transient or chronic tics	10

7.3.2 ADHD sample assessment

Assessments of ADHD, conduct disorder and other psychiatric disorders were undertaken by trained interviewers using the parent version of the Child and Adolescent Psychiatric Assessment (CAPA) (Angold et al. 1995). This is a semi-structured interview used to obtain diagnoses of ADHD and common co-morbid conditions according to ICD-10 (World Health Organisation 1993), DSM-IV (American Psychiatric Association 1994) and DSM-III-R (American Psychiatric Association 1987) criteria. The CAPA also gave information about the total number of ADHD symptoms, the number of hyperactivity/impulsivity and inattention symptoms and the number of Conduct Disorder (CD) symptoms, according to DSM-IV diagnostic criteria. The DSM-IV CD symptoms were coded as present or absent and summed to yield a total antisocial score. Items included behaviours such as "physical cruelty to other people", "sets fires" and "crime involving confrontation with the victim". All DSM-IV CD symptoms used to derive the total score in this sample were of childhood onset (onset less than 10 years). Pervasiveness of ADHD symptoms in a school setting was assessed by interviewing the class teacher over the telephone using the Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (ChATTI) (Holmes et al. 2004).

In order to assess the Intelligence Quotient (IQ), all boys with ADHD diagnosis were assessed using the Wechsler Intelligence Scale for Children, version III "WISC-III-UK" (Wechsler 1992). This gives scores for the Full Scale IQ (FSIQ), the Verbal IQ (VIQ) and the Performance IQ (PIQ). Only those children with FSIQ score of ≥ 70 were included in the study. Reading ability was assessed using the Wechsler Objective Reading Dimension "WORD" (Wechsler 1992).

A range of questionnaire measures were also completed by the parents, providing demographic and environmental information. Mothers stated the occupation of the main earner in the household. Social class was obtained from this description based on the UK Standard Occupational Classification (2nd Edition) (Office of Population Censures and Surveys 1995). In an attempt to make the categories more meaningful and easier to interpret, they were split into three categories; high, medium and low social class. High social class (21%) consisted of families from professional and managerial jobs in classes 1 and 2; medium social class (29%) consisted of families

from skilled occupations in classes 3 and 4; low social class (50%) consisted of families from unskilled jobs in class 5 and unemployed or unclassified individuals.

7.3.3 1958 UK birth cohort control sample

The control sample used in this study consists of 637 men drawn from the 1958 British Birth Cohort. This is a longitudinal study which takes as its subjects all the people born in England, Scotland and Wales in one week in March 1958. For information see the website of the Centre of Longitudinal Studies (http://www.cls.ioe.ac.uk/). Permission to genotype was obtained.

7.3.4 DNA samples and genotyping

DNA was obtained from venous blood or saliva samples for the ADHD patients. DNA samples were obtained from the 1958 UK Birth Cohort after request. Genotyping was performed with the MassARRAY and iPLEX systems (Sequenom, San Diego, CA) according to the manufacturers' recommendations. Assays were optimised in 30 male CEU individuals. All plates contained cases, controls, blanks, and CEU samples. Genotypes were called in duplicate blind to sample identity and blind to the other rater. All SNPs included in the association analysis had a call rate of >95%. Individuals were only included in analysis if they had genotypes for over 80% of the SNPs analysed. For more information about quality control procedures see Appendix I.

7.3.5 Selection of Y chromosome SNPs

The SNPs in the Y chromosome panel were chosen to reflect the origin of the individuals under study. Personal communication with researchers in the Y chromosome field and especially Professor Mark Jobling (University of Leicester) and Dr Peter Underhill (Stanford University Medical Center) was vital in choosing the Y chromosome haplogroups that were more likely to be found in a UK population. It is important to mention that previously there has not been a study of Y chromosome markers and/or haplogroups in a representative UK population. All previous studies

have focused mainly on genotyping a large number of SNPs in a small number of individuals from population isolates. The largest study has been performed by Capelli et al. (2003) and it involved genotyping 1,772 men. However, the sample still came from small urban locations in the UK and it did not represent the UK as a whole. A high degree of haplogroup breakdown would not be desirable for a psychiatric genetics study such as ours. It was decided that the haplogroups most likely to be frequent in the UK were I and R.

The next step, after deciding which haplogroups would be more frequent in the UK, was to use the Y Chromosome Consortium phylogenetic map (The Y Chromosome Consortium 2002) to locate the SNPs that would need to be typed in order to capture each haplogroup. By following the branches of the tree (see Figure 6-1 in Chapter 6) we were able to find the relevant SNPs. In cases where there were two or more SNPs providing similar information, the best SNP was chosen according to information about the sequence characteristics around the SNP, since they can affect the performance of the SNP in the genotyping platform. Recurrence was also taken into account. Using the Sequence Assay Design software a panel of 9 Y chromosome SNPs was designed. Figure 7-1 shows the SNPs genotyped (Y chromosome SNPs start with a capital M) and the haplogroups (with capital letters) they represent.

7.3.6 Categorisation of individuals in haplogroups

Y chromosome SNPs are used to categorise each individual to a single haplogroup. Each Y chromosome SNP has two alleles: an ancestral allele, which is the most ancient evolutionarily and a derived allele, which is the most recent. The ancestral allele is represented with 0 and the derived allele with 1. The ancestral and derived alleles for the SNPs in this panel are shown in Table 7-2. Each individual is categorised to one haplogroup according to whether they have the ancestral or derived allele of each Y chromosome SNP starting from the 'root marker', the SNP that defines the clade and continuing splitting the sample into different branches. For this particular set of Y chromosome SNPs, M89 is the 'root marker' or the most ancient marker. M9 and M173 also define major haplogroups. 'Root markers' can be seen in Figure 7-1, since they are in a "higher level" than the rest of the SNPs. For more

information about the combination of Y chromosome SNP alleles in each haplogroup see Appendix II.

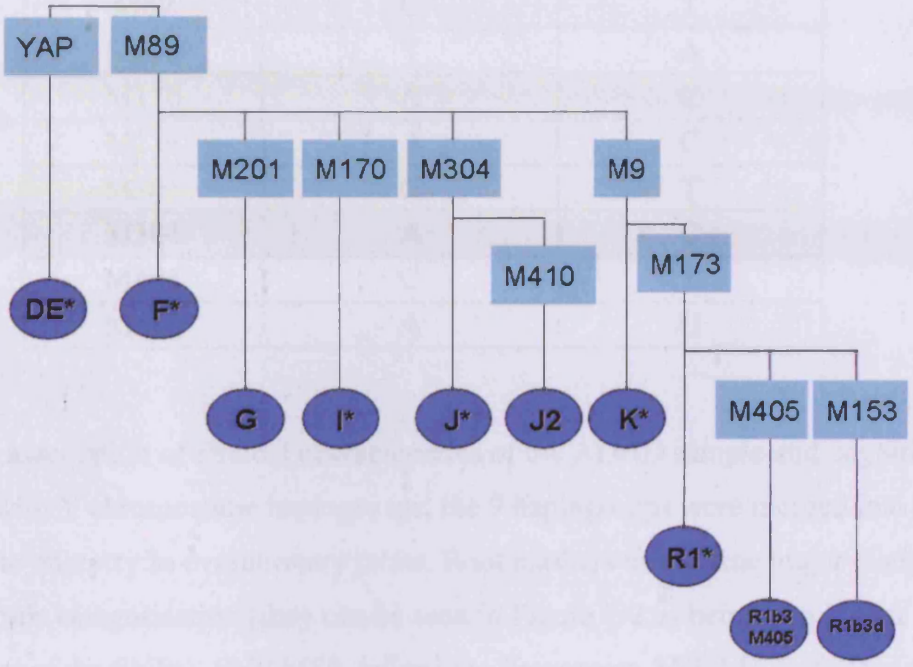

Figure 7-1. Y chromosome haplogroup tree showing the relationship between genotyped SNPs. Y chromosome haplogroups are represented by capital letters and SNPs start with M. There were no individuals belonging in haplogroup F*.

Table 7-2. Ancestral (0) and derived allele (1) for each Y chromosome SNP

Y chromosome SNPs	Ancestral allele (0)	Derived allele (1)
M9	C	G
M89	C	T
M153	T	A
M170	A	C
M173	A	C
M201	G	T
M304	A	C
M405	C	T
M410	A	G

To test for association of clinical characteristics of the ADHD sample and cognitive measures with Y chromosome haplogroups, the 9 haplogroups were merged into 3 according to ancestry in evolutionary terms. Root markers that define major clades facilitated this categorisation (they can be seen in Figure 7-2 as being in a higher level than the rest of the SNPs). SNP M89 defined the first group, SNP M173 defined the second group and SNP M405 defined the third group. This way there were enough individuals in each group to allow statistical analysis (Figure 7-2).

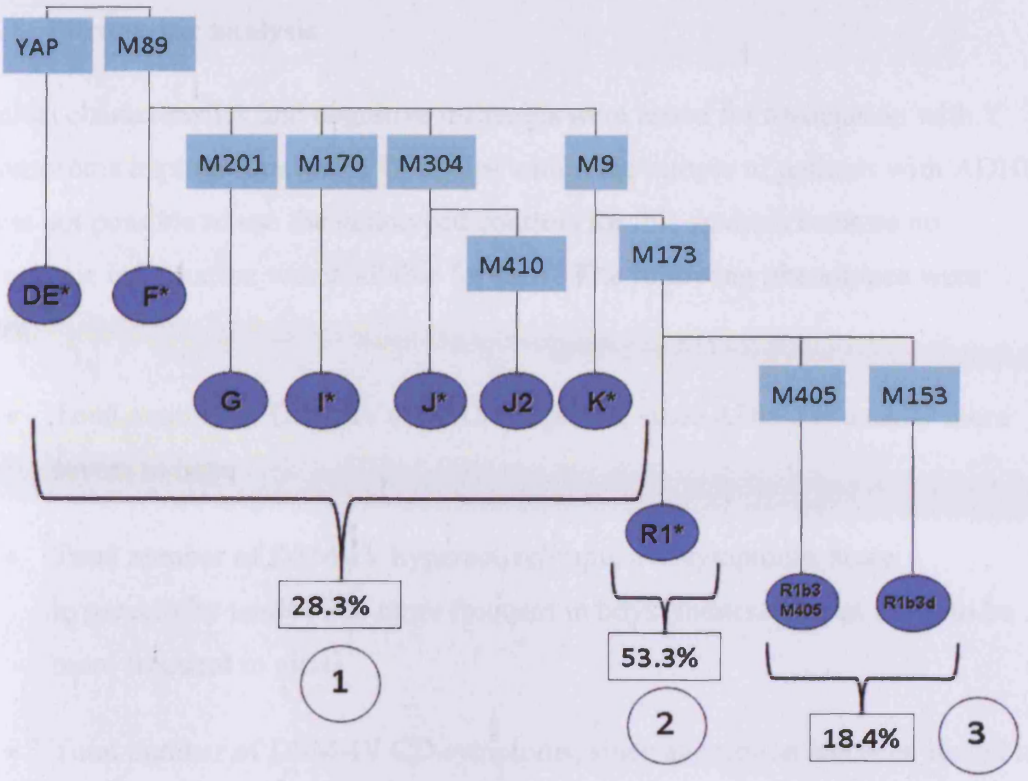

Figure 7-2. The Y chromosome haplogroups captured in this study were merged into three to facilitate phenotypic analysis. There were no individuals belonging in haplogroup F*. Percentages represent the frequencies of the three Y chromosome haplogroups in 4,950 UK men genotyped in total in this study.

7.3.7 Phenotypic analysis

Clinical characteristics and cognitive measures were tested for association with Y chromosome haplogroups and SNP alleles within the sample of patients with ADHD. It was not possible to use the genotyped controls for this analysis because no phenotypic information was available for them. The following phenotypes were tested:

- Total number of DSM-IV ADHD symptoms, since ADHD is usually more severe in boys
- Total number of DSM-IV hyperactive/impulsive symptoms, since hyperactivity tends to be more frequent in boys (inattentiveness tends to be more frequent in girls)
- Total number of DSM-IV CD symptoms, since aggression has been linked to the Y chromosome. Only children with ADHD and early aggression as indexed by CD symptoms present before the age of ten were used, since they are considered to be a distinct group from children with ADHD only (see Chapter 2 for more details)
- Total score on the WORD, since reading problems and dyslexia are also more prevalent in males
- Full scale IQ scores, performance IQ scores, verbal IQ scores as a general measure of cognitive function

7.3.8 Statistical analysis

The Pearson's chi-square test [degree of freedom (df=1)] was used to test for allelic association and also for haplogroup analysis. The threshold of significance was set at $p \leq 0.05$. Correction for multiple testing would be required in case of significant associations. Then, a Bonferroni correction or permutation testing would be applied. With this threshold of significance, this study had >95% statistical power to detect effect sizes of 1.5 when the risk allele frequency was 0.3-0.4. For effect sizes of 1.3

the power was reduced to 70% and for effect sizes of 1.2 the power was 41%. For rarer alleles with frequencies of 0.1-0.2 the power was 90% for effect sizes of 1.5, 54% for effect sizes of 1.3 and 30% for effect sizes of 1.2. The prevalence of ADHD was considered to be 4% according to Cuffe et al. (2005). Statistical power calculations were performed using the Genetic Power Calculator (Purcell et al. 2003).

Clinical characteristics of the ADHD sample and cognitive measures were tested for association with Y chromosome haplogroups. All phenotypes were chosen *a priori* due to evidence of association with the Y chromosome or because they were known to be sexually dimorphic. Normality of distribution for each of the dependent variables was assessed using measures of skewness and kurtosis (they should not exceed 1 and 3 respectively). The significance level of the Kolmogorov-Smirnov test was also taken into account.

Haplogroups were tested for association with all the above phenotypic measures using ANOVA after collapsing them into three groups (Figure 7-2) to increase the numbers in each group. When a positive association was detected, non-parametric tests were performed to confirm the results. Secondary analysis was also performed in order to exclude a confounding effect of social class or location where there were significant findings. Pearson's chi-square tests (df=4 for social class and df=2 for geographical location) were used to assess the effect of the two possible confounders on the distribution of Y chromosome haplogroups. Independent samples t-test was used to assess the effect of geographical location on PIQ and FSIQ and ANOVA was used to assess the effect of social class on PIQ and FSIQ. The threshold of significance was again set at $p \leq 0.05$.

Clinical variables were also tested for association with single SNPs that had an adequate number of individuals in order to perform the analysis (MAF>15%) using independent samples t-test or the non-parametric Mann-Whitney U test according to whether or not data were normally distributed. This was performed as a secondary analysis to confirm the results from the haplogroup analysis and further explore the mutation responsible for any association with specific Y chromosome haplogroups.

7.4 Results

7.4.1 Case-control association study

7.4.1.1 Allelic analysis

Nine Y chromosome SNPs were genotyped in the sample of 210 male patients with ADHD and 637 male control individuals from the 1958 Birth Cohort. Table 7-3 summarises these data and shows the frequencies of Y chromosome SNP alleles in cases and controls. There was no significant difference between cases and controls (Table 7-3). It was not possible to perform statistical analysis for SNPs with very low frequencies.

Table 7-3. Allelic analysis for 9 Y chromosome SNPs

Y chromosome SNPs	MAF (CASE) %	MAF (CONTROL) %	p value	OR	95% CI
M9	29.4	27.9	0.66	1.08	0.77 - 1.51
M89	3.6	3.4	0.90	1.05	0.46- 2.4
M153	0	0.2	0.56	n/a	n/a
M170	21.9	20.5	0.66	1.09	0.75 - 1.58
M173	31.4	28.5	0.42	1.15	0.82 - 1.6
M201	2.4	1.5	0.40	1.66	0.5 - 5.45
M304	2.8	2.3	0.71	1.20	0.46 - 3.13
M405	16.3	19.5	0.29	0.80	0.54 - 1.21
M410	1.8	1.9	0.96	0.97	0.31 - 3.04

MAF: Minor Allele Frequency; OR: Odds Ratio; CI: Confidence Interval; n/a: non-applicable

7.4.1.2 Haplogroup analysis

The frequencies of the nine Y chromosome haplogroups captured in this study (Figure 7-1) were compared between cases with ADHD and controls. There were no individuals belonging in haplogroup F*. Table 7-4 summarises the results. Pearson's chi-square test (df=1) showed no significant difference between cases and controls in terms of Y chromosome haplogroups (Table 7-4). It was not possible to perform statistical analysis for haplogroups with very low frequencies (J*, K* and R1b3d).

Table 7-4. Frequencies of Y chromosome haplogroups in 210 cases with ADHD and 637 controls

Y chromosome haplogroup	Cases (%)	Controls (%)	p value
DE	2.4	3.3	0.45
G*	1.9	1.4	0.62
I*	22.3	20.7	0.52
J*	0.9	0.6	n/a
J2	1.9	1.7	0.87
K*	0.9	0.5	n/a
R1*	53.1	52.1	0.79
R1b3M405	16.6	19.5	0.36
R1b3d	0	0.2	n/a

n/a: non-applicable

7.4.2 Phenotypic analysis using haplogroups

To test for an association of the previously described phenotypes with Y chromosome haplogroups, the 9 haplogroups were merged into 3 according to ancestry. As shown in the graphs, DSM-IV ADHD symptoms (N=210), DSM-IV hyperactive/impulsive symptoms (N=210) (Figure 7-3) and DSM-IV CD symptoms (N=78) (only in those with early CD) were approximately the same across each group (Figure 7-4).

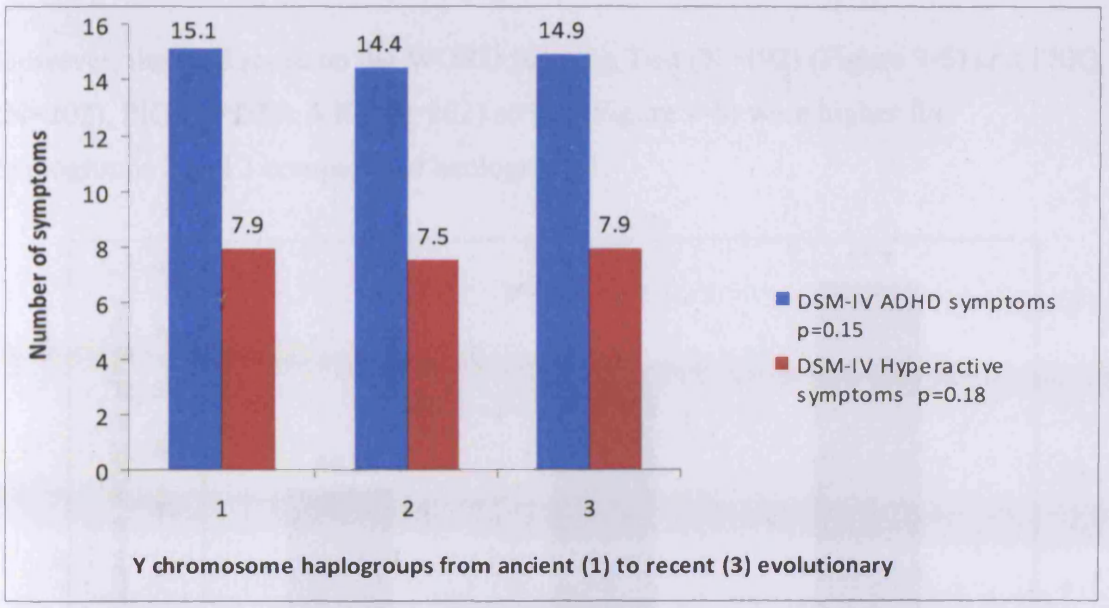

Figure 7-3. Total number of DSM-IV ADHD symptoms and DSM-IV hyperactive/impulsive symptoms (n=210) in 3 Y chromosome haplogroups within the sample of patients with ADHD. p values obtained from ANOVA comparing each of the variables separately in 3 Y chromosome haplogroups are shown on the graph

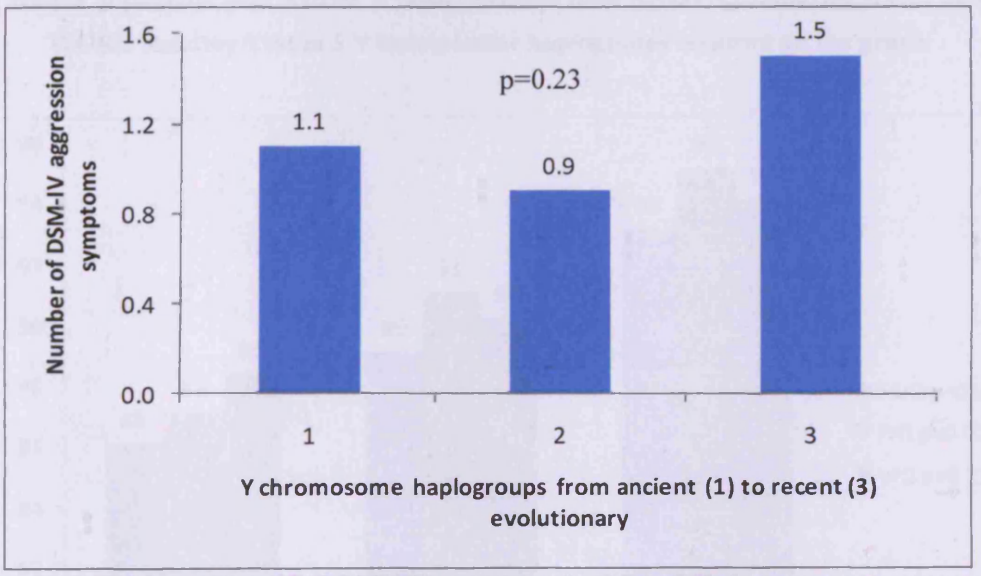

Figure 7-4. Total number of DSM-IV CD symptoms (n=80) in 3 Y chromosome haplogroups within the sample of patients with ADHD. p value obtained from ANOVA comparing DSM-IV CD symptoms in 3 Y chromosome haplogroups is shown on the graph

However, the total score on the WORD Reading Test (N=192) (Figure 7-5) and FSIQ (N=202), PIQ (N=202), VIQ (N=202) scores (Figure 7-6) were higher for haplogroups 2 and 3 compared to haplogroup 1.

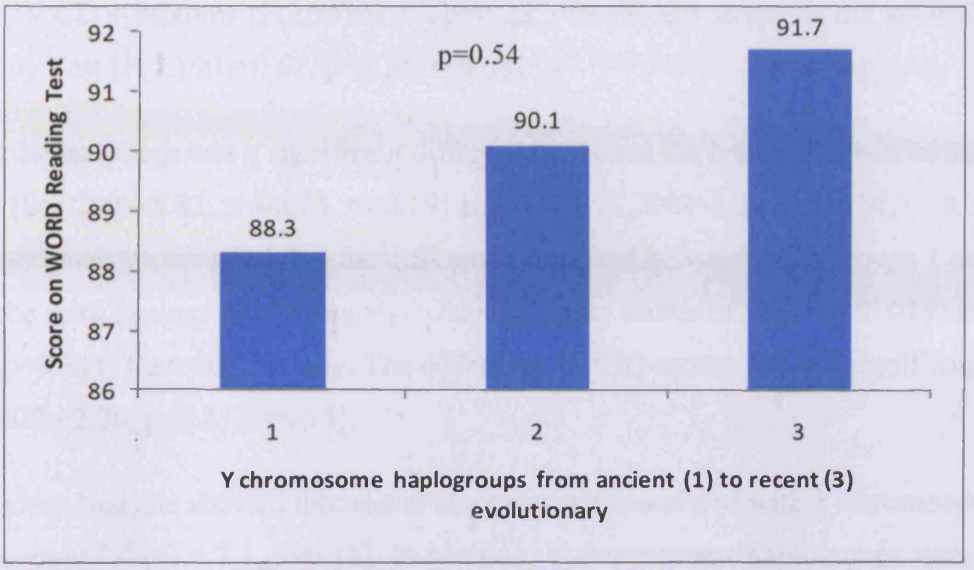

Figure 7-5. Scores on the WORD Reading Test (n=192) in 3 Y chromosome haplogroups within the sample of patients with ADHD. p value obtained from ANOVA comparing scores on the WORD Reading Test in 3 Y chromosome haplogroups is shown on the graph

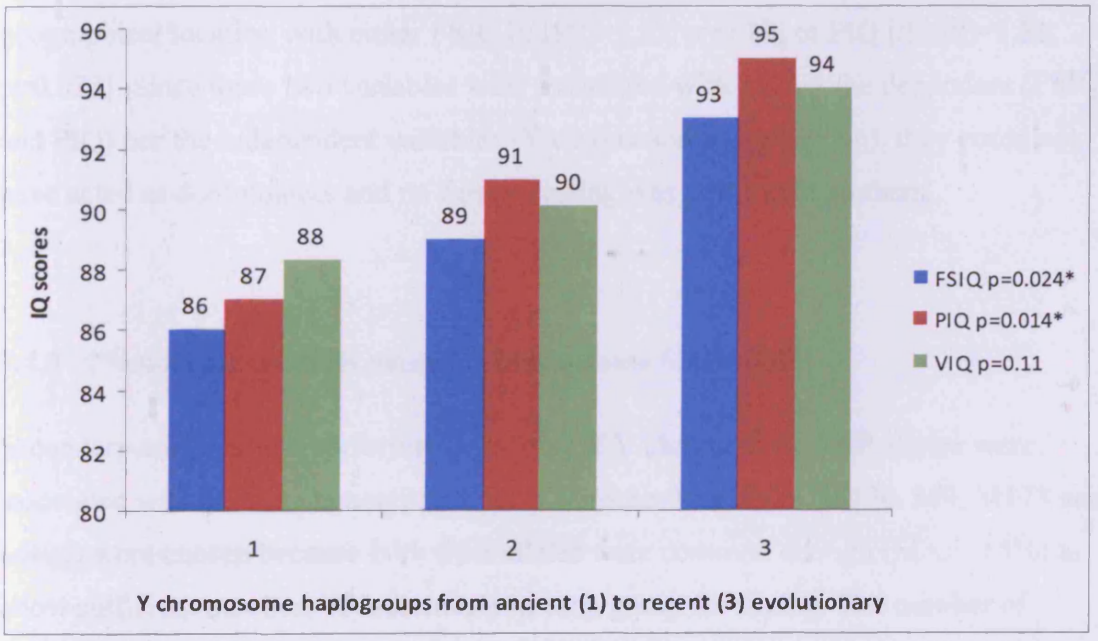

Figure 7-6. FSIQ, PIQ and VIQ scores (n=202) in 3 Y chromosome haplogroups within the sample of patients with ADHD. p values obtained from ANOVA comparing each of the variables separately in 3 Y chromosome haplogroups are shown on the graph; *: significant at the $p \leq 0.05$ level

Performing ANOVA showed that there was no significant difference in the number of DSM-IV ADHD symptoms [F(2,208)=1.95, p=0.15, r=0.14], the number of DSM-IV hyperactive/impulsive symptoms [F(2,208=1.74, p=0.18, r=0.13], the number of DSM-IV CD symptoms [F(2,76)=1.15, p=0.23, r=0.19] and scores on the WORD Reading Test [F(2,190)=0.62, p=0.54, r=0.08].

Nevertheless, there was a significant difference between the haplogroups in terms of FSIQ [F(2,200)=3.82, p=0.024, r =0.19] and PIQ [F(2,200)=4.39, p=0.014, r =0.21]. *Post hoc* analysis revealed that the difference appeared between haplogroups 1 and 3 with the third haplogroup having significantly higher scores of FSIQ (p=0.019) and PIQ (p=0.011) than the first one. The difference in VIQ scores was not significant [F(2,200)=2.26, p=0.11, r=0.15].

Secondary analysis showed that social class was not associated with Y chromosome haplogroups [χ^2 (4) = 7.1, p=0.13]. In addition, Y chromosome haplogroups were not associated with whether the individuals came from England or Wales [χ^2 (2) = 1.6, p=0.46]. There was also no association of social class with either FSIQ [F(2,179)=2.09, p=0.13] or PIQ [F(2,179)=2.41, p=0.093] and no association of geographical location with either FSIQ [t(199)=1.33, p=0.19] or PIQ [t(199)=1.81, p=0.072]. Since these two variables were associated with neither the dependent (FSIQ and PIQ) nor the independent variables (Y chromosome haplogroup), they could not have acted as confounders and no further testing was performed on them.

7.4.3 Phenotypic analysis using Y chromosome SNP alleles

Secondary analysis was performed by testing if Y chromosome SNP alleles were associated with certain phenotypes. Four Y chromosome SNPs (M170, M9, M173 and M405) were chosen because both their alleles were common enough (MAF>15%) to allow sufficient numbers of individuals in each group for testing. The number of individuals used for testing differs due to missing data for certain measures on some individuals.

Table 7-5. Testing for association of Y chromosome SNP alleles and clinical characteristics within the sample of patients with ADHD - Mann Whitney U test

	Y chromosome SNPs	Allele	N	Mean	95% CI around the mean	p value
Total number of DSM-IV ADHD symptoms	M170	Ancestral	164	14.5	14.1 to 14.9	0.018*
		Derived	46	15.5	14.8 to 16.1	
	M9	Ancestral	61	15.2	14.6 to 15.8	0.1
		Derived	149	14.5	14.1 to 14.9	
	M173	Ancestral	63	15.1	14.6 to 15.7	0.12
		Derived	147	14.5	14.1 to 14.9	
	M405	Ancestral	175	14.7	14.3 to 15	0.57
		Derived	35	14.9	14 to 15.8	
	Y chromosome SNPs	**Allele**	**N**	**Mean**	**95% CI around the mean**	**p value**
Total number of DSM-IV hyperactive/ impulsive symptoms	M170	Ancestral	164	7.6	7.4 to 7.9	0.13
		Derived	46	8	7.7 to 8.4	
	M9	Ancestral	61	7.9	7.6 to 8.2	0.5
		Derived	149	7.6	7.4 to 7.9	
	M173	Ancestral	63	7.9	7.6 to 8.2	0.42
		Derived	147	7.6	7.4 to 7.9	
	M405	Ancestral	175	7.6	7.4 to 7.9	0.34
		Derived	35	7.9	7.4 to 8.4	
	Y chromosome SNPs	**Allele**	**N**	**Mean**	**95% CI around the mean**	**p value**
Total number of DSM-IV CD symptoms within the sample of patients with ADHD and early CD symptoms	M170	Ancestral	59	1.1	0.8 to 1.3	0.82
		Derived	19	1.1	0.6 to 1.6	
	M9	Ancestral	23	1.1	0.7 to 1.5	0.57
		Derived	55	1	0.8 to 1.3	
	M173	Ancestral	25	1.1	0.8 to 1.5	0.55
		Derived	53	1.1	0.8 to 1.3	
	M405	Ancestral	64	1	0.7 to 1.2	0.065
		Derived	14	1.5	0.9 to 2.1	

CI: Confidence Interval; CD: Conduct Disorder;
*: significant at the $p \leq 0.05$ level

Table 7-6. Testing for association of Y chromosome SNP alleles and total score on the WORD - Independent samples t-test

Y chromosome SNPs	Allele	N	Mean	95% CI around difference in means	p value
M170	Ancestral	150	90.7	1 to 8.9	0.12
	Derived	42	86.7		
M9	Ancestral	57	14.3	-7.3 to 1.7	0.23
	Derived	135	14.5		
M173	Ancestral	59	88.3	-6.7 to 2.3	0.33
	Derived	143	90.5		
M405	Ancestral	167	91.7	-7.8 to 3.3	0.42
	Derived	35	89.4		

CI: Confidence Interval

Table 7-7. Testing for association of Y chromosome SNP alleles and cognitive measures within the sample of patients with ADHD - Independent samples t-test

	Y chromosome SNPs	Allele	N	Mean	95% CI around difference in means	p value
Full Scale IQ scores (FSIQ)	M170	Ancestral	160	90	1.1 to 8.7	0.012*
		Derived	42	85.1		
	M9	Ancestral	57	85.6	1.2 to 8	0.008**
		Derived	145	90.3		
	M173	Ancestral	59	86.2	0.5 to 7.3	0.023*
		Derived	143	90.1		
	M405	Ancestral	167	88.2	0.4 to 8.5	0.034*
		Derived	35	92.6		
	Y chromosome SNPs	**Allele**	**N**	**Mean**	**95% CI around difference in means**	**p value**
Performance IQ scores (PIQ)	M170	Ancestral	160	92	2.2 to 10.7	0.003**
		Derived	42	85.6		
	M9	Ancestral	57	86.6	1.8 to 9.5	0.004**
		Derived	145	92.2		
	M173	Ancestral	59	87.3	1 to 8.6	0.014*
		Derived	143	92.1		
	M405	Ancestral	167	89.7	0.7 to 10	0.025*
		Derived	35	96		
	Y chromosome SNPs	**Allele**	**N**	**Mean**	**95% CI around difference in means**	**p value**
Verbal IQ scores (VIQ)	M170	Ancestral	160	90.9	1.1 to 7.1	0.15
		Derived	42	87.9		
	M9	Ancestral	57	87.9	-7 to 0.4	0.077
		Derived	145	91.2		
	M173	Ancestral	59	11.4	-6.5 to 0.8	0.12
		Derived	143	12.1		
	M405	Ancestral	167	89.5	-8.5 to 0.3	0.067
		Derived	35	93.6		

IQ: Intelligence Quotient; CI: Confidence Interval;
*: significant at the $p \leq 0.05$ level; **: significant at the $p \leq 0.01$ level

We found no association between Y chromosome SNPs and number of ADHD symptoms, number of hyperactive/impulsive symptoms or number of CD symptoms (Table 7-5). There was only a significant association of SNP M170 and the number of DSM-IV ADHD symptoms but not for any of the other SNPs tested (Table 7-5). A significant association was demonstrated between all the SNPs tested and PIQ as shown by the independent samples t-test (Table 7-7). The same was true for FSIQ (Table 7-7). The results were confirmed by non-parametric tests and the p values obtained were similar. In three of the SNPs tested the derived allele was associated with higher PIQ and FSIQ scores. The ancestral allele was associated with higher IQ only for SNP M170.

7.5 Discussion

In this study, the hypothesis that Y chromosome variation is associated with ADHD or has a modifying effect by influencing clinical severity, cognitive performance or comorbidity was tested. The results of this Chapter indicate that there is no main effect of Y chromosome variation on ADHD. However, Y chromosome variation may have a modifying effect on ADHD by influencing cognitive performance, specifically, FSIQ [$F(2,200)=3.82$, $p=0.024$, $r=0.19$] and PIQ [$F(2,200)=4.39$, $p=0.014$, $r=0.21$] in patients with ADHD.

Since the Y chromosome is present only in men, the presence of Y chromosome genes can be responsible for sex differences in susceptibility to psychiatric disorders. However, the Y chromosome has been excluded from genetic studies. There have been very few studies of the Y chromosome in relation to psychiatric phenotypes (Jamain et al. 2002; Kittles et al. 1999; Shah et al. 2009) and most of them have used heterogeneous populations or very small sample sizes. The same is true for genome-wide association studies. No Y chromosome SNPs were included in genotyping arrays used to test for association with psychiatric disorders up until recently. The fact that a whole chromosome was excluded from studies that are designed to examine the whole genome shows the degree of neglect that the Y chromosome has received. For these reasons, this study is important since it is the first study to my knowledge to have genotyped Y chromosome markers in a systematic way in such a large sample of UK men. We were able to demonstrate that Y chromosome SNPs can be genotyped using the same genotyping platforms as for autosomal SNPs. The selection of SNPs for genotyping is the most critical aspect of Y chromosome studies and it differs considerably from the methods used for autosomal SNPs. Since linkage disequilibrium (LD) measures are not applicable for the Y chromosome, they cannot be used to select the minimum number of SNPs needed to cover a Y chromosome region. The small number of SNPs genotyped in HapMap individuals and the fact that most of the SNPs are not present in genetic databases, such as UCSC, poses another problem, since it is very difficult to obtain information about the frequency of Y chromosome SNPs in certain populations. A more appropriate way of selecting Y chromosome SNPs is to decide first on the haplogroups that need to be covered

according to the population to be studied and then resort to the Y Chromosome Consortium (The Y Chromosome Consortium 2002) in order to select the SNPs covering those haplogroups.

Phenotypic analysis was performed using both haplogroups and SNP alleles. These two types of analysis should yield the same pattern of results, since all the individuals were categorised in the Y chromosome haplogroups according to the Y chromosome SNP alleles that they possessed. The reason that both analyses were performed was because the aim was to test if any specific Y chromosome haplogroup had a modifying effect on ADHD and also to test if this effect was linked to a specific mutation on the Y chromosome.

In addition, by performing both analyses the evolutionary pattern of the results was observed. According to SNP analysis the ancestral allele of SNP M170 is associated with higher IQ scores in patients with ADHD (Table 7-7). Also, for SNP M9 the derived allele is associated with higher IQ scores in ADHD (Table 7-7). This points to haplogroups K, R1*, R1b3M405 and R1b3d being associated with higher IQ scores, because they are the ones to have this combination of alleles (see Table 1 in Appendix II). Since the derived allele of SNP M173 is associated with higher IQ scores, this limits the haplogroups that are associated with higher IQ scores to R1*, R1b3M405 and R1b3d. Finally, the derived allele of SNP M405 is associated with higher IQ scores, which means that haplogroup R1b3M405 is the only one with the appropriate combination of alleles. According to haplogroup analysis, haplogroup 3, which includes haplogroup R1b3M405, is associated with higher IQ scores (Figure 7-6). Hence, both SNP analysis and haplogroup analysis agree in pointing out that a haplogroup defined by a derived mutation of SNP M405 is associated with higher IQ scores.

There was also a significant association of SNP M170 and the number of DSM-IV ADHD symptoms (Table 7-5). However, none of the other Y chromosome SNPs were significant and the difference in the mean number of ADHD symptoms between the two alleles was very small. In addition, there was no evidence of association in the haplogroup analysis. For these reasons, we conclude that the data do not provide convincing support for an association with the number of DSM-IV ADHD symptoms.

In order to test for an association between Y chromosome variation and ADHD, a case-control association study was performed using nine Y chromosome markers. This is the first time that Y chromosome haplogroups and their effect have been investigated in ADHD. For this reason, even the fact that the case-control study showed that there was no evidence of an association of Y chromosome haplogroups with ADHD (Table 7-4) is of importance, since these are novel findings.

Our results indicated a possible modifying effect of Y chromosome variation on the phenotype of ADHD. Haplogroup 3 was associated with higher FSIQ ($p=0.019$) and PIQ ($p=0.011$) scores compared to haplogroup 1. The mean scores of all types of IQ increased from haplogroup 1 to 3, although not all the differences reached significance levels. Reading scores using the WORD Reading Test also increased from haplogroup 1 to 3 without reaching significance levels [$F(2,190)=0.62$, $p=0.54$, $r=0.08$]. Thus, there could be an effect of Y chromosome variation on cognitive functions. There can be two possible explanations for this association. One of them is that the functional polymorphism responsible for this association has appeared (in evolutionary terms) before the SNPs that define the associated haplogroup. Thus, all Y chromosomes belonging to these haplogroups would carry the causal variant. However, it is most likely that the functional polymorphism has appeared against a specific haplogroup background and for this reason, only certain individuals within this haplogroup will have it (Krausz et al. 2004). If the Y chromosome is truly associated with cognitive function, the classic notion that the Y chromosome is only important for reproductive functions would be challenged and its exclusion from genetic studies of brain disorders would be unjustifiable. This is also supported by animal studies linking the Y chromosome to aggression and parental behaviour (Guillot et al. 1995).

One of the strengths of this study is that the population used was of UK Caucasian origin. Participants were required to be of British Caucasian origin to the child's grandparents. The control individuals, the 1958 UK Birth Cohort, were again of the same origin. Sample homogeneity is important for any genetic study but it is vital for Y chromosome studies where the haplogroup distribution can vary significantly between countries leading to population stratification. The fact that there was no difference in the frequencies of Y chromosome haplogroups between the ADHD

cases and the controls suggests that it is valid for the 1958 UK Birth Cohort to be used as a control in genetic studies of the Y chromosome in the UK, since they are both of UK Caucasian origin.

Other advantages of this study are that the ascertainment of the sample was very careful using strict criteria for inclusion and providing a rich set of phenotypic measures to investigate. Finally, the Y chromosome markers that were genotyped were selected after very careful consideration and advice was sought from experts in the field due to the lack of information about the Y chromosome haplogroup distribution in the UK.

One of the most important issues that arise when interpreting these results is whether there is an association of Y chromosome haplogroups and IQ in the general population or whether this is only true for patients with ADHD. In order to test this alternative explanation, it would be necessary to test the same Y chromosome haplogroups for association with IQ in a non-clinical population of UK origin. A population cohort is the ideal sample to test the effect of Y chromosome on IQ. Although the control sample used in this study, the 1958 UK Birth Cohort, would be suitable, it was not possible to obtain IQ data. For this reason, another population cohort from the UK, the ALSPAC, was used and the results are discussed in Chapter 9.

Another issue is that different phenotypic measures were tested for association with Y chromosome SNPs and haplogroups and it is possible that this significant association is a result of multiple testing. However, the phenotypes that were studied were chosen *a priori* from the large number of phenotypic measures collected for our sample of ADHD patients due to evidence of association with the Y chromosome or because they were known to be sexually dimorphic. Since this study presents novel findings, replication of these findings in an independent sample is needed to provide strong supporting evidence. Using data from the ASLPAC sample, which had information on the presence of ADHD, we attempted to replicate our findings and this is discussed in Chapter 9. In addition, phenotypic analysis using SNPs was performed as a secondary analysis to confirm haplogroup analysis results and for this reason no multiple testing correction was performed.

It is possible that the association of Y chromosome variation and IQ in the ADHD sample is confounded by other factors or arises from population stratification effects. Social class can be one of these factors and it was tested for association with Y chromosome haplogroups. No evidence of association was found [χ^2 (4) = 7.1, p=0.13]. Another possible confounder would be the geographic origin in the UK that our individuals come from because Y chromosome haplogroups are highly geographically distributed and they can vary even within countries. The ADHD sample consisted of both English and Welsh individuals of Caucasian origin but no association with Y chromosome haplogroups was found [χ^2 (2) = 1.6, p=0.46]. However, our measure of geographic origin was limited by the fact that it was based on the postcode of the individuals at the time of assessment and this does not always correlate with the person's origin. Nevertheless, it was the only available measure of geographic location.

Another important issue is the sample size, as this can be one factor contributing to failure to detect other associations. It has become evident from recent genome-wide association studies (O'Donovan et al. 2008; Wolfs et al. 2009) that the effect sizes of common risk alleles in complex disorders are very small (OR<1.2) and as a consequence large sample sizes are required when multiple tests are performed. Although our study is the first to have genotyped such a large number of UK individuals for Y chromosome SNPs, the power to detect any association when the effect size of the risk allele is small (OR<1.2) is limited to 41% according to power calculations. This power is further compromised when the variant is rare. However, this study had >95% statistical power to detect effect sizes of 1.5 when the variant is common; thus it provides no evidence of the presence of risk alleles for ADHD of large effect sizes on the Y chromosome.

Despite the limitations of this study these results are interesting because they are the first systematic effort to assess the effect of Y chromosome on male susceptibility to ADHD. The unique properties of the Y chromosome and the problems they pose when studying it should not deter researchers from including it in genetic studies, since it is a male-only chromosome that can influence the phenotype of psychiatric disorders with a male preponderance. These results indicate an influence of Y chromosome variants on IQ in patients with ADHD but they need to be replicated.

Chapter 8 The role of Y chromosome in schizophrenia

8.1 Introduction

Schizophrenia is a debilitating psychiatric disorder which exhibits striking sex differences in age of onset and morbidity. Although the male/female prevalence ratio is moderately increased, males have an earlier age of onset that is associated with worse outcome and more impaired cognition. The presentation of the disorder is also different with men having more negative symptoms. A longitudinal study following patients with schizophrenia over a period of twenty years found that women consistently showed better functioning than men (Grossman et al. 2008). Sex differences in schizophrenia exist even in relation to brain morphology (see Chapter 1 on sex differences for more details).

As previously discussed in this thesis, one of the main differences between male and female genomes is the presence of Y chromosome genes in males. This could be a factor contributing to sex differences. However, the Y chromosome has not been previously studied in relation to schizophrenia. In this study, Y chromosome SNPs were genotyped in patients with schizophrenia and controls in order to determine if the two groups differ in terms of Y chromosome variation.

8.2 Aims of the study

The present study investigated the role of the Y chromosome variation in schizophrenia. Y chromosome haplogroups appropriate for UK populations were chosen as described in Chapter 7.

The hypothesis that the Y chromosome SNPs and haplogroups are associated with schizophrenia was tested by performing a case-control association study. Both allelic and haplogroup analyses were performed, although they provide similar information, in order to confirm any positive findings.

The hypothesis that Y chromosome variation is associated with modifying effects on schizophrenia by affecting clinical presentation or cognitive ability was tested within

the group of patients with schizophrenia after stratifying them according to their Y chromosome SNP alleles or haplogroup.

8.3 Materials and methods

8.3.1 Schizophrenia sample description

The schizophrenia sample used in this study consists of 313 unrelated British Caucasian men. The mean age at first psychiatric contact was 23.6 (SD 6.8) years and the mean age at ascertainment was 44.8 (SD 13.1) years. Both Multi Center Research Ethics Committee and Local Research Ethics Committee approval were obtained and subjects gave written informed consent to participate.

8.3.2 Schizophrenia sample assessment

All patients had a consensus diagnosis of schizophrenia according to DSM-IV criteria made by two independent raters following a semi-structured interview by trained psychiatrists or psychologists using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (Wing et al. 1990) and review of case records. Cases were screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition.

SAPS/SANS (Schedule for the Assessment of Positive/Negative Symptoms)

There are six subscales (2 for each of positive, negative and disorganised symptoms) assessing severity and frequency of symptoms ranging from 0 (symptom not present at all) to five (maximum). Positive symptom subscales assess hallucinations and delusions, negative symptoms subscales assess affective flattening and poverty of speech, disorganised symptoms subscales assess inappropriate affect and thought disorder. Each of the two subscales for the three categories of symptoms are combined to form the total number of positive, negative and disorganised symptoms (ranging from 0 to 10).

Global assessment scale (GAS)

The global assessment scale (GAS) was used to rate the subject's lowest level of functioning at their worst point (scale from 1 to 100, with 1 representing worst functioning and 100 superior functioning) (Endicott et al. 1976).

Course of illness

Subjects were categorised into five groups according to their course of illness with 1 corresponding to a single episode with good recovery; 2 corresponding to multiple episodes with good recovery between; 3 corresponding to multiple episodes with partial recovery between; 4 corresponding to continuous chronic illness and 5 corresponding to continuous chronic illness with deterioration. In order to reduce the number of categories and facilitate analysis, individuals were placed in 2 groups *a priori* according to whether they had continuous chronic illness or not; the first group was considered to have good outcome (categories 1-3) and the second one was considered to have poor outcome (categories 4-5).

Educational level

Information was also obtained regarding the highest level of education achieved. This was assessed by the self-declaration of the highest level of qualifications attained. Subjects with no qualification and those with qualifications obtained at school up to but excluding A level were categorised in the "low education" group while those with A levels and any further educational qualifications were placed in the "high education" group.

8.3.3 1958 UK birth cohort control sample

The control sample used in this study was the same as in the ADHD association study described in Chapter 7. It consists of 637 men drawn from the 1958 British Birth Cohort.

8.3.4 DNA samples and genotyping

DNA was obtained from venous blood or mouthwash sample for patients with schizophrenia. DNA samples were obtained from the 1958 UK Birth Cohort after request. Genotyping was performed as described in Chapter 7. For more information about quality control procedures see Appendix I.

8.3.5 Selection of SNPs and categorisation of individuals in haplogroups

Nine Y chromosome SNPs were selected, as described previously, in order to capture the major haplogroups present in the UK (Figure 8-1). Y chromosome SNPs were used to categorise each individual to only one haplogroup (see Chapter 7).

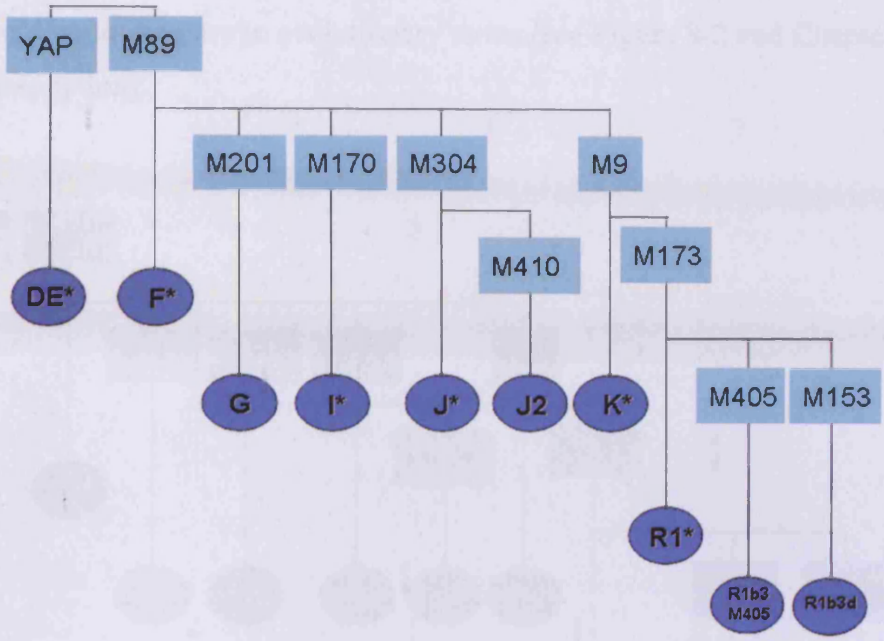

Figure 8-1. Y chromosome haplogroup tree showing the relationship between genotyped SNPs. Y chromosome haplogroups are represented by capital letters and SNPs start with M. There were no individuals belonging in haplogroup F*. Figure is the same as in Chapter 7 and is repeated to facilitate reading.

To test for an association of clinical characteristics of the schizophrenia sample and educational level with Y chromosome haplogroups, the 9 haplogroups were merged into 3 according to ancestry in evolutionary terms (see Figure 8-2 and Chapter 7 for detailed description).

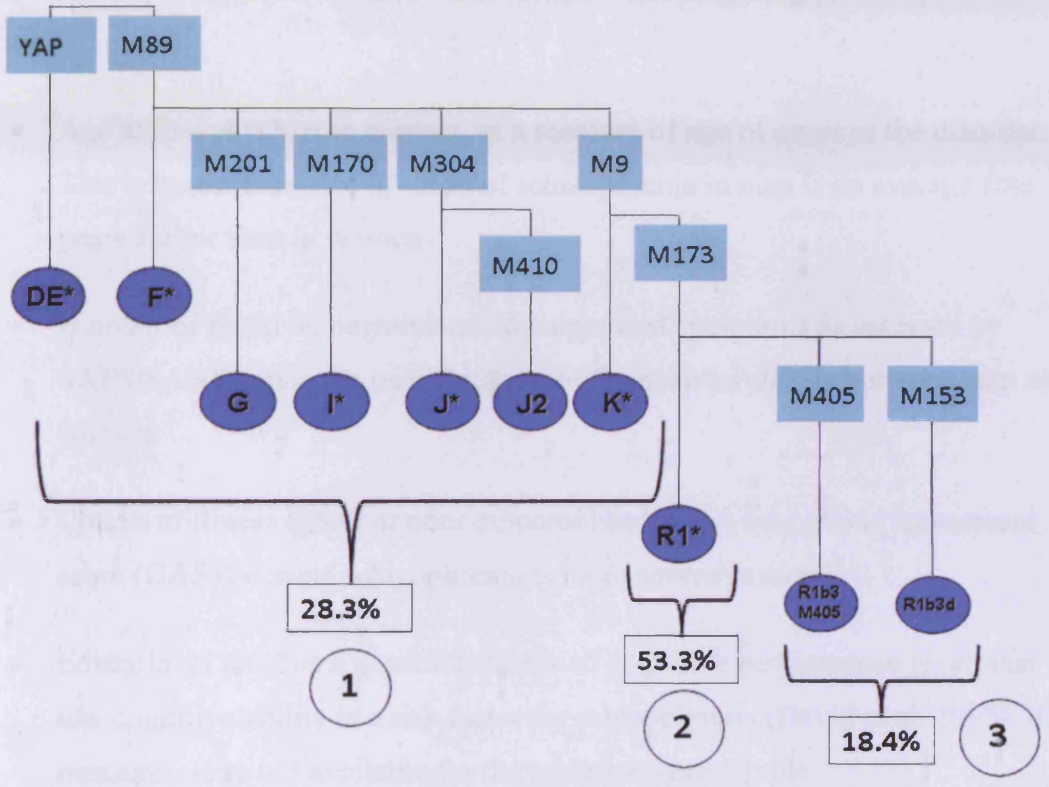

Figure 8-2. The Y chromosome haplogroups captured in this study were merged into three to facilitate phenotypic analysis. There were no individuals belonging in haplogroup F*. Percentages represent the frequencies of the three Y chromosome haplogroups in 4,950 UK men genotyped in total in this study. Figure is the same as in Chapter 7 and is repeated to facilitate reading.

8.3.6 Phenotypic analysis

Clinical characteristics and educational level were tested for association with Y chromosome haplogroups and SNP alleles within the sample of patients with schizophrenia. It was not possible to use the genotyped controls for this analysis because no phenotypic information was available. The following phenotypes were tested:

- Age at first psychiatric contact, as a measure of age of onset of the disorder. This is because the age of onset of schizophrenia in men is on average four years earlier than in women
- Number of positive, negative and disorganised symptoms as assessed by SAPS/SANS, since the manifestation of the disorder differs between men and women
- Course of illness (good or poor outcome) and worst ever global assessment score (GAS) because schizophrenia is more severe in men
- Educational level as a general measure of cognitive performance given that low cognitive ability is a risk factor for schizophrenia (David et al. 1997). IQ measures were not available for the schizophrenia sample

8.3.7 Statistical analysis

The Pearson's chi-square test [degree of freedom (df=1)] was used to test for allelic association and also for haplogroup analysis. The threshold of significance was set at $p \leq 0.05$. Correction for multiple testing would be required in case of significant associations. Then, a Bonferroni correction or permutation testing would be applied. With this threshold of significance, this study had >95% statistical power to detect effect sizes of 1.5 when the risk allele frequency was 0.3-0.4. For effect sizes of 1.3 the power was reduced to 72% and for effect sizes of 1.2 the power was 45%. For rarer alleles with frequencies of 0.1-0.2 the power was 92% for effect sizes of 1.5, 56% for effect sizes of 1.3 and 35% for effect sizes of 1.2. The prevalence of

schizophrenia was considered to be 1%. Statistical power calculations were performed using the Genetic Power Calculator (Purcell et al. 2003).

Clinical characteristics of the schizophrenia sample and educational level were tested for association with Y chromosome haplogroups. Normality of distribution for each of the dependent variables was assessed using measures of skewness and kurtosis (they should not exceed 1 and 3 respectively). The significance level of the Kolmogorov-Smirnov test was also taken into account.

Haplogroups were tested for association with all the previous phenotypic measures using ANOVA after collapsing them into three groups (Figure 8-2), based on evolutionary structure, to increase numbers in each group. When a positive association was detected, non-parametric tests were performed to confirm the results. Pearson chi-square test (df=2) was used to test for an association of categorical variables (course of illness and educational level) with haplogroups. The threshold of significance was again set at $p \leq 0.05$.

All variables were tested for association with single SNPs that had an adequate number of individuals in order to perform the analysis (MAF>15%) using independent samples t-test or the non-parametric Mann-Whitney U test according to whether or not data were normally distributed. For categorical variables (course of illness and educational level) Pearson chi-square test (df=1) was used. This was performed as a secondary analysis to confirm the results from the haplogroup analysis and further explore the mutation responsible for any association with specific Y chromosome haplogroups.

8.4 Results

8.4.1 Case-control association study

8.4.1.1 Allelic analysis

Nine Y chromosome SNPs were genotyped in the sample of 313 male patients with schizophrenia and 637 male control individuals from the 1958 Birth Cohort. Table 8-1 summarises these data and shows the frequencies of Y chromosome SNP alleles in

cases and controls. There was no significant difference between cases and controls (Table 8-1). It was not possible to perform statistical analysis for SNPs with very low frequencies.

Table 8-1. Allelic analysis for 9 Y chromosome SNPs

Y chromosome SNPs	MAF (CASE) %	MAF (CONTROL) %	p value	OR	95% CI
M9	27.2	27.9	0.84	0.97	0.72 - 1.31
M89	4.2	3.4	0.58	1.22	0.61 - 2.45
M153	0	0.2	0.49	0.00	n/a
M170	18.5	20.5	0.48	0.88	0.63 - 1.25
M173	27.8	28.5	0.82	0.97	0.72 - 1.3
M201	2.1	1.5	0.47	1.46	0.52 - 4.15
M304	2.9	2.3	0.61	1.24	0.54 - 2.87
M405	16.6	19.5	0.28	0.82	0.58 - 1.17
M410	1	1.9	0.29	0.51	0.14 - 1.83

MAF: Minor Allele Frequency; OR: Odds Ratio; CI: Confidence Interval; n/a: non-applicable

8.4.1.2 Haplogroup analysis

The frequencies of the nine Y chromosome haplogroups captured in this study (Figure 8-1) were compared between cases with schizophrenia and controls. There were no individuals belonging in haplogroup F*. Table 8-2 summarises the results. Pearson chi-square test (df=1) test showed no significant difference between cases and controls in terms of Y chromosome haplogroups (Table 8-2). It was not possible to perform statistical analysis in haplogroups with very low frequencies (J*, K* and R1b3d).

Table 8-2. Frequencies of Y chromosome haplogroups in 313 cases with schizophrenia, controls and 637 controls

Y chromosome haplogroup	Cases (%)	Controls (%)	p value
DE	4.2	3.3	0.5
G*	1.9	1.4	0.56
I*	18.5	20.7	0.43
J*	1.9	0.6	n/a
J2	1	1.7	0.36
K*	0.3	0.5	n/a
R1*	55.6	52.1	0.31
R1b3M405	16.6	19.5	0.29
R1b3d	0	0.2	n/a

n/a: non-applicable

8.4.2 Phenotypic analysis using haplogroups

To test for an association of the previous phenotypes with Y chromosome haplogroups, the 9 haplogroups were merged into 3 according to ancestry. As shown in the graphs, age at first psychiatric contact increased as haplogroups became more recent in evolutionary terms but the difference was not significant [$F(2,296)=0.44$, $p=0.65$, $r=0.054$] (Figure 8-3).

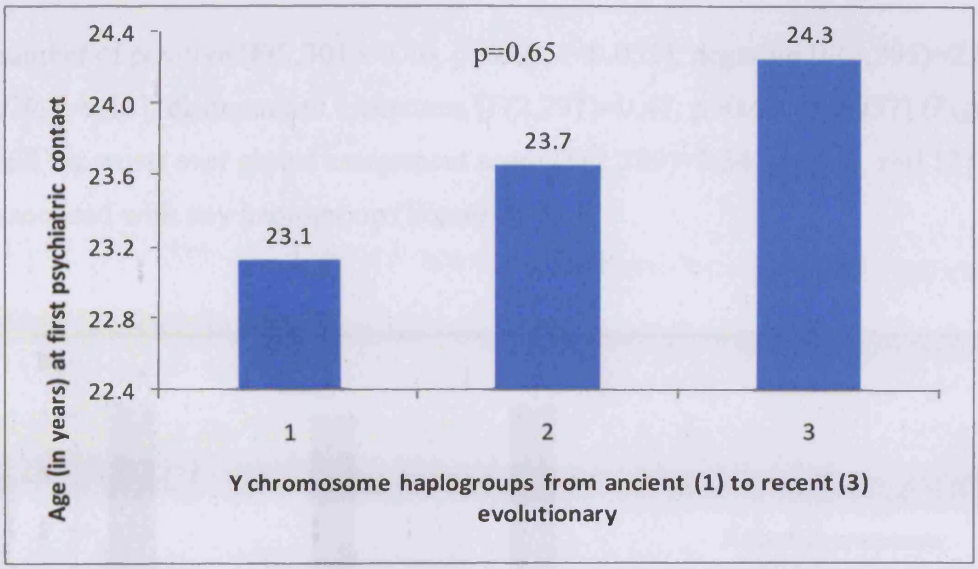

Figure 8-3. Age at first psychiatric contact in 3 Y chromosome haplogroups within the sample of patients with schizophrenia. . p value obtained from ANOVA comparing age at first psychiatric contact in 3 Y chromosome haplogroups is shown on the graph

The number of positive [F(2,301)=0.40, p=0.67, r=0.051], negative [F(2,295)=2.59, p=0.076, r=0.13], disorganised symptoms [F(2,291)=0.47, p=0.63, r=0.057] (Figure 8-4) and the worst ever global assessment score [F(2,289)=2.54, p=0.08, r=0.13] were not associated with any haplogroup (Figures 8-5).

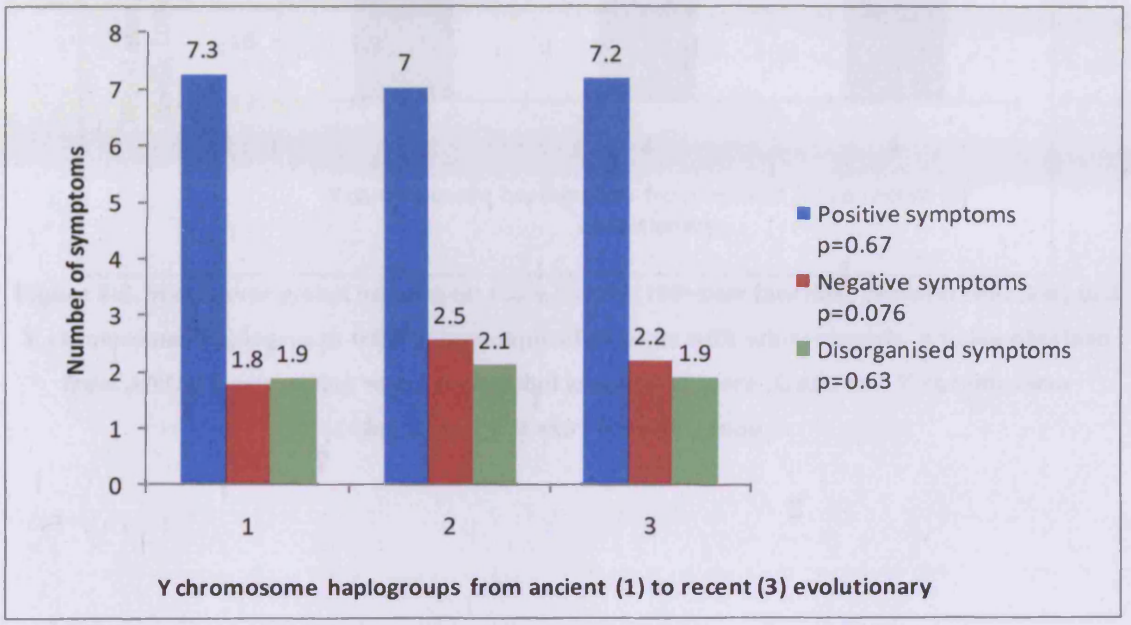

Figure 8-4. Number of positive, negative and disorganised symptoms in 3 Y chromosome haplogroups within the sample of patients with schizophrenia. p values obtained from ANOVA comparing each of the variables separately in 3 Y chromosome haplogroups are shown on the graph

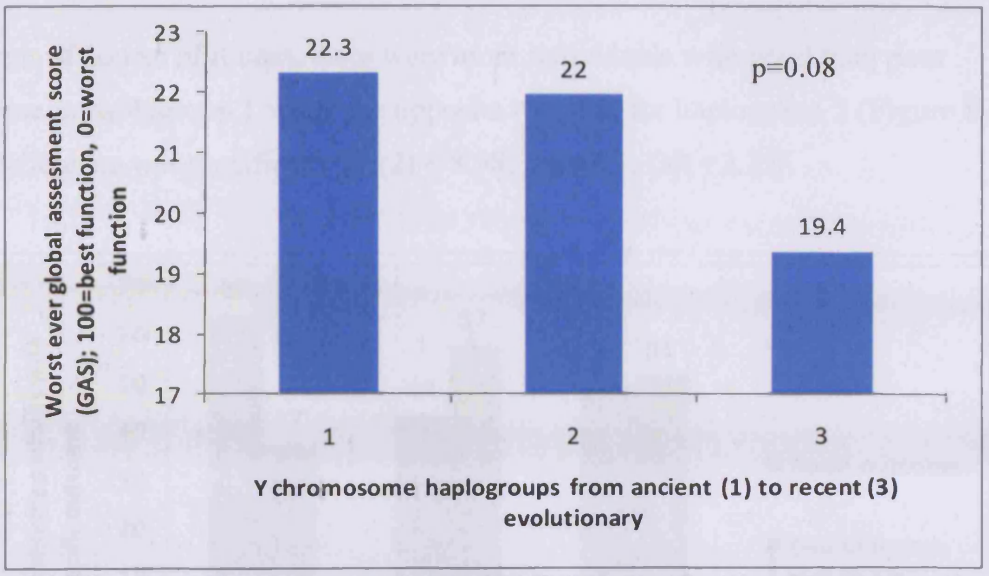

Figure 8-5. Worst ever global assessment score (GAS) (100=best function, 0=worst function) in 3 Y chromosome haplogroups within the sample of patients with schizophrenia. p value obtained from ANOVA comparing worst ever global assessment score (GAS) in 3 Y chromosome haplogroups is shown on the graph

In terms of course of illness, there were more individuals with good than poor outcome in haplogroup 1 while the opposite was true for haplogroup 2 (Figure 8-6). The difference was significant [χ^2 (2) = 8.98, p=0.011, OR =2.25].

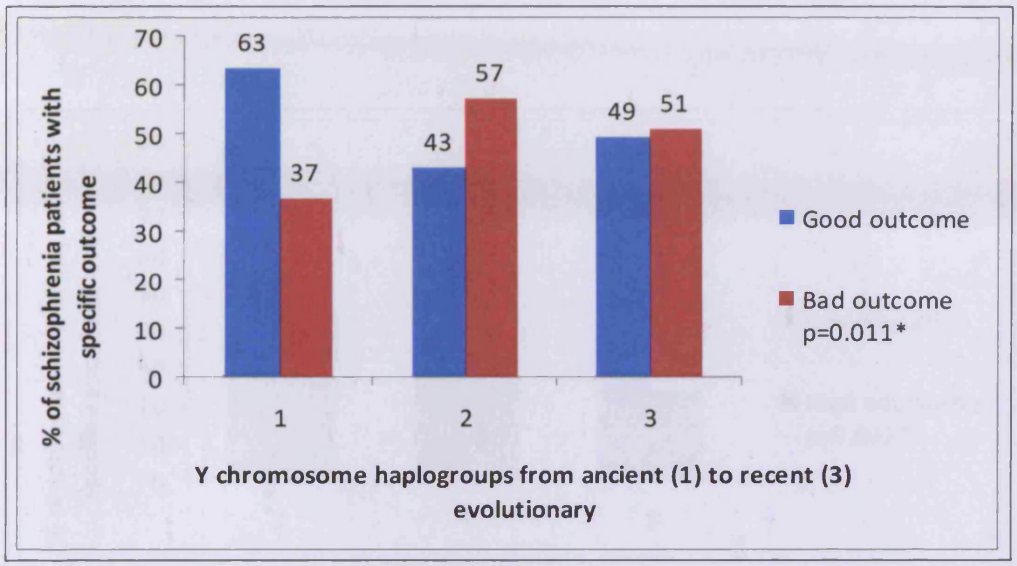

Figure 8-6. % of patients with schizophrenia in each of the three Y chromosome haplogroups that have good or poor outcome of illness. p value obtained from Pearson chi-square test comparing % of patients with good and poor outcome in 3 Y chromosome haplogroups is shown on the graph; *: significant at the $p \leq 0.05$ level

The educational level was very interesting with more individuals with high education than low education in haplogroup 1 and the opposite for haplogroup 3 (Figure 8-7). The difference was significant [χ^2 (2) = 6.35, p=0.042, OR =2.5]. Haplogroup 2 had approximately the same frequency of individuals for high and low education (Figure 8-7).

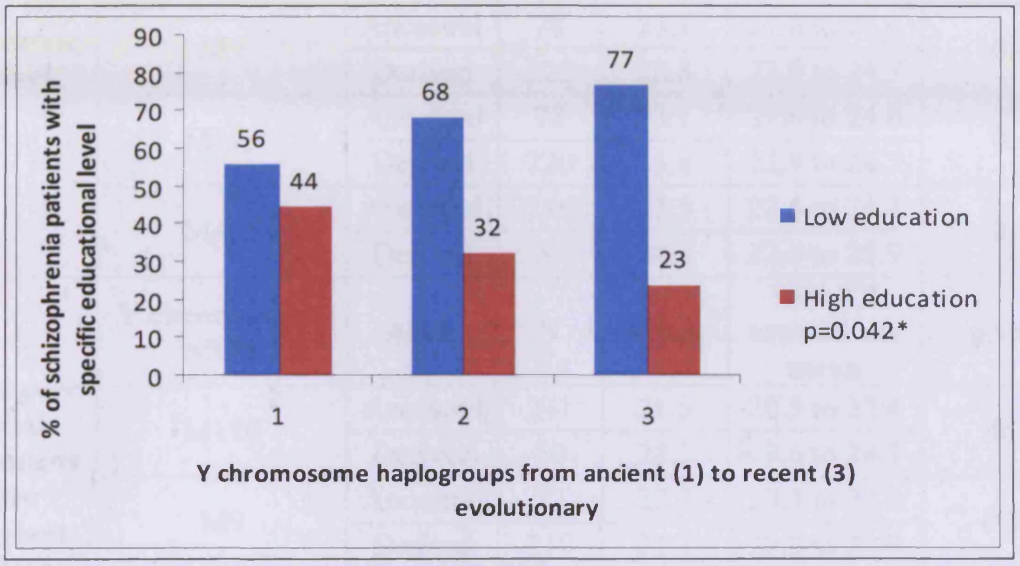

Figure 8-7. % of patients with schizophrenia in each of the three Y chromosome haplogroups that have low or high educational level (low education: no qualification or qualifications obtained at school up to but excluding A level, high education: A level or any further qualifications). p value obtained from Pearson chi-square test comparing % of patients with low and high education in 3 Y chromosome haplogroups is shown on the graph; *: significant at the p≤0.05 level

8.4.3 Phenotypic analysis using Y chromosome SNP alleles

Secondary analysis was performed by testing if Y chromosome SNP alleles were associated with certain phenotypes. Four Y chromosome SNPs (M170, M9, M173 and M405) were chosen because both their alleles were common enough (MAF>15%) to allow sufficient numbers of individuals in each group for testing. The number of individuals used for testing differs due to missing data for certain measures on some individuals.

Table 8-3. Testing for association of Y chromosome SNP alleles with age at first psychiatric contact and worst ever global assessment score (GAS) (100=best function, 0=worst function) within the sample of patients with schizophrenia - Mann Whitney U test

	Y chromosome SNPs	Allele	N	Mean	95% CI around the mean	p value
Age at first psychiatric contact	M170	Ancestral	245	24.0	23.1 to 24.8	0.053
		Derived	53	22.0	20.4 to 23.6	
	M9	Ancestral	78	23.1	21.6 to 24.6	0.44
		Derived	220	23.8	22.9 to 24.7	
	M173	Ancestral	78	23.1	21.6 to 24.6	0.44
		Derived	220	23.8	22.9 to 24.7	
	M405	Ancestral	246	23.5	22.6 to 24.4	0.17
		Derived	52	24.3	22.6 to 25.9	
Worst ever global assessment score (100=best function, 0=worst function)	**Y chromosome SNPs**	**Allele**	**N**	**Mean**	**95% CI around the mean**	**p value**
	M170	Ancestral	241	21.5	20.5 to 22.4	0.38
		Derived	50	22.2	19.6 to 24.7	
	M9	Ancestral	72	22.3	20.4 to 24.2	0.251
		Derived	219	21.3	20.3 to 22.4	
	M173	Ancestral	72	22.3	20.4 to 24.2	0.251
		Derived	219	21.3	20.3 to 22.4	
	M405	Ancestral	239	22.1	21.1 to 23.1	0.02*
		Derived	52	19.4	17.2 to 21.5	

CI: Confidence Interval;

*: significant at the $p \leq 0.05$ level;

Table 8-4. Testing for association of Y chromosome SNP alleles with the number of positive, negative and disorganised symptoms within the sample of patients with schizophrenia - Mann Whitney U test

	Y chromosome SNPs	Allele	N	Mean	95% CI around the mean	p value
Positive symptoms	M170	Ancestral	247	7.1	6.8 to 7.3	0.44
		Derived	56	7	6.9 to 7.8	
	M9	Ancestral	81	7.3	6.8 to 7.7	0.4
		Derived	222	7.1	6.8 to 7.3	
	M173	Ancestral	81	7.3	6.8 to 7.7	0.4
		Derived	222	7.1	6.8 to 7.3	
	M405	Ancestral	251	7.1	6.8 to 7.3	0.88
		Derived	52	7.2	6.7 to 7.7	
Negative symptoms	**Y chromosome SNPs**	**Allele**	**N**	**Mean**	**95% CI around the mean**	**p value**
	M170	Ancestral	242	2.5	2.1 to 2.8	0.09
		Derived	55	1.4	0.9 to 1.9	
	M9	Ancestral	79	1.8	1.3 to 1.6	0.067
		Derived	218	2.5	2.1 to 2.8	
	M173	Ancestral	79	1.8	1.3 to 1.6	0.067
		Derived	218	2.5	2.1 to 2.8	
	M405	Ancestral	245	2.3	2 to 2.6	0.68
		Derived	52	2.2	1.5 to 2.9	
Disorganised symptoms	**Y chromosome SNPs**	**Allele**	**N**	**Mean**	**95% CI around the mean**	**p value**
	M170	Ancestral	240	2.0	1.8 to 2.3	0.77
		Derived	53	1.9	1.4 to 2.4	
	M9	Ancestral	77	1.9	1.4 to 2.3	0.550
		Derived	216	2.0	1.8 to 2.3	
	M173	Ancestral	77	1.9	1.4 to 2.3	0.550
		Derived	216	2.0	1.8 to 2.3	
	M405	Ancestral	241	2.0	1.8 to 2.3	0.51
		Derived	52	1.9	1.3 to 2.5	

CI: Confidence Interval

Table 8-5. Testing for association of Y chromosome SNP alleles with the course of illness (good or poor outcome) within the sample of patients with schizophrenia – Pearson chi-square test (df=1)

Y chromosome SNP	Allele	N	Good outcome (%)	Poor outcome (%)	p value
M170	Ancestral	244	45.5	54.5	0.005**
	Derived	54	66.7	33.3	
M9	Ancestral	79	63.3	36.7	0.004**
	Derived	219	44.3	55.7	
M173	Ancestral	79	63.3	36.7	0.004**
	Derived	219	44.3	55.7	
M405	Ancestral	247	49.4	50.6	0.961
	Derived	51	49	51	

**: significant at the $p \leq 0.01$ level

Table 8-6. Testing for association of Y chromosome SNP alleles with educational level (low education: no qualification or qualifications obtained at school up to but excluding A level, high education: A level or any further qualifications) within the sample of patients with schizophrenia – Pearson chi-square test (df=1)

Y chromosome SNP	Allele	N	Low education (%)	High education (%)	p value
M170	Ancestral	227	66.5	33.5	0.675
	Derived	52	63.5	36.5	
M9	Ancestral	78	55.1	44.9	0.017*
	Derived	201	70.1	29.9	
M173	Ancestral	79	55.7	44.3	0.023*
	Derived	200	70	30	
M405	Ancestral	232	63.8	36.2	0.091
	Derived	47	76.6	23.4	

*: significant at the $p \leq 0.05$ level

A significant association was demonstrated between three of the SNPs tested and course of illness (Table 8-5) as shown by Pearson chi-square analysis. The ancestral allele of SNP M170 had an increased frequency of individuals with poor outcome while the derived allele had an increased frequency of individuals with good outcome. The ancestral alleles of SNPs M9 and M173 had increased frequencies of individuals with good outcome while their derived alleles had increased frequencies of individuals with poor outcome. The frequencies of individuals with good and poor outcomes were very similar in the two alleles of SNP M405.

A significant association was also demonstrated between two of the SNPs tested and educational level as shown by Pearson chi-square analysis (Table 8-6). The ancestral alleles of SNPs M9 and M173 had increased frequencies of individuals with high educational level while their derived alleles had increased frequencies of individuals with low educational levels. The same was true for SNP M405, although the difference did not reach significance.

There was also a significant association of SNP M405 and worst ever GAS (Table 8-3).

No association was shown for the age at first psychiatric contact (Table 8-3) and the number of positive, negative and disorganised symptoms (Table 8-4).

8.5 Discussion

In this study, the hypothesis that Y chromosome variation can have a direct effect on schizophrenia or a modifying effect by influencing clinical presentation or cognitive performance was tested. The results of this Chapter indicate that there is no main effect of Y chromosome variation on schizophrenia. However, Y chromosome variation may have a modifying effect on schizophrenia by influencing cognitive performance (as indexed by educational level) [χ^2 (2) = 6.35, p=0.042, OR =2.5] and course of illness [χ^2 (2) = 8.98, p=0.011, OR =2.25].

This is the first time that Y chromosome haplogroups and their effect have been investigated in schizophrenia. This was achieved by performing a case-control association study using nine Y chromosome markers. Although there was no evidence of an association of Y chromosome haplogroups with schizophrenia, the results are still important since they are novel.

Our results indicated a modifying effect of Y chromosome on clinical outcome. Haplogroup 1 had a significantly higher frequency of individuals with good outcome compared to haplogroup 2, which had a higher frequency of individuals with poor outcome. The OR was 2.25, which means that the odds of an individual having a good outcome were 2.25 times higher, if they belonged to haplogroup 1 than if they belonged in haplogroup 2. There was no difference in haplogroup 3.

There was also a difference between the Y chromosome haplogroups in terms of educational level. The frequencies of individuals with high and low educational levels were different in haplogroups 1 and 3. The OR was 2.5, which means that the odds of an individual having high education were 2.5 times higher, if they belonged to haplogroup 1 than if they belonged in haplogroup 3. Thus, there could be an effect of Y chromosome on our proxy for cognitive ability.

Phenotypic analysis was performed using both haplogroups and SNP alleles. These two types of analysis should yield the same pattern of results, since all the individuals were categorised in the Y chromosome haplogroups according to the Y chromosome SNP alleles that they possessed. The reason that both analyses were performed was

because the aim was to test if any specific Y chromosome haplogroup had a modifying effect on ADHD and also to test if this effect was linked to a specific mutation on the Y chromosome. This is indeed observed for clinical outcome and educational level, where both SNP analyses (Tables 8-5 and 8-6) and haplogroup analysis (Figures 8-6 and 8-7) provide significant results. There was also a significant association of SNP M405 and worst ever GAS (Table 8-3). However, none of the other Y chromosome SNPs were significant and the difference in worst ever GAS between the two alleles was very small. In addition, there was no evidence of association in the haplogroup analysis. For these reasons, we conclude that the data do not provide evidence for an association with worst ever GAS.

The pattern of the results in the schizophrenia sample is interesting. Haplogroup 1 seems to be associated with good disease outcome and high education. These results seem to be contradictory to the findings in the ADHD sample, where haplogroup 3 was associated with higher IQ scores compared to haplogroup 1. However, in the ADHD sample, IQ scores were available, which provide a more direct estimation of cognitive function than educational level. Educational level is not always correlated with IQ. Especially in a disorder such as schizophrenia it is very difficult to estimate the educational level that the patients would have achieved if they had not been affected by schizophrenia. Even when the disease is not very severe, the age of onset (especially in men who have an earlier age of onset than women) could coincide with an important period for education in an individual's life. Thus, the disruption that is caused can have devastating effects on the qualifications that the individual can obtain. Furthermore, cognitive impairment is common in schizophrenia and it persists, even when there are no other symptoms present. Thus, it could also affect the educational level that patients with schizophrenia can achieve.

Another explanation for the different pattern of results in ADHD and schizophrenia is that they are two aetiologically different disorders and the mechanism by which the Y chromosome operates to modify the phenotype could be different in these two disorders. This is likely because both disorders are complex with very poorly understood neuropathology. The reason that they were both studied in this study is that they are both neurodevelopmental disorders with a male preponderance. It is not implied at any point that they are aetiologically similar. It is also likely that Y

chromosome variants have pleiotropic effects. Finally, one or even both sets of findings may be a false positive.

There are some important advantages in this study. Firstly, both the case and the control sample were of British Caucasian origin. The Y chromosome markers genotyped were chosen after careful consideration to represent UK populations. The pattern of haplogroup distribution in patients and controls samples was as expected for a UK population and it was comparable with the ADHD samples discussed in Chapter 7. The ascertainment of the patients with schizophrenia and the collection of phenotypic measures were very careful and strict.

Apart from the fact that educational level was used as a proxy for cognitive ability in schizophrenia, there are other limitations in this study. Although the phenotypes that were studied were chosen *a priori* from the large number of collected phenotypic measures because they were known to be sexually dimorphic, it is possible that the significant findings are the result of multiple testing. Since this study presents novel findings and there have been no previous studies on the Y chromosome in schizophrenia, the need to correct for multiple testing for the number of phenotypic measures tested was considered less of a priority compared to investigating the role of Y chromosome in male susceptibility to schizophrenia. Replication of these findings in an independent sample is required to provide supporting evidence. In addition, phenotypic analysis using SNPs was performed as a secondary analysis to confirm haplogroup analysis results and for this reason no multiple testing correction was performed.

Another important issue is the limited power to detect associations with risk alleles of small effect sizes. It is likely that complex disorders, like schizophrenia, are caused by a large number of risk variants with small effects but detecting them requires large sample sizes. The power of this study to detect an association with an OR<1.2 was limited to 45% provided that the risk allele was relatively common according to power calculations. However, for variants of large effect sizes (similar to the ones detected for clinical outcome and educational level) the power was increased to >99%.

It is also possible that the significant associations are confounded by other factors that have not been measured. Social class and location of the patients could be some of these factors but we were not able to obtain information about these factors for the patients with schizophrenia.

This is the first study to assess the effect of Y chromosome on male susceptibility to schizophrenia. Our results indicate a possible influence of Y chromosome variants on the course of the illness and the educational level achieved in patients with schizophrenia but they would need to be replicated in order to provide solid evidence. IQ scores would be more suitable to assess the effect of Y chromosome on cognitive performance in schizophrenia than the educational level which might be severely compromised by the symptoms of the disorder.

Chapter 9 Investigating the role of Y chromosome in cognitive ability using a population cohort

9.1 Aims of the study

The aims of the present study were to investigate the role of the Y chromosome variation in cognitive ability using a population cohort and seek replication of our previous results about Y chromosome haplogroups influencing cognitive ability in ADHD.

In Chapter 7, an association between Y chromosome haplogroups and performance and full scale IQ was found in individuals with ADHD. Since this analysis was performed within the sample of patients with ADHD, it was not possible to delineate whether Y chromosome variation is associated with IQ only in individuals with ADHD or whether this is true for non-affected populations as well. In addition, it was desirable to perform phenotypic analysis using the Y chromosome haplogroups in an independent sample of patients with ADHD in order to try and replicate our significant findings.

IQ data from the control sample used to perform the case-control association studies on ADHD and schizophrenia, the 1958 UK Birth Cohort, could not be provided for the genotyped individuals. For this reason, we decided to use data from the AVON Longitudinal Study of Parents and Children (ALSPAC) (Golding 2004). Two Y chromosome SNPs were genotyped in 3,749 male individuals from the ALSPAC study and the genotypes were used to categorise each individual to one of the three haplogroups used in the previous phenotypic analysis for ADHD (Figure 9-1). The haplogroups were first tested for association with full scale (FSIQ), performance (PIQ) and verbal IQ (VIQ) using all available genotyped individuals in order to investigate the effect of Y chromosome haplogroups on IQ in a general population. Then, the same analysis was performed for those with an ADHD diagnosis to seek replication for our previous findings.

9.2 Materials and methods

9.2.1 ALSPAC sample and data description

The Avon Longitudinal Study of Parents and Children (Golding 2004) is a general population cohort enrolling pregnant women resident in Avon, South West England, with delivery dates between December 1992 and April 1999. The cohort consisted of 14,062 live births from 14,541 enrolled pregnant women with 13,985 surviving offspring at twelve months. Questionnaires were sent at regular intervals during pregnancy and following childbirth and biological samples were taken from parents and children including blood samples from which DNA was extracted. Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the Local Research Ethics Committees.

From age 7, all children were invited for an annual assessment of physical and psychological development, and up to 8,000 children continue to attend each year. 10% of the sample, known as Children in Focus, attended clinics at the University of Bristol at various intervals. The present study concerns cognitive assessments undertaken at age 8. IQ was assessed at age 8 using the WISC-III for the children attending the Children in Focus clinics. DSM-IV ADHD symptoms were assessed at age 7.5 years from parent reports using the Development and Well-being Assessment (DAWBA) (Goodman et al. 2000). The DAWBA is a package of questionnaires, interviews, and rating techniques designed to generate ICD-10 and DSM-IV psychiatric diagnoses on 5-16 year-olds. Diagnosis was also confirmed by a clinician.

To increase sample homogeneity, only children of Caucasian ethnicity (95% of the cohort) were included in our study.

The total sample used for analysis consists of 3,749 male individuals. The DAWBA at 7.5 years was completed for 74% of those individuals. 90 (3.3%) individuals met criteria for DSM-IV ADHD diagnosis. The mean full scale IQ score was 105.2 (SD 16.7, range 49-151). 2,491 individuals (66%) completed the WISC-III at the Children in Focus clinics. In order to make the ALSPAC ADHD cases as comparable as possible with the Cardiff ones, children with IQ<70 and/or autism were excluded. 63

children from ALSPAC had an ADHD diagnosis, full IQ data and met the inclusion criteria.

9.2.2 Selection of Y chromosome SNPs

In Chapter 7, I have described how the initial nine haplogroups were merged into three haplogroups according to ancestry in order to facilitate statistical analysis. Three SNPs defined these three haplogroups (M89, M173 and M405) (Figure 9-1).

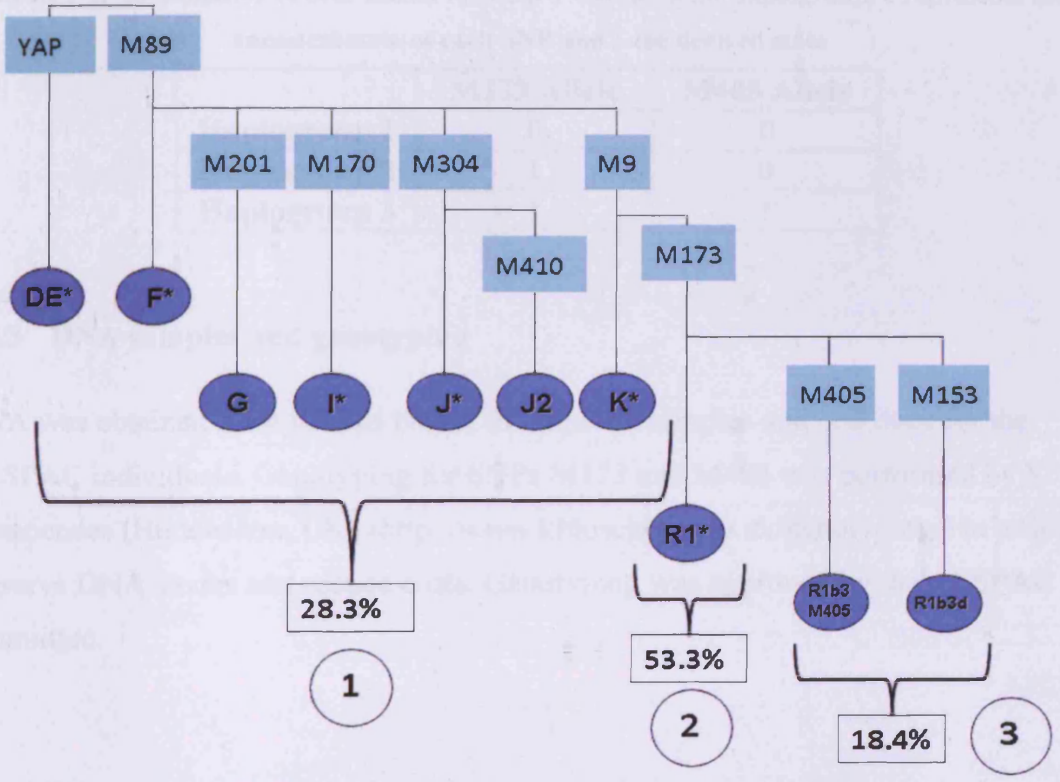

Figure 9-1. The Y chromosome haplogroups captured in this study were merged into three to facilitate phenotypic analysis. There were no individuals belonging in haplogroup F*. Percentages represent the frequencies of the three Y chromosome haplogroups in 4,950 UK men genotyped in total in this study. Figure is the same as in Chapter 7 and is repeated to facilitate reading.

In order to categorise the ALSPAC individuals into these three haplogroups, two SNPs were genotyped, M173 and M405. M173 defines haplogroup 2 and M405 defines haplogroup 3. Each individual was categorised in only one haplogroup according to whether they possessed the ancestral (0) or derived (1) allele of these two SNPs (Table 9-1). There was no need to genotype SNP M89 because if individuals possessed the ancestral alleles for these two SNPs, they would definitely belong to haplogroup 1.

Table 9-1. Combination of SNP alleles for each Y chromosome haplogroup. 0 represents the ancestral state of each SNP and 1 the derived state

	M173 Allele	M405 Allele
Haplogroup 1	0	0
Haplogroup 2	1	0
Haplogroup 3	1	1

9.2.3 DNA samples and genotyping

DNA was obtained from venous blood, mouthwash samples and cell lines for the ALSPAC individuals. Genotyping for SNPs M173 and M405 was performed by K Biosciences (Hoddesdon, UK) (http://www.kbioscience.co.uk/genotyping/) in order to preserve DNA stocks and reduce costs. Genotyping was approved by the ALSPAC committee.

9.2.4 Statistical analysis

Y chromosome haplogroups were tested for association with PIQ, VIQ and FSIQ using ANOVA. The threshold of significance was set at $p \leq 0.05$. With this threshold of significance, this study had >95% statistical power to detect effect sizes of 1.5 when the risk allele frequency was 0.3-0.4. For effect sizes of 1.3 the power was reduced to 70% and for effect sizes of 1.2 the power was 41%. For rarer alleles with frequencies of 0.1-0.2 the power was 90% for effect sizes of 1.5, 54% for effect sizes of 1.3 and 30% for effect sizes of 1.2. The prevalence of ADHD was considered to be 4% according to Cuffe et al. (2005). Statistical power calculations were performed

using the Genetic Power Calculator (Purcell et al. 2003). Normality of distribution for each of the dependent variables was assessed using measures of skewness and kurtosis (they should not exceed 1 and 3 respectively). The significance level of the Kolmogorov-Smirnov test was also taken into account.

9.3 Results

9.3.1 Frequencies of Y chromosome haplogroups in ALSPAC

The frequencies of Y chromosome haplogroups in all the genotyped individuals from ALSPAC (n=3,749) are shown in Table 9-2.

Table 9-2. Y chromosome haplogroup distribution in all individuals from ASLPAC

Y chromosome haplogroup	Frequency (%) in total ALSPAC sample
1	27.1
2	49.5
3	23.4

9.3.2 Testing for association of Y chromosome haplogroups with IQ in a general population cohort

The full sample of 3,749 male individuals from ALSPAC was stratified according to the Y chromosome haplogroup they belonged to. Due to incomplete data, 2,495 individuals had PIQ scores, 2,491 individuals had VIQ scores and 2,482 individuals had FSIQ scores.

As shown in the graph, PIQ, VIQ and FSIQ scores (Figure 9-2) do not vary in the three haplogroups. Performing ANOVA showed that there is no significant difference between the haplogroups in terms of FSIQ [F(2,2479)=2.52, p=0.081, r =0.045], PIQ [F(2,2492)=1.36, p=0.26, r =0.039] and VIQ [F(2,2488)=2.13, p=0.12, r =0.037].

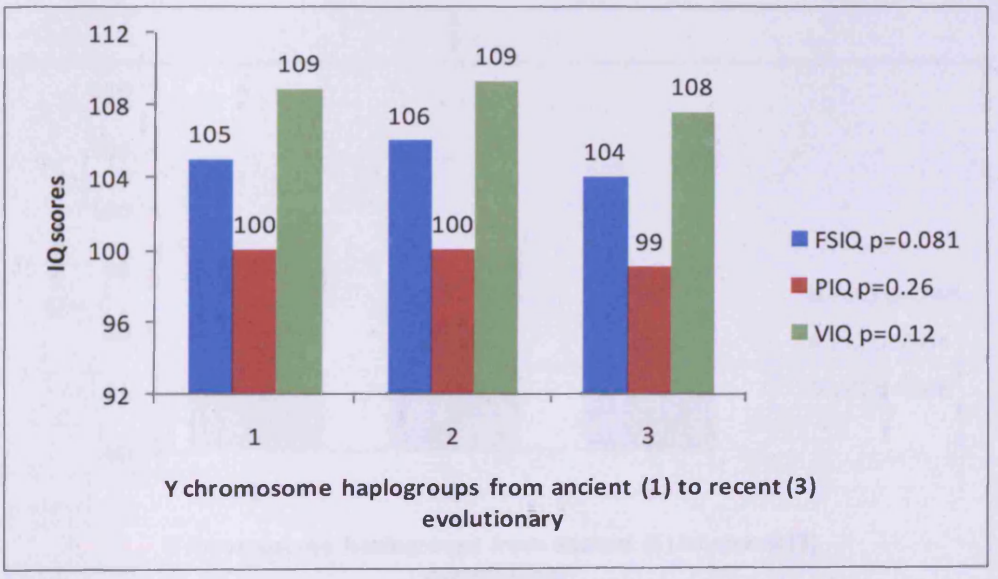

Figure 9-2. FSIQ, PIQ and VIQ scores in 3 Y chromosome haplogroups within all available individuals from ALSPAC population cohort. p values obtained from ANOVA comparing each of the variables separately in 3 Y chromosome haplogroups are shown on the graph

9.3.3 Testing for association Y chromosome haplogroups with IQ in ALSPAC ADHD patients

90 males from ALSPAC met criteria for DSM-IV ADHD diagnosis. 21 children did not complete the WISC-III while 6 children had an IQ score below 70. These 63 individuals were stratified according to the Y chromosome haplogroup they belonged to. Then, PIQ, VIQ and FSIQ scores were compared in these three groups using ANOVA.

As shown in the graph, scores of PIQ, VIQ and FSIQ, (Figure 9-3) do not vary in the three haplogroups. Performing ANOVA showed that there is no significant difference between the haplogroups in terms of FSIQ [F(2,60)=0.54, p=0.59, r =0.12], PIQ [F(2,60)=0.16, p=0.85, r =0.1] and VIQ [F(2,60)=0.34, p=0.71, r =0.1].

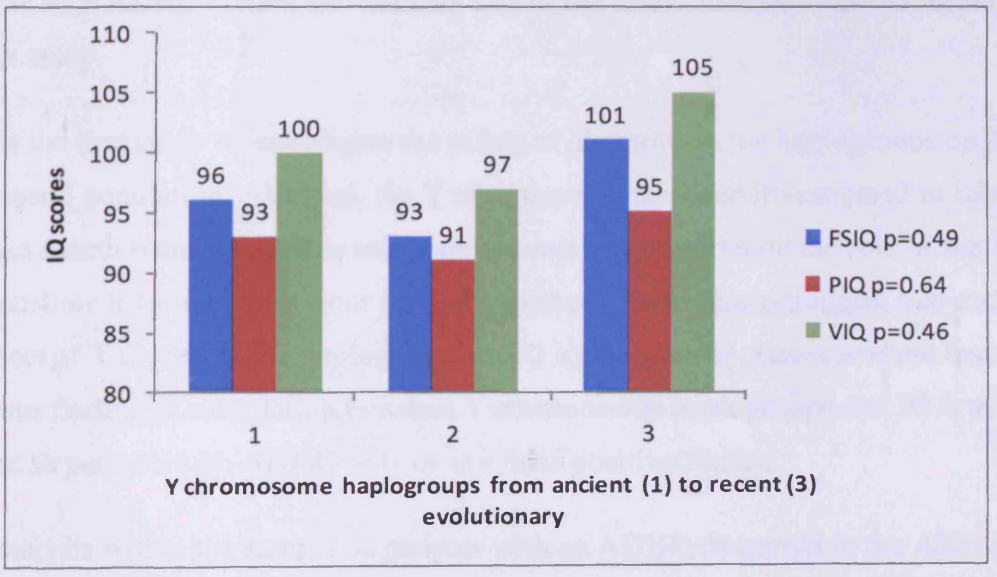

Figure 9-3. FSIQ, PIQ and VIQ scores in 3 Y chromosome haplogroups within 63 individuals from ALSPAC that meet diagnosis for ADHD. p values obtained from ANOVA comparing each of the variables separately in 3 Y chromosome haplogroups are shown on the graph

9.4 Discussion

The aims of this study were to investigate the association between Y chromosome haplogroups and IQ in a non-clinical population and to attempt to replicate the finding that Y chromosome haplogroup 3 is associated with higher FSIQ and PIQ scores compared to Y chromosome haplogroup 1 in an independent sample of patients with ADHD. Our findings indicate that there is no association between Y chromosome haplogroups and IQ in the normal population (FSIQ [$F(2,2479)=2.52$, $p=0.081$, $r=0.045$], PIQ [$F(2,2492)=1.36$, $p=0.26$, $r=0.039$] and VIQ [$F(2,2488)=2.13$, $p=0.12$, $r=0.037$]). Although there was no significant difference between the Y chromosome haplogroups in terms of IQ in the sample of ADHD patients from ALSPAC (FSIQ [$F(2,60)=0.54$, $p=0.59$, $r=0.12$], PIQ [$F(2,60)=0.16$, $p=0.85$, $r=0.1$] and VIQ [$F(2,60)=0.34$, $p=0.71$, $r=0.1$]), an effect of Y chromosome haplogroups on IQ in patients with ADHD cannot be excluded due to the small number of available patients for the study.

This is the first study to investigate the effect of Y chromosome haplogroups on IQ in the general population. Although the Y chromosome has been investigated in relation to brain function and behaviour using animal models, research on the role of the Y chromosome in human behaviour has been limited. These results suggest that there is no effect of Y chromosome haplogroups on IQ in the general population and that our previous finding of association between Y chromosome haplogroups and IQ is at best limited to patients with ADHD only or is a false positive finding.

Our analysis within the sample of patients with an ADHD diagnosis in the ALSAC sample showed no significant difference in IQ scores in the different Y chromosome haplogroups. However, the number of individuals that could be included in the study was very small. Only 63 males in ALSPAC met criteria for ADHD diagnosis while satisfying our inclusion criteria and it is likely that this sample was underpowered to detect an association. Assuming an effect size of 1.2-1.3 (similar to the one that was achieved for the Cardiff ADHD cases), the power to detect an association with such a small sample was limited to <20%. In order to have >90% power to detect this effect, a large independent sample of UK ADHD cases (>800) with available data on IQ would be required.

An important issue in this attempt to replicate our results is whether the two samples of patients with ADHD are comparable. In an effort to make the samples as comparable as possible the same exclusion criteria as for the Cardiff ADHD cases were applied to ALSPAC as well. Children with IQ<70 and/or autism were excluded. However, it is interesting that the mean FSIQ score in ADHD cases in ALSPAC was 105 compared to a mean FSIQ score of 89 for the Cardiff sample. The reasons for this are likely to reflect differences in ascertainment and assessment.

In the ALSPAC sample there were only 90 male individuals that met DSM-IV criteria for ADHD diagnosis. Of those, 21 individuals (22%) failed to complete the WISC-III at the Children in Focus Clinics. It is likely that people who failed to complete the WISC-III are those who had more attention problems or found it more difficult to do so. For these reasons, it is possible that people with more severe ADHD and lower IQ scores failed to complete the tests and thus were not included in the study.

Another reason for the difference in the mean IQ scores between the Cardiff and the ALSPAC ADHD samples is that the Cardiff cases consist of children referred to district child and adolescent psychiatry and paediatric clinics while the ALSPAC ADHD sample is part of a general population cohort, so it can include children who have not yet received ADHD diagnosis by a health professional. In addition, children with ADHD and low IQ (but still above 70) are more likely to come to the attention of psychiatric services compared to children with ADHD and higher IQ, where the impairment caused by the disorder can be masked by the IQ. Thus, it can be argued that the Cardiff ADHD sample includes more severe ADHD cases that are selected for low IQ.

To conclude, there is the possibility of selection attrition where people with low IQ scores and ADHD are less likely to attend the Children in Focus Clinic and/or fully complete the WISC-III but are more likely to come to the attention of the Cardiff research team.

The main advantage of this study is that a large general population sample of UK origin was used. This was essential in ruling out an association of Y chromosome haplogroups with IQ in the general population. It is also important that we attempted to replicate our results in a non-clinical sample of patients with ADHD applying the

same exclusion criteria to make the samples as comparable as possible. Although the results did not reach statistical significance, the number of ADHD cases in ALSPAC is small and a large sample of cases with ADHD and IQ data would be required to increase confidence in our results.

Chapter 10 Testing for an indirect effect of Y chromosome variation on ADHD and schizophrenia – Interaction with Tyrosine Hydroxylase

10.1 Introduction

10.1.1 SRY (Sex Determining Region of chromosome Y)

One of the most important genes on the Y chromosome is the *SRY* gene (Sex Determining Region of chromosome Y). The *SRY* gene, apart from being solely responsible for male sex differentiation, has been implicated in brain function. It is transcribed in the hypothalamus and the frontal and temporal cortex of the adult male human brain (Lahr et al. 1995). The expression of SRY shows developmental changes with circular, untranslatable transcripts seen prenatally and linear, translatable transcripts postnatally (Figure 10-1) (Mayer et al. 2000). However, due to the very low levels of mRNA there is only one report of SRY protein detected in human embryonic testes and none for any other species or tissue (Poulat et al. 1995).

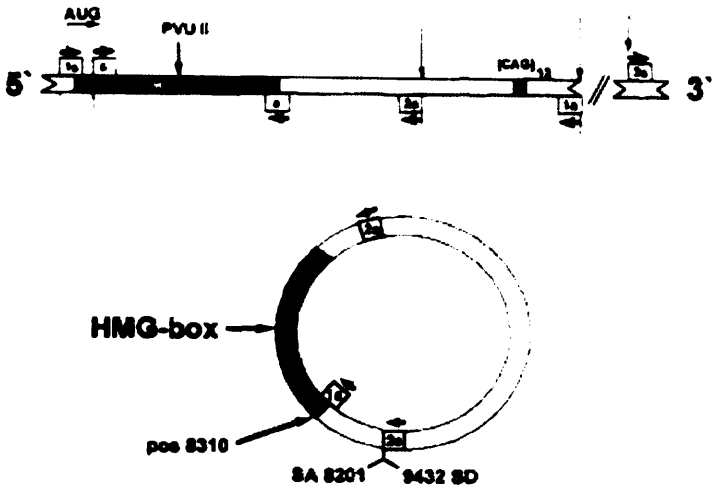

Figure 10-1. Different developmental profile of SRY. Linear transcripts that are translated postnatally and circular transcripts that are not translated prenatally. Taken from Mayer et al. (2000)

10.1.2 Tyrosine Hydroxylase (*TH*) function

One way that SRY can affect brain function is by interacting with autosomal genes which are known to be expressed in brain. One such hypothesised interaction is with the Tyrosine Hydroxylase (*TH*) gene (see later). The TH gene codes for an enzyme that catalyses the conversion of tyrosine to dihydroxyphenylalanine (L-DOPA) (Figure 10-2). This initial step is the rate-limiting step in the production of catecholamines. Dopamine is the next catecholamine to be produced followed by norepinephrine and epinephrine (Figure 10-2) (Naoi and Parvez 1993). TH is an oxygenase and is found in the cytosol of all cells containing catecholamines.

The dopaminergic pathway has been implicated in a number of neuropsychiatric disorders with higher male prevalence, such as ADHD, Parkinson's and schizophrenia as well as cardiovascular disease, which is also more prevalent in males. Specifically, stimulant medication used in ADHD acts by increasing the availability of extracellular dopamine (Volkow and Swanson 2003). Further evidence about the involvement of dopamine in ADHD comes from nueroimaging studies (Spencer et al. 2005) and animal studies, including that of the *DAT* knockout mouse (Gainetdinov 2007). Dopamine is thought to be involved in schizophrenia according to the hypothesis that the disorder is caused by increased dopaminergic transmission

(Pearlson 2000). Antipsychotic drugs work by blocking dopamine receptors and dopamine agonists, such as amphetamines, can provoke psychotic symptoms (Pearlson 2000).

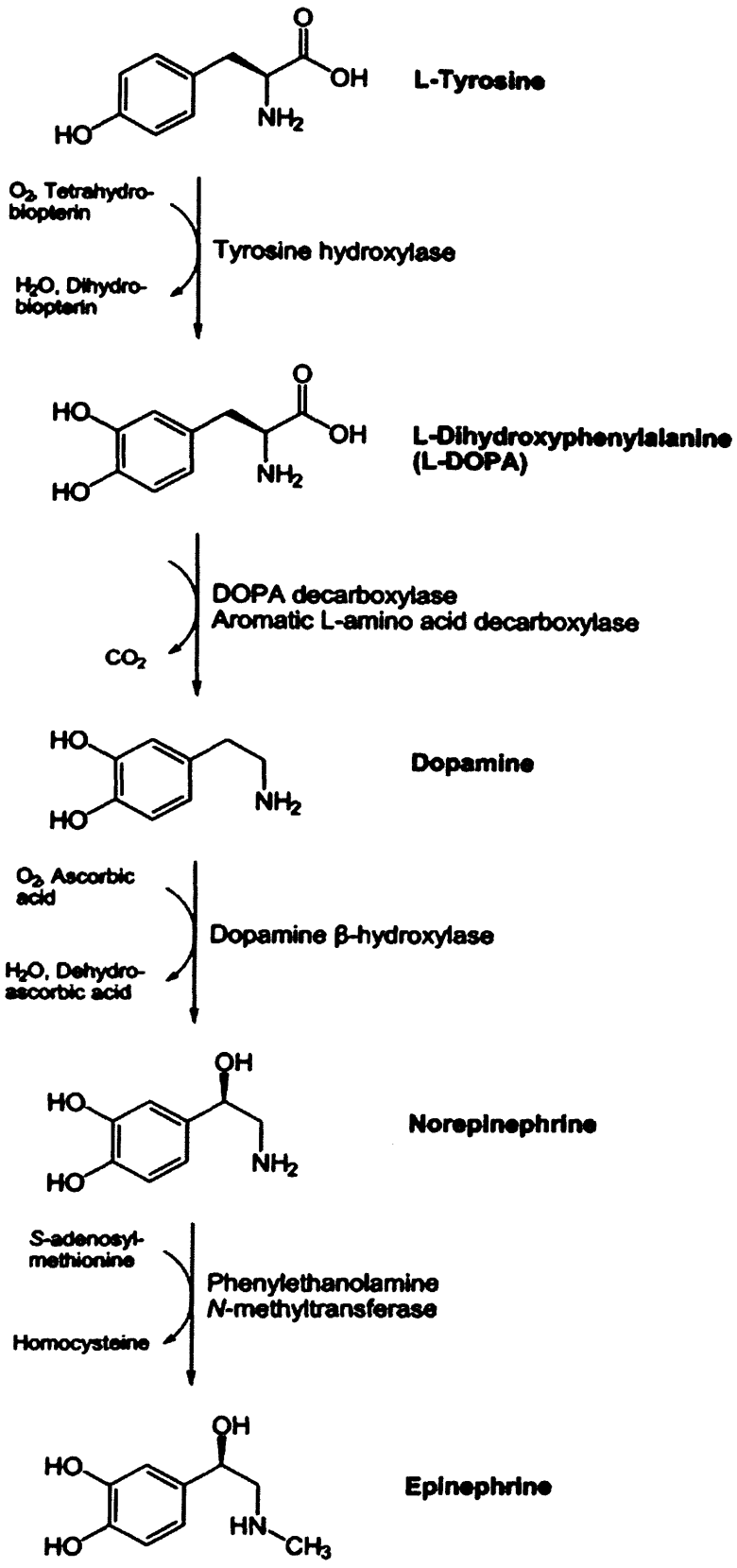

Figure 10-2. Catecholamine synthesis (Edinformatics.com 1999)
(http://www.edinformatics.com/interactive_molecules/info/adrenaline.htm)

10.1.3 *TH* gene

The TH gene, located at 11p15.5, has 14 exons and spans approximately 8 kb. There are four different *TH* isoforms resulting from alternative splicing of a single gene. All isoforms are expressed in the human brain and specifically, the substantia nigra and locus coeruleus. Isoform 1 has the highest TH activity of all other isoforms (Nagatsu and Ichinose 1991).

10.1.4 TH associations

The most studied polymorphism in the *TH* gene is an intronic polymorphic TCAT repeat (HUMTH01). This microsatellite is functional and has been found to regulate transcription (Albanese et al. 2001). Mutations in the *TH* gene have been associated with autosomal recessive Segawa Syndrome, a form of infantile parkinsonism (Verbeek et al. 2007).

The chromosomal region 11p harbours another two genes, *IGF2* (Insulin-like growth factor 2) and *INS* (insulin), in close proximity with *TH*. Due to the association of these genes with obesity, diabetes and hypertension, *TH* has been studied in relation to these conditions as well (Rodriguez et al. 2004).

Due to the possible involvement of dopamine in psychiatric disorders, *TH* has been studied in schizophrenia, alcohol (Dahmen et al. 2005) and nicotine dependence (Rodriguez et al. 2006), bipolar disorder (Byerley et al. 1992), autism and ADHD (Comings et al. 1995). In schizophrenia, association studies have focused on the tetranucleotide repeat in the first intron of the gene (Meloni et al. 2002). Positive (Jacewicz et al. 2006; Kurumaji et al. 2001; Pae et al. 2003) and negative associations (Andreou et al. 2009; Ishiguro et al. 1998; Jonsson et al. 1998; Ota et al. 2001) of *TH* variants with schizophrenia have been reported, thus no safe conclusions can be drawn. A meta-analysis by the SzGene database showed no evidence of association with the SNP rs6356 in *TH* including four case-control studies performed in Asian populations (Schizophrenia Research Forum 2005). Studies on the HUMTH01 tetranucleotide repeat were not included in the meta-analysis. In ADHD, only three studies have examined the association between the HUMTH01 polymorphism on *TH*

and ADHD and they have all been negative (Barr et al. 2000; Comings et al. 1995; Payton et al. 2001).

Although this intronic microsatellite has been extensively studied in relation to psychiatric disorder, there is a lack of studies on the possible association of other polymorphisms in the *TH* gene and the surrounding chromosomal region. Despite this polymorphism being functional, it accounts for a small amount of the variation in *TH* gene, which is an important candidate gene for involvement in many psychiatric disorders. Another interesting point is that the *TH* region has not been represented in genotyping arrays used for genome-wide association studies up until recently; no *TH* SNPs are present in the Affymetrix GeneChip Human mapping 500K array, which has been used for most of the schizophrenia genome-wide association studies. This means that recent genome-wide association studies on ADHD, schizophrenia as well as other psychiatric disorder have not examined polymorphisms in *TH*.

10.1.5 Interaction of *SRY* with *TH*

SRY can increase reporter gene activity under the control of the tyrosine hydroxylase promoter and therefore might mediate regulation of the dopaminergic pathway (Milsted et al. 2004). Of particular interest are data suggesting that SRY increases the number of dopaminergic neurons in males and compensates for their lack of oestrogen. This hypothesis stemmed from the fact that oestrogen in females is known to increase TH-expressing neurons (Raab et al. 1995). Most interestingly, Dewing et al. (2006) reported that downregulation of SRY using antisense oligonucleotides caused a decrease in TH expression and this led to motor deficits in male rats (Dewing et al. 2006). This interaction between a Y chromosome gene and an autosomal gene expressed in the brain points to another pathway leading to sex differences.

Although testing for an interaction of *SRY* and *TH* in our samples of patients with ADHD and schizophrenia was of great interest, it was not possible to perform due to the limited amount of variation in the *SRY* gene. In humans, *SRY* has only one exon and it spans 897 bp. We were able to locate only 3 SNPs in the *SRY* region using UCSC genome browser. Two of these were not polymorphic in our samples and one

failed to optimise in the genotyping platform used. For this reason, the previously genotyped Y chromosome SNPs were used to test for interaction with *TH* SNPs that showed evidence of association with the disorders. Since the Y chromosome SNPs are co-inherited, there would be a high degree of correlation between them and any *SRY* variation. While it was not possible to directly test for interaction between *TH* and polymorphisms in *SRY*, it was possible to test for interaction between TH variation and presence or absence of *SRY* i.e. sex.

10.2 Aims of the study

The aim of this study was to investigate if *TH* is implicated in sex differences in ADHD and schizophrenia and test for an interaction with Y chromosome variants. First, we tested for a main effect of *TH* on ADHD and schizophrenia. This was undertaken by genotyping 10 *TH* SNPs in 331 cases with ADHD, 487 cases with schizophrenia, 1,076 controls and performing two case-control association studies, one for ADHD and one for schizophrenia using a common set of controls. Secondly, the hypothesis that *TH* can have a stronger effect in males was tested by stratifying the cases and the control individuals by sex and performing separate association testing for males and females. Thirdly, we tested for an interaction between the Y chromosome SNPs we had already genotyped and *TH* SNPs that showed evidence of association with the disorders.

10.3 Materials and methods

10.3.1 ADHD sample description

The ADHD sample used in this study consists of 331 individuals (294 males and 37 females). It is the same sample used for the Y chromosome study described in Chapter 7 with the addition of 121 new male cases and 37 female cases. The sample characteristics and assessment remains the same.

10.3.2 Schizophrenia sample description

The schizophrenia sample used in this study consists of 487 individuals (312 males and 175 females). It is the same sample used for the Y chromosome study described in Chapter 7 with the addition of 175 female cases. The sample characteristics and assessment remains the same.

10.3.3 1958 UK birth cohort control sample

The control sample used in this study consists of 1,076 individuals (646 males and 430 females) drawn from the 1958 British Birth Cohort as discussed in Chapter 7.

10.3.4 Selection of *TH* SNPs

TH has 14 exons and spans 3.8 kb. A region including approximately 35 kb around the gene was targeted for selecting SNPs in order to capture variation in the promoter and any other regulatory regions. The SNPs were selected to include all SNPs that were located in the actual gene area and then more tagging SNPs were added to maximise the coverage in the surrounding 35 kb area. These extra SNPs were chosen using Haploview Tagger 4.1 (pairwise tagging with $r^2 > 0.8$). The MassARRAY Assay Design software was used to design a panel incorporating the SNPs in question.

We were able to successfully genotype 10 SNPs in both patient and controls samples. Table 10-1 shows the SNPs genotyped and their alleles.

Table 10-1. SNPs in *TH* 35 kb area genotyped in ADHD cases, schizophrenia cases and controls

SNP rs number	Minor allele	Major allele
rs10770140	C	T
rs10840489	T	C
rs10840490	G	C
rs10840491	A	G
rs11042978	T	G
rs11564703	G	A
rs11564709	T	C
rs2070762	G	A
rs4320932	C	T
rs7924316	T	G

10.3.5 DNA samples and genotyping

DNA was obtained from venous blood, mouthwash sample or saliva samples for the ADHD and schizophrenia patients. DNA samples were obtained from the 1958 UK Birth Cohort after request. Genotyping was performed with the MassARRAY and iPLEX systems (Sequenom, San Diego, CA) according to manufacturers' recommendations. Assays were optimised in 30 CEU individuals. All plates contained cases, controls, blanks, and CEU samples. Genotypes were called in duplicate blind to sample identity and blind to the other rater. All SNPs included in the association analysis had a call rate of >95%. Individuals were only included in analysis if they had genotypes for over 80% of the SNPs analysed. All SNPs included in the analysis were in Hardy-Weinberg (HW) equilibrium at $p>0.001$ for both cases and controls. For more information about quality control procedures see Appendix I.

10.3.6 Statistical analysis

In order to test for differences in allelic frequencies between cases and controls in the ADHD and the schizophrenia case-control samples, PLINK 1.06 (Purcell et al. 2007) (http://pngu.mgh.harvard.edu/purcell/plink/), an open-source genome-wide association analysis toolset was used. The same package was used to assess Hardy-Weinberg (HW) equilibrium for each SNP for cases and controls and to test if the

difference in call rates for each SNP between cases and controls was statistically significant. Secondary analyses including testing for a sex-specific effect and testing for an interaction of Y chromosome SNPs with *TH* SNPs were also performed using PLNK 1.06. The same analysis toolset was used to test for interaction of *TH* SNPs with sex.

The threshold of significance was set at $p \leq 0.005$ after applying a Bonferroni correction for the number of genotyped SNPs. With this threshold of significance, this study had >95% statistical power to detect effect sizes of 1.5 when the risk allele frequency was 0.3-0.4. For effect sizes of 1.3 the power was reduced to 85% and for effect sizes of 1.2 the power was 56%. For rarer alleles with frequencies of 0.1-0.2 the power was 86% for effect sizes of 1.5, 47% for effect sizes of 1.3 and 25% for effect sizes of 1.2. Power estimations were similar for both disorders. The prevalence of ADHD was considered to be 4% according to Cuffe et al. (2005) and the prevalence of schizophrenia was considered to be 1%. Statistical power calculations were performed using the Genetic Power Calculator (Purcell et al. 2003) (http://pngu.mgh.harvard.edu/~purcell/gpc/).

10.4 Results

10.4.1 Case-control association study in ADHD

10.4.1.1 Allelic analysis

10 *TH* SNPs were genotyped in 331 cases with ADHD and 1076 controls. Allelic analysis was performed using PLINK 1.06, which compares the frequencies of each SNP allele in cases and controls by applying a Pearson's chi-square test (df=1). Table 10-2 summarises these results and shows the frequencies of *TH* SNP alleles in cases and controls.

Table 10-2. Allelic analysis for ADHD case-control association study

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	41.5	37.6	3.32	1.18	0.99 – 1.41	0.068
rs10840489	12.8	14.3	0.95	0.88	0.68 – 1.14	0.33
rs10840490	12.8	14.1	0.7	0.9	0.69 – 1.16	0.4
rs10840491	13	13.3	0.03	0.98	0.75 – 1.27	0.87
rs11042978	47.7	49.3	0.5	0.94	0.79 – 1.12	0.48
rs11564703	39.1	41.2	0.86	0.92	0.77 – 1.1	0.35
rs11564709	10.8	8.8	2.42	1.26	0.94 – 1.68	0.12
rs2070762	46.1	50.1	3.3	0.85	0.7 – 1.01	0.069
rs4320932	20	18.8	0.46	1.08	0.87 – 1.35	0.5
rs7924316	48	45.9	0.96	1.09	0.92 – 1.3	0.33

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval

10.4.1.2 Case-control association study within each sex

The ADHD case-control sample was split according to sex. The male sample consisted of 294 ADHD cases and 646 controls while the female sample consisted of 37 ADHD cases and 430 controls. Allelic analysis was performed separately in males and females using PLINK 1.06 as before, although the number of females is too small for meaningful interpretation. Table 10-3 summarises the results in males while Table 10-4 summarises the results in females.

Table 10-3. Allelic analysis for ADHD case-control association study in males only

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	41	37.8	1.69	1.14	0.93 – 1.4	0.19
rs10840489	12.9	13.5	0.15	0.94	0.71 – 1.26	0.7
rs10840490	13.1	13.5	0.07	0.96	0.72 – 1.28	0.79
rs10840491	13.3	12.6	0.17	1.06	0.8 – 1.42	0.68
rs11042978	48	49.5	0.41	0.94	0.77 – 1.14	0.52
rs11564703	39.5	39.9	0.03	0.98	0.80 – 1.2	0.85
rs11564709	11.5	8.7	3.66	1.37	0.99 – 1.88	0.056
rs2070762	45.7	50.5	3.74	0.82	0.68 - 1	0.053
rs4320932	20.5	19.1	0.51	1.09	0.86 – 1.4	0.48
rs7924316	48.5	46	1.01	1.11	0.91 – 1.34	0.32

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval

Table 10-4. Allelic analysis for ADHD case-control association study in females only

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	45.8	37.2	2.12	1.43	0.88 – 2.32	0.15
rs10840489	12.2	15.5	0.58	0.76	0.37 – 1.56	0.45
rs10840490	10.8	15	0.96	0.69	0.32 – 1.47	0.33
rs10840491	11.1	14.3	0.57	0.75	0.35 – 1.6	0.45
rs11042978	46	49	0.25	0.89	0.55 – 1.43	0.62
rs11564703	36.1	43	1.28	0.75	0.46 – 1.24	0.26
rs11564709	5.4	9.0	1.08	0.58	0.21 – 1.64	0.3
rs2070762	48.7	49.4	0.02	0.97	0.6 – 1.56	0.9
rs4320932	16.2	18.4	0.22	0.86	0.45 – 1.63	0.64
rs7924316	44.6	45.7	0.03	0.96	0.59 – 1.54	0.86

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval

According to the results, there is no evidence of a significant association of *TH* SNPs and ADHD diagnosis. There is neither a male-specific effect of *TH* variation on ADHD. There are too few females to draw meaningful conclusions.

10.4.2 Case-control association study in schizophrenia

10.4.2.1 Allelic analysis

10 *TH* SNPs were genotyped in 487 cases with schizophrenia and 1,076 controls. Allelic analysis was performed using PLINK 1.06, which compares the frequencies of each SNP allele in cases and controls by applying a Pearson's chi-square test (df=1). Table 10-5 summarises these results and shows the frequencies of *TH* SNP alleles in cases and controls.

Table 10-5. Allelic analysis for schizophrenia case-control association study

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	41.6	37.6	4.50	1.18	1.01 – 1.38	0.034*
rs10840489	13.5	14.3	0.36	0.93	0.75 – 1.17	0.55
rs10840490	13.6	14.1	0.18	0.95	0.76 – 1.19	0.67
rs10840491	13	13.3	0.04	0.98	0.78 – 1.22	0.84
rs11042978	46.1	49.3	2.82	0.88	0.75 – 1.02	0.093
rs11564703	43.6	41.2	1.68	1.11	0.95 – 1.29	0.19
rs11564709	11.2	8.8	4.51	1.31	1.02 – 1.68	0.034*
rs2070762	48.4	50.1	0.81	0.93	0.8 – 1.09	0.37
rs4320932	17.3	18.8	1.04	0.90	0.74 – 1.1	0.31
rs7924316	46.5	45.9	0.11	1.03	0.88 – 1.2	0.74

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval;

*: significant at the $p \leq 0.05$ level

10.4.2.2 Case-control association study within each sex

The schizophrenia case-control sample was split according to sex. The male sample consisted of 312 schizophrenia cases and 646 controls while the female sample consisted of 175 schizophrenia cases and 430 controls. Allelic analysis was performed separately in males and females using PLINK 1.06 as before. Table 10-6 summarises the results in males while Table 10-7 summarises the results in females.

Table 10-6. Allelic analysis for schizophrenia case-control association study in males only

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	44.2	37.8	7.13	1.3	1.07 – 1.58	0.008**
rs10840489	12.4	13.5	0.45	0.9	0.68 – 1.21	0.5
rs10840490	13	13.5	0.12	0.95	0.72 – 1.26	0.73
rs10840491	12.4	12.6	0.01	0.99	0.74 – 1.32	0.92
rs11042978	43.8	49.5	5.44	0.8	0.66 – 0.96	0.02*
rs11564703	43.9	39.9	2.71	1.18	0.97 – 1.43	0.1
rs11564709	11.7	8.7	4.43	1.4	1.02 – 1.91	0.035*
rs2070762	47.1	50.5	1.98	0.87	0.72 – 1.06	0.16
rs4320932	18.1	19.1	0.24	0.94	0.73 – 1.21	0.62
rs7924316	48.2	46	0.86	1.1	0.90 – 1.33	0.35

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval;
*: significant at the $p \leq 0.05$ level; **: significant at the $p \leq 0.01$ level

Table 10-7. Allelic analysis for schizophrenia case-control association study in females only

SNP rs number	MAF (cases) %	MAF (controls) %	χ^2	OR	95% CI	p value
rs10770140	36.9	37.2	0.01	0.99	0.76 – 1.28	0.92
rs10840489	15.4	15.5	0.001	0.99	0.70 – 1.41	0.97
rs10840490	14.6	15	0.04	0.97	0.68 – 1.37	0.85
rs10840491	14.1	14.3	0.01	0.98	0.69 – 1.4	0.91
rs11042978	50	49	0.11	1.04	0.81 – 1.34	0.74
rs11564703	43.1	43	0.003	1.01	0.78 – 1.29	0.96
rs11564709	10.3	9	0.52	1.17	0.77 – 1.77	0.47
rs2070762	50.6	49.4	0.13	1.05	0.82 – 1.34	0.72
rs4320932	15.8	18.4	1.16	0.83	0.59 – 1.16	0.28
rs7924316	43.4	45.7	0.51	0.91	0.71 – 1.17	0.47

MAF: Minor allele frequency, OR: Odds Ratio; CI: Confidence Interval

In the total sample, there is evidence of a nominally significant association for SNPs rs10770140 (p=0.034, OR=1.18) and rs11564709 (p=0.034, OR=1.31) with schizophrenia. These associations become stronger in the male-only sample of patients with schizophrenia (p=0.008, OR=1.3 for rs10770140 and p=0.035, OR=1.4 for rs11564709) and also a third SNP, rs11042978, becomes nominally significant (p=0.02, OR=0.8). The direction of effects was the same between the male-only and the combined sample. There is no evidence of association of any SNP in the female-only sample of patients with schizophrenia. The three SNPs that showed evidence of significance were tested for interaction with sex. None of them significantly interacted with sex (p=0.099 for rs10770140, p=0.1 for rs11042978 and p=0.57 for rs11564709).

The SNPs that give nominally significant p values in the sample of all schizophrenia patients, rs10770140 and rs11564709, are not in high LD (r^2=0.22) (see Figure 10-3). The SNP that is only significant in the male-only sample, rs11042978, is in LD with rs10770140 (r^2=0.66) but not with rs11564709 (r^2=0.16) (see Figure 10-3). These three SNPs are located upstream of the TH gene.

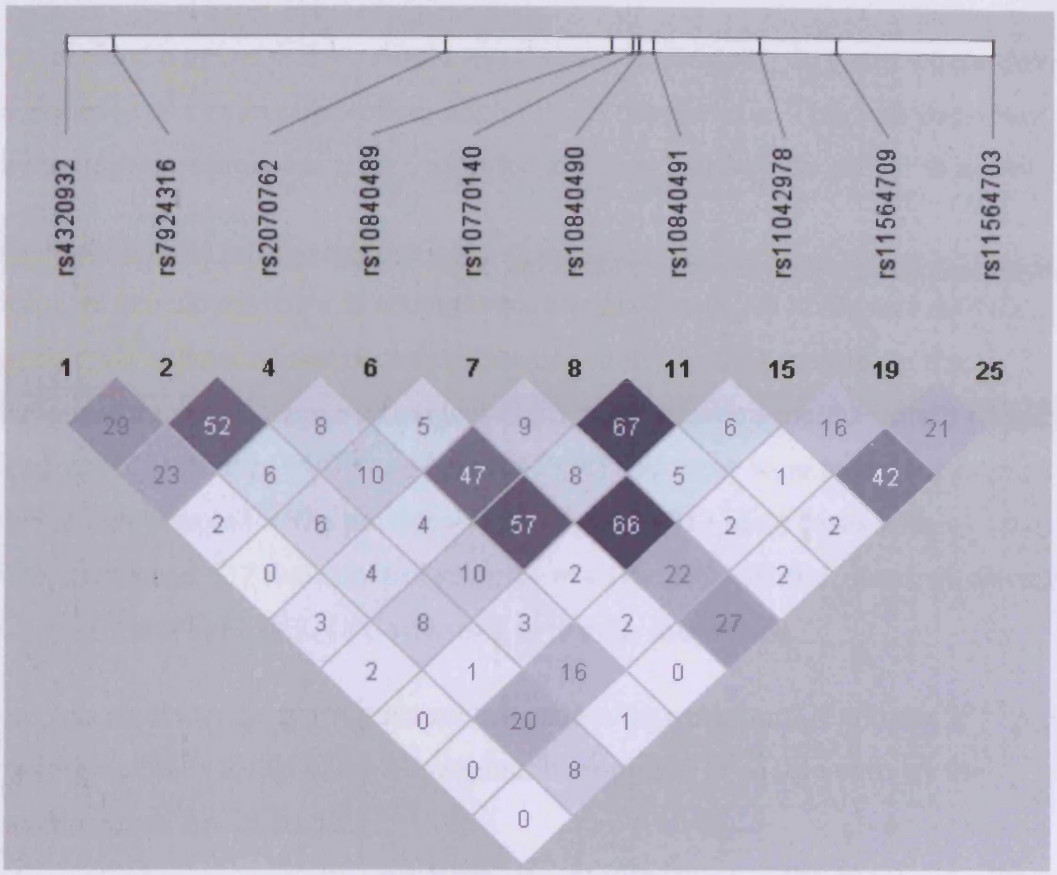

Figure 10-3. Pairwise LD between the SNPs genotyped. The Haploview LD plot shows the r^2 values between each SNP pair in the CEU population using data from HapMap Phases 1 and 2

10.4.3 Testing for an interaction between Y chromosome and *TH* SNPs

It was decided *a priori* that we would test for interactions only in cases where there was evidence of a main effect of certain SNPs on the disorder. This was important in order to limit the number of tests performed and thus increase the power to detect significant interactions.

Since there was no evidence of a significant association of *TH* SNPs and ADHD diagnosis, no interaction testing was performed in the ADHD sample. In the schizophrenia sample, there was evidence of a main effect on the disorder for three SNPs (rs10770140, rs11564709 and rs11042978) and these were tested for interaction with 9 Y chromosome SNPs previously genotyped in 313 male cases with schizophrenia and 637 male control samples from the 1958 Birth Cohort, as discussed in Chapter 7 and 8. PLINK 1.06 was used to test for interaction.

There was no evidence of a significant interaction between the *TH* and the Y chromosome SNPs tested in the schizophrenia sample. For a table with all the interactions, see Appendix III.

10.5 Discussion

In this chapter the hypothesis that *TH* is implicated in sex differences in ADHD and schizophrenia was tested. The main reason for hypothesising that *TH* can be implicated in sex differences is because of previous evidence of interaction with the *SRY* gene on the Y chromosome. In addition, *TH* is involved in dopamine metabolism which is important for both ADHD and schizophrenia. Although *TH* seems an obvious candidate for neuropsychiatric disorders, there has been limited research on the role of polymorphisms in the *TH* gene apart from an intronic microsatellite. In addition, the *TH* gene region has not been examined in genome-wide studies because it was not covered by previous genotyping arrays used for genome-wide association studies of psychiatric disorders.

In our study, there was no main effect of *TH* SNPs on ADHD. This was also true when the sample was split according to sex. These results agree with the previous studies where no association between *TH* and ADHD was found (Barr et al. 2000; Comings et al. 1995; Payton et al. 2001). However, these studies are different to this one because they just focused on a microsatellite repeat in intron 1, while in our study SNPs that cover *TH* and the area around the gene were chosen. Although *TH* seems to be a very strong candidate for involvement in ADHD, the neurobiology of ADHD is largely unknown and there is no reason why other neurotransmitter pathways or indeed completely novel mechanisms could not be implicated in the pathogenesis of ADHD.

In the schizophrenia case-control association study there were two SNPs which showed some evidence of association [rs10770140 (p=0.034, OR=1.18) and rs11564709 (p=0.034, OR=1.31], although they did not reach the required level of significance when correcting for multiple testing. The associations were driven by the male individuals of the schizophrenia sample, as shown by the fact that the association with the previously associated SNPs becomes more strongly significant (p=0.008, OR=1.3 for rs10770140 and p=0.035, OR=1.4 for rs11564709) and also one more SNP, s11042978, became nominally significant (p=0.02, OR=0.8). However, all the results in the male-specific sample of the schizophrenia patients

were only nominally significant. In addition, there were 312 male patients and only 175 female patients with schizophrenia. When testing *TH* SNPs for interaction with sex, there were no significant results. This means that the association of *TH* SNPs with schizophrenia could exist in the female sample as well but there was not enough power to detect it due to the limited number of female patients with schizophrenia. It is also possible that these are false positive results especially since they do not achieve the required level of significance when correcting for multiple testing. These results are in agreement with previous studies that have failed to find an association between *TH* and schizophrenia (Andreou et al. 2009; Ishiguro et al. 1998; Jonsson et al. 1998; Ota et al. 2001). The studies that have found an association were all performed in non-European populations (Jacewicz et al. 2006; Kurumaji et al. 2001; Pae et al. 2003). In addition, all the studies have focused on the tetranucleotide repeat of the first intron while our study achieved a more comprehensive coverage of the gene.

It is possible that *TH* has a significant effect not on the disorder itself but on some aspect of the phenotype, as it was the case for the Y chromosome haplogroups. However, in the case of Y chromosome haplogroups we had *a priori* reasons to choose the specific measures due to the fact that were sexually dimorphic and thus could be modified by Y chromosome variation. There was no previous knowledge of any phenotypic aspects of ADHD or schizophrenia that is affected by *TH* variation and thus it was not possible to select measures more likely to be modified by *TH.*

One of the strongest points of this study was that there was a strong *a priori* reason for testing *TH* for involvement in sex differences in ADHD and schizophrenia. There is evidence that *SRY* regulates *TH* expression (Milsted et al. 2004) and animal studies have shown that *TH* interacts with *SRY* (Dewing et al. 2006). Since oestrogen increase *TH* expression in females, it is intriguing that *SRY* was considered to be the factor that compensates for the lack of oestrogen in males. Another advantage of this study is that we achieved good coverage of the variation not only in the *TH* gene but also of a 35kb area around the gene, which includes the promoter region. We decided, instead of focusing on the functional polymorphism in intron 1, which had been extensively genotyped without strong evidence of involvement in psychiatric disorders, to genotype other SNPs which have not been examined by candidate gene or genome-wide association studies. Finally, the ADHD and schizophrenia samples were large

enough to have adequate power to detect alleles with large to medium effect sizes (>86%). The level of significance was *a priori* set at $p<0.005$ to reflect the number of tests performed in each sample.

However, there are some limitations to this study. Firstly, it was not possible to genotype four *TH* SNPs due to technical reasons. One of the SNPs leads to a non-synonymous aminoacid sequence change and it would have been of great interest to incorporate in the study. Another possible reason for the results is that *TH* has such a small effect size that our statistical power is very limited. The power to detect effect sizes smaller than 1.2 is less than 25% unless the risk allele is very common.

Another issue that deserves to be discussed is that SNPs in the *SRY* gene that has been found to interact with *TH* were not genotyped. The *SRY* gene was so small (897kb) that only three polymorphisms were found. Two of them were actually non-polymorphic in our samples. It would have been interesting to investigate whether certain *SRY* alleles interact with *TH* SNP alleles to increase risk to psychiatric disorders.

The issue of testing for interactions has been a matter of debate recently. Power is limited when testing for interactions and even more in this case where the sample had to be limited to male cases to test for interaction of *TH* with Y chromosome SNPs. For this reason it was decided that we would test for interactions only when a main effect is found. In this way the number of tests performed was kept to a minimum and the power to detect any significant association was not compromised. However, the objection here is that some genes may not have any main effect on the disorder and yet they could be important for the pathophysiology of the disorder when other factors, such as other gene alleles or environmental factors are taken into account.

To conclude, there is no evidence of association of *TH* in ADHD. There is weak evidence of a male-specific effect of *TH* on schizophrenia but no evidence of these variants interacting with Y chromosome SNPs to increase risk to schizophrenia. Thus, replication in larger studies achieving a good coverage of the *TH* gene and the surrounding chromosomal region is needed.

Chapter 11 General discussion

11.1 Summary of findings

The main aim of this thesis was to investigate the genetic causes of sex differences in ADHD and schizophrenia focusing on the previously neglected Y chromosome. Both disorders were studied in this thesis due to their neurodevelopmental nature and male excess. In ADHD there are consistently higher rates of male patients and the manifestation of the disorder also differs between male and female patients (Cuffe et al. 2005). The male excess in schizophrenia is not as profound, although there is a substantially earlier age of onset, more severe disorder and poorer outcome in male patients (Grossman et al. 2008; Ober et al. 2008). The causes of sex differences in psychiatric disorders as well as in brain and behaviour are likely to be multiple and complex. More than 600 genes are differentially expressed in mouse brain (Yang et al. 2006) and hormonal, genetic and environmental factors all contribute to the sexual dimorphism of the brain. One of the least studied factors (especially in the field of psychiatric genetics) is the Y chromosome.

The Y chromosome is 60 Mb in length and most of its part is covered by the male-specific region (MSY), which does not recombine with the X chromosome. The lack of recombination contributes to the geographic differentiation of Y chromosome haplogroups making the choice of the appropriate haplogroups for the population studied crucial in Y chromosome studies. Animal studies have implicated the Y chromosome in brain function and behaviour (De Vries et al. 2002; Gatewood et al. 2006; Tordjman et al. 1995) and individuals with certain Y chromosome abnormalities are at increased risk of psychiatric disorders (Gotz et al. 1999; Mulligan et al. 2008; Yoshitsugu et al. 2003). One of the most interesting genes on the Y chromosome is the SRY gene, which is responsible for male sex differentiation. The SRY gene is expressed in mouse and human brain (Dewing et al. 2006; Mayer et al. 1998) and it has been found to interact with the rate–limiting enzyme in dopamine metabolism, tyrosine hydroxylase (*TH*) (Dewing et al. 2006).

The role of Y chromosome in ADHD and schizophrenia was investigated by testing for a main effect on the disorders, testing for a modifying effect by influencing

clinical presentation and/or cognitive performance and finally testing for an interaction with *TH*. The main findings are summarised below:

- No evidence of a main effect of Y chromosome SNPs and haplogroups on ADHD and schizophrenia.

- There was evidence of a modifying effect of Y chromosome haplogroups on ADHD by influencing IQ scores. Specifically, Full Scale IQ (FSIQ) [$F(2,199)=3.82$, $p<.05$, $r=0.19$] and Performance IQ (PIQ) [$F(2,199)=4.39$, $p<.05$, $r=0.21$] scores were significantly higher in haplogroup 3 compared to haplogroup 1.

- No evidence of an effect of Y chromosome haplogroups on IQ scores in the general population.

- No evidence of a modifying effect of Y chromosome haplogroups on IQ scores in an independent sample of patients with ADHD. However, the sample size was very small to detect any association.

- There was evidence of a modifying effect of Y chromosome haplogroups on schizophrenia, specifically an association with the course of illness and the educational level achieved. Specifically, haplogroup 1 had a higher percentage of individuals with good than poor outcome while the opposite was true for haplogroup 2 [χ^2 (2) = 8.98, p=0.011, OR =2.25]. Haplogroup 1 had also a higher percentage of individuals with high than low educational level while the opposite was true for haplogroup 3 [χ^2 (2) = 6.35, p=0.042, OR =2.5].

- No evidence of an association of *TH* SNPs with ADHD was found. There was also no evidence of a sex-specific effect of *TH* SNPs on ADHD.

- There was evidence of a nominally significant association for *TH* SNPs rs10770140 (p=0.034, OR=1.18) and rs11564709 (p=0.034, OR=1.31) with schizophrenia diagnosis. These associations became stronger in the male-only sample of patients with schizophrenia (p=0.008, OR=1.3 for rs10770140 and p=0.035, OR=1.4 for rs11564709) and also a third SNP, rs11042978, became nominally significant (p=0.02, OR=0.8). There was no evidence of association

in the female patients with schizophrenia. There was no evidence of an interaction of *TH* and Y chromosome SNPs increasing risk for schizophrenia.

11.2 Y chromosome haplogroup distribution in the UK

As discussed before, the Y chromosome haplogroup distribution depends highly on geography. The frequencies of these haplogroups can differ extensively even within areas of the same country. Research on the Y chromosome haplogroup distribution in the UK has mainly focused on studying small isolated populations in order to achieve a high degree of haplogroup breakdown (Capelli et al. 2003). However, this is not desirable for psychiatric genetics studies which involve a large number of individuals from different areas of the UK. The Y chromosome haplogroup distribution in the UK would be useful for researchers wishing to study the effect of Y chromosome on a number of disorders using populations of UK Caucasian origin.

During the course of this project a large number of UK Caucasian individuals have been genotyped for Y chromosome variants. Specifically, 1,160 UK Caucasian males, including 210 patients with ADHD, 313 patients with schizophrenia and 637 individuals from the 1958 Birth Cohort were genotyped for nine Y chromosome SNPs, which covered nine Y chromosome haplogroups. This is the first time to my best knowledge that Y chromosome haplogroups have been genotyped in such a large sample of UK Caucasian men. The frequencies of the nine different haplogroups are very similar between these three different groups of UK men (Table 11-1).

Table 11-1. Frequencies of Y chromosome haplogroups in patients with ADHD, patients with schizophrenia and controls

Y chromosome haplogroup	ADHD cases (%)	SZ cases (%)	Controls (%)	Total (%)
DE	2.4	4.2	3.3	3.3
G*	1.9	1.9	1.4	1.7
I*	22.3	18.5	20.7	20.5
J*	0.9	1.9	0.6	1.1
J2	1.9	1	1.7	1.5
K*	0.9	0.3	0.5	0.6
R1*	53.1	55.6	52.1	53.6
R1b3M405	16.6	16.6	19.5	17.6
R1b3d	0	0	0.2	0.1

The similarity in the haplogroup distribution in the three groups of males supports our notion that the Y chromosome SNPs that are suitable for UK populations were chosen after careful consideration and after consulting members of the Y Chromosome Consortium (The Y Chromosome Consortium 2002) for their advice in this matter. It also confirms that the 1958 UK Birth Cohort is a suitable control population for Y chromosome studies in the UK, since there is no evidence of population stratification of the Y chromosome haplogroups.

The ADHD sample included individuals from both England and Wales. The postcodes of the individuals were available and, although they do not always correlate with the person's origin, they provided a measure of geographic location. There was no difference in the haplogroup distribution between English and Welsh individuals, a finding providing more evidence that this is a panel of Y chromosome SNPs appropriate for any UK population.

The Y chromosome haplogroups that were expected to be frequent in the UK were I* and R*. As seen in Table 11.1, this is true for haplogroup I* with a total frequency of 20.5% and haplogroup R*, which includes haplogroups R1* with a total frequency of 53.6% and R1b3M405 with a total frequency of 17.6%.

In addition to the 1,160 UK men genotyped for nine Y chromosome SNPs, another 3,790 males from ALSPAC were genotyped for Y chromosome SNPs M173 and M405. These two SNPs define the three major haplogroups that resulted when the nine haplogroups were merged according to ancestry in order to increase the numbers in each group (see Chapter 7 for more details). This means that 4,950 UK male individuals were genotyped for these two SNPs. It is the first time that any Y chromosome polymorphism has been genotyped in such a large sample of men. Frequencies of haplogroups 1-3 can be seen in Table 11-2 and they look remarkably similar between ALSPAC and the rest of the UK individuals.

Table 11-2. Frequencies of merged Y chromosome haplogroups in patients with ADHD, patients with schizophrenia and controls from the 1958 Birth Cohort and ALSPAC

Merged Y chromosome haplogroup	ADHD cases (%)	SZ cases (%)	1958 BC Controls (%)	ALSPAC (%)	Total (%)
1	30.3	27.2	28.3	27.1	28.3
2	53.1	55.3	52.1	49.5	53.3
3	16.6	17.4	19.6	23.4	18.4

11.3 The role of the Y chromosome on ADHD

The aims of this project included testing for a main effect and a modifying effect of the Y chromosome on ADHD. Although this study provided no evidence for a main effect of Y chromosome haplogroups on ADHD, there was evidence for a possible modifying effect by influencing cognitive performance. IQ scores were higher in haplogroup 3 and the difference in FSIQ and PIQ between haplogroup 1 and haplogroup 3 was significant.

Although there is no knowledge about the mechanism through which the Y chromosome haplogroups could modify the phenotype of ADHD, it is possible to speculate. The most likely scenario, since the most recent in evolutionary terms haplogroup is associated with higher IQ scores, is that a mutation appeared at some point during evolution against a specific Y chromosome background. This mutation was beneficial to patients with ADHD by protecting them against cognitive

impairment. The way this is achieved could be through interactions with other autosomal or X chromosome variants that increase risk to ADHD. This means that possession of certain variants that increase risk to ADHD and a certain Y chromosome haplogroup background leads to less cognitive impairment and thus higher IQ scores in patients with ADHD. Since both ADHD and IQ are complex traits, it is possible that environmental factors are also involved in this process creating an even more complicated pathway.

Another explanation for this association could be that these results are false positive findings. For the significance threshold used in this study ($p \leq 0.05$), there is a 5% chance that these results would be observed even if there are no true associations of Y chromosome haplogroups with any of the phenotypic measures studied. The only way to increase confidence on these results would be to replicate them in a large independent sample of patients with ADHD with available IQ scores.

11.4 The role of the Y chromosome on schizophrenia

The aims of this project included testing for a main effect and a modifying effect of the Y chromosome on schizophrenia. No main effect of Y chromosome on schizophrenia was found. However, there was a modifying effect of Y chromosome haplogroups on schizophrenia. Y chromosome haplogroups were associated with the course of illness and the educational level achieved. Haplogroup 1 had a significantly higher frequency of individuals with good outcome than haplogroup 2 while the same haplogroup also had a significantly higher frequency of individuals with high education. Haplogroup 1 is the most ancient in evolutionary terms from the three that were compared in schizophrenia. It can be speculated that in this case a mutation appeared against a specific Y chromosome haplogroup background that increased the susceptibility to a poor outcome and more impaired cognition in patients with schizophrenia. This mutation might have been beneficial or neutral to the general population; but coupled with other risk factors for schizophrenia (genetic or environmental), it was possible to modify the phenotype of the disorder. However, there are some issues with using educational level as a proxy for cognitive ability in schizophrenia. The onset of the disorder usually coincides with important periods for education in an individual's life, thus it is difficult to estimate the educational level

that would have been achieved otherwise. In addition, even in periods of recovery cognitive impairment can still be present and it will have a considerable effect on the educational level achieved. Nevertheless, the educational level was the only available measure of cognitive performance in the sample of patients with schizophrenia and the direction of the effect is in agreement with the results obtained for the course of illness (the same haplogroup is associated with both good outcome and high educational level).

11.5 Y chromosome interaction with Tyrosine Hydroxylase (*TH*)

It was hypothesised that the Y chromosome can interact with *TH*, which is the rate-limiting enzyme in dopamine metabolism, and increase risk to the disorders. Evidence for interaction of *TH* with the *SRY* gene has been found in male rats (Dewing et al. 2006). This hypothesis was tested by performing two case-control association studies, one for ADHD and one for schizophrenia using *TH* tag SNPs. Then, these SNPs were tested for a sex-specific effect on the disorders and, finally, *TH* SNPs were tested for interaction with already genotyped Y chromosome SNPs only when a main effect was detected.

Studies on the role of *TH* in psychiatric disorders have focused on an intronic microsatellite. Research on other polymorphisms in and around the gene area has been very limited. As the name suggests genome-wide association studies are designed to survey the entire genome for common variants that might increase risk to a complex disorder. They utilise the large number of SNPs that have been discovered and the information on linkage disequilibrium patterns across the genome as provided by HapMap. However, genotyping arrays used for genome-wide association studies of psychiatric disorders (like the Affymetrix GeneChip Human Mapping 500K Array) examined ~500,000 SNPs. This coverage is not 100% and this is without taking into account the removal of SNPs due to quality control measures which would reduce the coverage further. One of the regions not covered by genotyping arrays that have been used for genome-wide association studies of psychiatric disorders is the *TH* gene region.

Our study achieved good coverage (>80%) of the variation not only in the *TH* gene but also of a 35kb area adjacent to the gene. However, there was no association of *TH* SNPs with ADHD and this was also true when the sample was split according to sex. The fact that no association was found between *TH* and ADHD does not exclude an involvement of *TH* in ADHD because the power to detect an association was limited especially when the effect size was small. This is one of the biggest problems in research of complex disorders at the moment. Detecting associations with alleles of small effects requires a large sample size. Even a common allele with a relatively large effect size (OR=1.5) needs to be genotyped in a large sample (1500 cases and 1500 controls) to reach genome-wide significance. When multiple test correction is taken into account (as with genome-wide association studies), the sample sizes for detecting alleles of small effect sizes are increased to numbers that are possible to achieve only in large collaborative studies. In addition, variants with large effect sizes but very low frequencies in the population would again be very difficult to detect unless sample sizes are large enough. Thus, negative results cannot be conclusive and an involvement of *TH* in ADHD cannot be excluded.

The case-control association study in schizophrenia revealed three SNPs that were nominally significant in the male-only sample. First of all, the required level of significance was not achieved, thus these results could be a false positive finding. However, if this is not the case, these results could indicate a possible role of *TH* in schizophrenia and especially in male patients. Since there was no association in the female patients with schizophrenia, it is reasonable to speculate that *TH* interacts with a male-specific factor to increase risk to schizophrenia in male patients only. One of the most obvious male-specific factors is the Y chromosome. Animal studies have already shown an interaction of *TH* with the *SRY* gene on the Y chromosome. Thus, it is possible that the mechanism by which *TH* increases risk to schizophrenia is by interacting with polymorphisms on the *SRY* gene or other parts of the Y chromosome, although this study did not provide any evidence towards this.

11.6 Implications of research results

This thesis identified a possible modifying effect of Y chromosome haplogroups on ADHD and schizophrenia. Previous association studies of the Y chromosome and psychiatric phenotypes have failed to find any evidence of association. However, most of the studies have not used homogeneous samples and Y chromosome markers appropriate for the populations they were studying. In addition, the sample sizes in these studies have been small compared with the individuals genotyped for this project. However, the most important issue is that few studies have investigated the effect of Y chromosome haplogroups on different phenotypic measures. This could be an important reason for their failure to detect association, since the Y chromosome might not have a main effect on the disorder but it could be modifying the phenotype by affecting clinical presentation or cognitive performance.

One of the implications of this research project is that it provides a set of Y chromosome markers that can be used for future association studies in UK populations. Since these markers and the haplogroups they cover represent the UK and not specific areas within the country, they can be used for any UK population provided that the individuals are Caucasian.

The most important point highlighted by these results is that not taking the Y chromosome into account in genetic studies is not justified, since these results suggest the possibility of a modifying effect of Y chromosome haplogroups on the phenotype of ADHD and schizophrenia. For a number of years it has been thought that the Y chromosome was only important for male sex differentiation and reproduction. This view has changed in recent years with animal models showing that the Y chromosome is implicated in brain function and behaviour (De Vries et al. 2002; Gatewood et al. 2006; Guillot et al. 1995) and evidence of its involvement in another phenotype with increased male susceptibility, hypertension (Charchar et al. 2003). In the previous years, the genetics of complex disorders have been mainly investigated using candidate gene association studies and linkage studies. The Y chromosome is suitable for neither of these two approaches: it would not be one of the obvious candidates (as, for example, genes related to neurotransmitter pathways) that have been the focus of candidate gene studies for psychiatric disorders and it is not possible to include in

linkage studies due to the lack of recombination. In addition, before the systematic effort of the Y Chromosome Consortium (The Y Chromosome Consortium 2002) the confusing nomenclature made choosing Y chromosome SNPs for each population a very difficult task. Nevertheless, today there are Y chromosome SNPs appropriate for most populations and new ones are still being discovered.

The approach of testing for association using a large number of SNPs across the whole genome has revolutionised the way the genetics of complex disorders are investigated. However, once again the Y chromosome has been excluded from these studies. As already mentioned, most of the genotyping arrays for genome-wide association studies have not included Y chromosome SNPs. This situation has changed with recent genotyping arrays having Y chromosome SNPs, although most genome-wide association studies still ignore Y chromosome SNPs in their analysis. It is striking that in a disorder with such a profound male excess as in ADHD, all genome-wide association studies performed so far have not analysed sex chromosomes.

To conclude, this thesis provides evidence that there are valid reasons to include the Y chromosome in genetic studies. The Y chromosome and other non-autosomal sources of DNA, such as the X chromosome and the mitochondrial DNA, have unique characteristics and it can be challenging to find the appropriate way of incorporating them into analyses. However, these are sources of DNA which could have an effect on a number of disorders and the effort invested in studying them and their interactions with autosomes can be worthwhile.

11.7 Moving from genetic association to understanding the biology

Identifying gene alleles that are associated with disease phenotypes is the first step in a long process towards the investigation of risk pathways that are relevant to disease and could help elucidate the pathophysiology or develop therapeutic targets.

Before even considering further genetic or functional studies to enhance biological knowledge about the disease, it is important to confirm the initial association. In the case of the association of Y chromosome haplogroups with phenotypic measures in

ADHD and schizophrenia, this would be possible only if large enough samples of patients with ADHD and schizophrenia and available phenotypic measures could be used. Genome-wide association studies have increased in number and recent genotyping arrays will be including SNPs on the Y chromosome. Utilising data from these studies, which tend to use large sample sizes to maximise power, could provide the means to achieve an independent replication of our current results provided that the studies have relevant phenotypic measures available.

If independent replication was achieved for any of the phenotypes in ADHD and schizophrenia associated with Y chromosome haplogroups, the next step would involve further genotyping of SNPs within the associated haplogroup in an attempt to narrow down the possible mutations involved.

Bioinformatic approaches would then be employed to gain more information about the chromosomal region in which the mutation is located. Greatest credibility for a variant being causal is usually assigned to non-synonymous coding SNPs and biological credibility is increased when genes are expressed in relevant tissues or code for a protein that takes part in interesting biological pathways. Sequencing efforts or utilisation of already available sequencing information could potentially increase confidence in specific genes or chromosomal regions, if rare mutations in the region are also found to be associated with the disorder.

Going a step further in implicating possible mechanisms behind the genetic association, would require information about the possible function of the mutation. Microarray technology has allowed the simultaneous assessment of the expression of the majority of genes in the genome and these have been named eQTLs (expression Quantitative Trait Locus), since the level of expression of a gene can be considered a quantitative trait. eQTL data are publically available for a number of tissues (Stranger et al. 2007) including human brain (Myers et al. 2007). Combining eQTL data with results from genome-wide association studies can yield findings that would otherwise be missed, as is demonstrated by a genome-wide association study into Crohn's disease. The strongest associated SNP in this study was located in a gene desert but examination of an eQTL database showed that it influences expression of certain genes (Libioulle et al. 2007).

Apart from *in silico* investigation, function can be evaluated in living organisms, although this can be particularly challenging. Developing cell lines is usually the first stage when investigating the functional impact of putative causal variants in living organisms but deciding on cell type, developmental stage and specific environmental conditions is not straightforward and can have a great impact on the experimental outcome. Even greater challenges are expected when using animal models to assess the function of putative causal variants because variants involved in complex disorders can have subtle effects that are not easy to observe in animals (McCarthy and Hirschhorn 2008).

In summary, the association of Y chromosome haplogroups with phenotypic measures in ADHD and schizophrenia needs to be replicated in large independent samples with available phenotypic measures. If and when the association is replicated, bioinformatic approaches and detailed investigation of the region the mutation is located could provide more information about its possible role with the aim to develop cell lines or animal models where the function of the mutation can be studied in living organisms.

11.8 Future directions

The work presented in this thesis has provided evidence about the possible role of the Y chromosome in ADHD and schizophrenia. The Y chromosome can be considered a part of the complex picture of the causes of sex differences in psychiatric disorders. Research on the sex differences of psychiatric disorders can be performed separately or simultaneously with studies investigating the genetics of complex disorders.

Since the most recent versions of genotyping arrays for genome-wide association studies include markers on non-autosomal sources of DNA, such as the sex chromosomes and mitochondrial DNA, they should be taken into account in future analysis. In some instances this will involve modifying the quality control criteria and the software used in order to accommodate the characteristics of these SNPs. In addition, the threshold of significance should be taken into consideration, especially for the Y chromosome which does not recombine. The stringent p value of $\sim p<10^{-8}$, which is required for a variant to achieve genome-wide significance (Pe'er et al.

2008), might be not necessary for Y chromosome SNPs that are not independent from each other, since they are co-inherited. Genotyping arrays for genome-wide association studies can also be used to investigate copy number variation (CNVs). The degree of penetrance and the pattern of CNVs in the genome can differ between males and females and now it is possible to investigate this at the same time as genome-wide association studies. Another approach that can lead to identification of sex-specific risk variants, not necessarily located on the sex chromosomes, is testing for association separately in each sex when performing genome-wide or candidate-gene association studies, since sex has been found to affect association results (Biederman et al. 2008; Guimaraes et al. 2007; Rommelse et al. 2008).

Studying the causes of sex differences in psychiatric disorders can be another way to obtain clues about the pathophysiology of disorders we know little about. Since it is possible that the variants leading to sex differences in psychiatric disorders would modify the disease phenotype without having a main effect on the disorder, it is of great importance to look deeper into the disease phenotype. This can be achieved by taking into account clinical presentation, such as comorbidity with other disorders and cognitive measures, which can be used as endophenotypes.

It is generally accepted that there are limitations in the current disease categories as defined by the diagnostic criteria. One way to overcome the lack of biological validity of disease diagnosis based on clinical symptoms is to further explore the phenotypic heterogeneity of the disorder. Comorbidity in most psychiatric disorders is the rule and not the exception. Identifying groups of patients with certain comorbidities that exhibit more profound sex differences can facilitate the identification of factors contributing to them. One example of such a successful approach is the identification of the *COMT* val/val genotype as a risk factor for conduct disorder in children with ADHD (Thapar et al. 2005). This finding has been replicated across three independent samples and it justifies considering patients with ADHD and conduct disorder a distinct group to those with just ADHD, since there is no association of this polymorphism in patients with ADHD but no conduct disorder (Caspi et al. 2008).

Apart from comorbidity, looking deeper into the phenotype can also yield results. Endophenotypes, or intermediate phenotypes, are quantitative, heritable traits that are thought to be more directly associated with risk factors than the disease diagnosis

(Castellanos and Tannock 2002). Investigating the potential of endophenotypes requires considerable effort in order to collect the appropriate information about the neuropsychological or cognitive performance of the patients. However, it is an approach that coupled with collaborative efforts to increase sample sizes can yield important results about the involvement of genetic factors in psychiatric disorders.

The advance of technology will revolutionise genetic studies and consequently the search for genetic causes of sex differences. The recent development of high-throughput sequencing or next-generation sequencing is becoming increasingly affordable and allows the parallel analysis of millions of sequence reads compared to the 96 that conventional capillary systems can process (Mardis 2008). This technology will be applied to sequence ~1000 human genomes as part of the 1000 genomes project (Siva 2008). The aim of this project is to discover >95 % of the variants (SNPs, CNVs, indels) with minor allele frequencies as low as 1%. For the first time an unprecedented amount of genetic variation would be available for testing for association with complex disorders as well as sex differences.

One of the main challenges when investigating the genetic causes of sex differences is to find a way of dissociating them from the hormonal effects. While this is not easy to do with genetic studies, animal models have provided the first evidence of hormone-independent effects of Y chromosome on brain and behaviour using the four-core genotypes model (De Vries et al. 2002) and it is likely that they will continue yielding results in the field of sex differences. The effect of the *SRY* gene, which is very difficult to assess with genetic studies due to the limited amount of variation, has been investigated in rats. By microinjecting *Sry* antisense in specific areas of the rat brain, it is possible to knockdown the *Sry* gene in certain brain regions without affecting the rest of the brain. Testing these rats on behavioural tasks will provide evidence about the cognitive effects of the *Sry* gene on the brain (Kopsida 2009 Personal communication).

To conclude, future directions about research in this field should include a careful consideration of the disease phenotype, the possibility of using large sample sizes while maintaining good quality in the sample characterisation, maximising the potential of new technologies and applying them to non-autosomal sources of DNA and making use of animal models.

11.9 Conclusions

The research described in this thesis involved systematically investigating the Y chromosome in ADHD and schizophrenia for the first time. The possible modifying effect of Y chromosome on ADHD and schizophrenia supports the notion that it is involved in brain function and behaviour and it is vital that both sex chromosomes as well as other sources of non-autosomal DNA are investigated in future genetic studies. Although sex differences in psychiatric disorders are well-documented, research on their causes is still in its infancy and it will require maximising the potential of existing and new technologies to gain an insight into them. Nevertheless, the effort will definitely be worthwhile, since research on the causes of sex differences is likely to provide clues about the pathophysiology of psychiatric disorders.

Chapter 12 References

Agate, R. J. et al. 2003. Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proc Natl Acad Sci U S A* 100(8), pp. 4873-4878.

Albanese, V. et al. 2001. Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite. *Hum Mol Genet* 10(17), pp. 1785-1792.

Aleman, A. et al. 2003. Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch Gen Psychiatry* 60(6), pp. 565-571.

Allen, N. C. et al. 2008. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 40(7), pp. 827-834.

American Psychiatric Association 1987. *Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition revised.* Washington, DC: American Psychiatric Association.

American Psychiatric Association 1994. *Diagnostic and Statistical Manual of Mental Disorders, 4th edition.* Washington, DC: American Psychiatric Association.

American Psychiatric Association 2000. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).* Washington, DC: American Psychiatric Publishing.

Andreou, D. et al. 2009. Tyrosine hydroxylase Val81Met polymorphism: lack of association with schizophrenia. *Psychiatr Genet* 19(5), pp. 273-274.

Angold, A. et al. 1995. The Child and Adolescent Psychiatric Assessment (CAPA). *Psychol Med* 25(4), pp. 739-753.

Anney, R. J. et al. 2008. Conduct disorder and ADHD: evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1369-1378.

Arcos-Burgos, M. et al. 2004. Attention-deficit/hyperactivity disorder in a population isolate: linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. *Am J Hum Genet* 75(6), pp. 998-1014.

Arnold, A. 2004. Sex chromosomes and brain gender. *Nature reviews* 5, pp. 1-8.

Arnold, A. P. 1997. Experimental analysis of sexual differentiation of the zebra finch brain. *Brain Res Bull* 44(4), pp. 503-507.

Asherson, P. et al. 1992. No evidence for a pseudoautosomal locus for schizophrenia. Linkage analysis of multiply affected families. *Br J Psychiatry* 161, pp. 63-68.

Asherson, P. et al. 2008. A high-density SNP linkage scan with 142 combined subtype ADHD sib pairs identifies linkage regions on chromosomes 9 and 16. *Mol Psychiatry* 13(5), pp. 514-521.

Autism Genome Project Consortium 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39(3), pp. 319-328.

Badner, J. A. and Gershon, E. S. 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7(4), pp. 405-411.

Bakker, J. et al. 2006. Alpha-fetoprotein protects the developing female mouse brain from masculinization and defeminization by estrogens. *Nat Neurosci* 9(2), pp. 220-226.

Bakker, S. C. et al. 2003. A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *Am J Hum Genet* 72(5), pp. 1251-1260.

Barr, C. L. et al. 2000. Further evidence from haplotype analysis for linkage of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. *Am J Med Genet* 96(3), pp. 262-267.

Barrett, J. C. 2009. Haploview: Visualization and analysis of SNP genotype data. *CSH Protoc* 2009(10), p. pdb ip71.

Becker, K. et al. 2008. Interaction of dopamine transporter genotype with prenatal smoke exposure on ADHD symptoms. *J Pediatr* 152(2), pp. 263-269.

Bhutta, A. T. et al. 2002. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 288(6), pp. 728-737.

Biederman, J. et al. 2008. Sexually dimorphic effects of four genes (COMT, SLC6A2, MAOA, SLC6A4) in genetic associations of ADHD: a preliminary study. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1511-1518.

Biederman, J. et al. 2002. Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *Am J Psychiatry* 159(1), pp. 36-42.

Blanco, P. et al. 2000. Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm Genome* 11(10), pp. 906-914.

Blaschke, R. J. and Rappold, G. 2006. The pseudoautosomal regions, SHOX and disease. *Curr Opin Genet Dev* 16(3), pp. 233-239.

Boisen, K. A. et al. 2001. Are male reproductive disorders a common entity? The testicular dysgenesis syndrome. *Ann N Y Acad Sci* 948, pp. 90-99.

Bray, N. J. et al. 2003. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum Genet* 113(2), pp. 149-153.

Bray, N. J. et al. 2005. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. *Hum Mol Genet* 14(14), pp. 1947-1954.

Brookes, K. et al. 2006a. The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol Psychiatry* 11(10), pp. 934-953.

Brookes, K. J. et al. 2008. Association of the steroid sulfatase (STS) gene with attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1531-1535.

Brookes, K. J. et al. 2006b. A common haplotype of the dopamine transporter gene associated with attention-deficit/hyperactivity disorder and interacting with maternal use of alcohol during pregnancy. *Arch Gen Psychiatry* 63(1), pp. 74-81.

Brown, C. J. and Greally, J. M. 2003. A stain upon the silence: genes escaping X inactivation. *Trends Genet* 19(8), pp. 432-438.

Byerley, W. et al. 1992. Tyrosine hydroxylase gene not linked to manic-depression in seven of eight pedigrees. *Hum Hered* 42(4), pp. 259-263.

Cahill, L. 2006. Why sex matters for neuroscience. *Nat Rev Neurosci* 7(6), pp. 477-484.

Cahill, L. et al. 2004. Sex-related hemispheric lateralization of amygdala function in emotionally influenced memory: an FMRI investigation. *Learn Mem* 11(3), pp. 261-266.

Cantwell, D. P. 1975. Genetics of hyperactivity. *J Child Psychol Psychiatry* 16(3), pp. 261-264.

Capelli, C. et al. 2003. A Y chromosome census of the British Isles. *Curr Biol* 13(11), pp. 979-984.

Cappuccio, F. P. 1997. Ethnicity and cardiovascular risk: variations in people of African ancestry and South Asian origin. *J Hum Hypertens* 11(9), pp. 571-576.

Cardno, A. G. and Gottesman, II 2000. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 97(1), pp. 12-17.

Cardon, L. R. and Bell, J. I. 2001. Association study designs for complex diseases *Nature reviews* 2, pp. 91-99.

Cardon, L. R. and Palmer, L. J. 2003. Population stratification and spurious allelic association. *Lancet* 361(9357), pp. 598-604.

Carrel, L. and Willard, H. F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434(7031), pp. 400-404.

Carruth, L. L. et al. 2002. Sex chromosome genes directly affect brain sexual differentiation. *Nat Neurosci* 5(10), pp. 933-934.

Carvalho, C. M. et al. 2003. Lack of association between Y chromosome haplogroups and male infertility in Japanese men. *Am J Med Genet A* 116A(2), pp. 152-158.

Caspi, A. et al. 2008. A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 65(2), pp. 203-210.

Castellanos, F. X. and Tannock, R. 2002. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nat Rev Neurosci* 3(8), pp. 617-628.

Center, R. et al. 1993. Y chromosome loss and rearrangement in non-small-cell lung cancer. *Int J Cancer* 55(3), pp. 390-393.

Charchar, F. J. et al. 2004. Association of the human Y chromosome with cholesterol levels in the general population. *Arterioscler Thromb Vasc Biol* 24(2), pp. 308-312.

Charchar, F. J. et al. 2002. The Y chromosome effect on blood pressure in two European populations. *Hypertension* 39(2 Pt 2), pp. 353-356.

Charchar, F. J. et al. 2003. Y is there a risk to being male? *Trends Endocrinol Metab* 14(4), pp. 163-168.

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. *Am J Hum Genet* 75(5), pp. 807-821.

Cheuk, D. K. and Wong, V. 2006. Meta-analysis of association between a catechol-O-methyltransferase gene polymorphism and attention deficit hyperactivity disorder. *Behav Genet* 36(5), pp. 651-659.

Chevret, E. et al. 1997. Meiotic behaviour of sex chromosomes investigated by three-colour FISH on 35,142 sperm nuclei from two 47,XYY males. *Hum Genet* 99(3), pp. 407-412.

Choi, B. G. and McLaughlin, M. A. 2007. Why men's hearts break: cardiovascular effects of sex steroids. *Endocrinol Metab Clin North Am* 36(2), pp. 365-377.

Clepet, C. et al. 1993. The human SRY transcript. *Hum Mol Genet* 2(12), pp. 2007-2012.

Comings, D. E. et al. 1995. No association of a tyrosine hydroxylase gene tetranucleotide repeat polymorphism in autism, Tourette syndrome, or ADHD. *Biol Psychiatry* 37(7), pp. 484-486.

Cooke, H. J. and Noel, B. 1979. Confirmation of Y/autosome translocation using recombinant DNA. *Hum Genet* 50(1), pp. 39-44.

Costa e Silva, J. A. 2008. Autism, a brain developmental disorder: some new pathopysiologic and genetics findings. *Metabolism* 57 Suppl 2, pp. S40-43.

Cowell, P. E. et al. 1996. Sex differences in neuroanatomical and clinical correlations in schizophrenia. *Am J Psychiatry* 153(6), pp. 799-805.

Craddock, N. et al. 2006a. Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophr Bull* 32(1), pp. 9-16.

Craddock, N. et al. 2008. Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. *Mol Psychiatry.*

Craddock, N. et al. 2006b. The catechol-O-methyl transferase (COMT) gene as a candidate for psychiatric phenotypes: evidence and lessons. *Mol Psychiatry* 11(5), pp. 446-458.

Crow, T. 2002. Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherinXY to hominid speciation and the aetiology of psychosis. *British Journal of Psychiatry* 181, pp. 295-297.

Crow, T. J. et al. 1996. Cerebral lateralization is delayed in children who later develop schizophrenia. *Schizophr Res* 22(3), pp. 181-185.

Cuffe, S. P. et al. 2001. Prevalence of attention-deficit/hyperactivity disorder in a community sample of older adolescents. *J Am Acad Child Adolesc Psychiatry* 40(9), pp. 1037-1044.

Cuffe, S. P. et al. 2005. Prevalence and correlates of ADHD symptoms in the national health interview survey. *J Atten Disord* 9(2), pp. 392-401.

Curtis, A. L. et al. 2006. Sexually dimorphic responses of the brain norepinephrine system to stress and corticotropin-releasing factor. *Neuropsychopharmacology* 31(3), pp. 544-554.

Dahmen, N. et al. 2005. Tyrosine hydroxylase Val-81-Met polymorphism associated with early-onset alcoholism. *Psychiatr Genet* 15(1), pp. 13-16.

Daly, G. et al. 1999. Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. *Mol Psychiatry* 4(2), pp. 192-196.

David, A. S. et al. 1997. IQ and risk for schizophrenia: a population-based cohort study. *Psychol Med* 27(6), pp. 1311-1323.

Davies, W. et al. 2005a. Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nature Genetics* 37, pp. 625-629.

Davies, W. et al. 2005b. Imprinted gene expression in the brain. *Neuroscience and Biobehavioral Reviews* 29, pp. 421-430.

Davies, W. and Wilkinson, L. 2006. It is not all hormones: Alternative explanations for sexual differentiation of the brain. *Brain Research.*

De Vries, G. J. et al. 2002. A model system for study of sex chromosome effects on sexually dimorphic neural and behavioral traits. *J Neurosci* 22(20), pp. 9005-9014.

Degreef, G. et al. 1992. Volumes of ventricular system subdivisions measured from magnetic resonance images in first-episode schizophrenic patients. *Arch Gen Psychiatry* 49(7), pp. 531-537.

Delbridge, M. L. and Graves, J. A. 1999. Mammalian Y chromosome evolution and the male-specific functions of Y chromosome-borne genes. *Rev Reprod* 4(2), pp. 101-109.

Derks, E. M. et al. 2008. Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. *Behav Genet* 38(1), pp. 11-23.

Dewing, P. et al. 2006. Direct regulation of adult brain function by the male-specific factor SRY. *Curr Biol* 16(4), pp. 415-420.

Dewing, P. et al. 2003. Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. *Brain Res Mol Brain Res* 118(1-2), pp. 82-90.

DiMaio, S. et al. 2003. Dopamine genes and attention-deficit hyperactivity disorder: a review. *J Psychiatry Neurosci* 28(1), pp. 27-38.

Domschke, K. et al. 2007. Meta-analysis of COMT val158met in panic disorder: ethnic heterogeneity and gender specificity. *Am J Med Genet B Neuropsychiatr Genet* 144B(5), pp. 667-673.

Drevets, W. C. 2003. Neuroimaging abnormalities in the amygdala in mood disorders. *Ann N Y Acad Sci* 985, pp. 420-444.

Duan, J. et al. 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 12(3), pp. 205-216.

Durston, S. et al. 2004. Magnetic resonance imaging of boys with attention-deficit/hyperactivity disorder and their unaffected siblings. *J Am Acad Child Adolesc Psychiatry* 43(3), pp. 332-340.

Eastwood, S. L. and Harrison, P. J. 2005. Interstitial white matter neuron density in the dorsolateral prefrontal cortex and parahippocampal gyrus in schizophrenia. *Schizophr Res* 79(2-3), pp. 181-188.

Edinformatics.com, T. I. L. 1999. *Biosynthesis of catecholamines* [Online]. Edinformatics.com. Available at: http://www.edinformatics.com/interactive_molecules/info/adrenaline.htm [Accessed: 7 December 2009]

Elia, J. et al. 2009. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry.*

Ellis, J. A. et al. 2000. Association of the human Y chromosome with high blood pressure in the general population. *Hypertension* 36(5), pp. 731-733.

Ely, D. et al. 1997. Spontaneously hypertensive rat Y chromosome increases indexes of sympathetic nervous system activity. *Hypertension* 29(2), pp. 613-618.

Endicott, J. et al. 1976. The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 33(6), pp. 766-771.

Erickson, R. P. 1997. Does sex determination start at conception? *Bioessays* 19(11), pp. 1027-1032.

Ewis, A. A. et al. 2002. Linkage between prostate cancer incidence and different alleles of the human Y-linked tetranucleotide polymorphism DYS19. *J Med Invest* 49(1-2), pp. 56-60.

Faraone, S. V. et al. 2007. Linkage analysis of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet.*

Faraone, S. V. et al. 2001. Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. *Am J Psychiatry* 158(7), pp. 1052-1057.

Faraone, S. V. et al. 2005. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57(11), pp. 1313-1323.

Feng, Y. et al. 2005. The SNAP25 gene as a susceptibility gene contributing to attention-deficit hyperactivity disorder. *Mol Psychiatry* 10(11), pp. 998-1005, 1973.

Ferreira, M. A. et al. 2008. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 40(9), pp. 1056-1058.

Fisher, S. E. et al. 2002. A genomewide scan for loci involved in attention-deficit/hyperactivity disorder. *Am J Hum Genet* 70(5), pp. 1183-1196.

Flaquer, A. et al. 2008. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet* 16(7), pp. 771-779.

Franke, B. et al. 2009. Genome-wide association studies in ADHD. *Hum Genet* 126(1), pp. 13-50.

Frankel, W. N. and Schork, N. J. 1996. Who's afraid of epistasis? *Nat Genet* 14(4), pp. 371-373.

Gainetdinov, R. R. 2007. Dopamine transporter mutant mice in experimental neuropharmacology. *Naunyn Schmiedebergs Arch Pharmacol.*

Gatewood, J. D. et al. 2006. Sex chromosome complement and gonadal sex influence aggressive and parental behaviors in mice. *J Neurosci* 26(8), pp. 2335-2342.

Gauthier, J. et al. 2005. NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *Am J Med Genet B Neuropsychiatr Genet* 132B(1), pp. 74-75.

Giouzeli, M. et al. 2004. ProtocadherinX/Y, a candidate gene-pair for schizophrenia and schizoaffective disorder: a DHPLC investigation of genomic sequence. *Am J Med Genet B Neuropsychiatr Genet* 129B(1), pp. 1-9.

Gizer, I. R. et al. 2009. Candidate gene studies of ADHD: a meta-analytic review. *Hum Genet* 126(1), pp. 51-90.

Golding, J. 2004. The Avon Longitudinal Study of Parents and Children (ALSPAC)--study design and collaborative opportunities. *Eur J Endocrinol* 151 Suppl 3, pp. U119-123.

Goldstein, J. M. et al. 2001. Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cereb Cortex* 11(6), pp. 490-497.

Goodman, R. et al. 2000. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 41(5), pp. 645-655.

Goodman, R. and Stevenson, J. 1989. A twin study of hyperactivity--II. The aetiological role of genes, family relationships and perinatal adversity. *J Child Psychol Psychiatry* 30(5), pp. 691-709.

Gottesman, I. I. 1991. *Schizophrenia Genesis: The Origins of Madness*. New York: Freeman.

Gotz, M. J. et al. 1999. Criminality and antisocial behaviour in unselected men with sex chromosome abnormalities. *Psychol Med* 29(4), pp. 953-962.

Graves, J. A. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124(5), pp. 901-914.

Green, E. et al. 2009. Bipolar disorder risk allele at CACNA1C also confers risk to recurrent major depression and to schizophrenia. *Molecular Psychiatry* in press.

Grossman, L. S. et al. 2006. Sex differences in outcome and recovery for schizophrenia and other psychotic and nonpsychotic disorders. *Psychiatr Serv* 57(6), pp. 844-850.

Grossman, L. S. et al. 2008. Sex differences in schizophrenia and other psychotic disorders: a 20-year longitudinal study of psychosis and recovery. *Compr Psychiatry* 49(6), pp. 523-529.

Gubbay, J. et al. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346(6281), pp. 245-250.

Guillot, P. V. et al. 1995. Intermale aggression tested in two procedures, using four inbred strains of mice and their reciprocal congenics: Y chromosomal implications. *Behav Genet* 25(4), pp. 357-360.

Guimaraes, A. P. et al. 2007. Serotonin genes and attention deficit/hyperactivity disorder in a Brazilian sample: preferential transmission of the HTR2A 452His allele to affected boys. *Am J Med Genet B Neuropsychiatr Genet* 144(1), pp. 69-73.

Hardy, J. and Singleton, A. 2009. Genomewide association studies and human disease. *N Engl J Med* 360(17), pp. 1759-1768.

Harley, V. R. et al. 2003. The molecular action and regulation of the testis-determining factors, SRY (sex-determining region on the Y chromosome) and SOX9 [SRY-related high-mobility group (HMG) box 9]. *Endocr Rev* 24(4), pp. 466-487.

Harrison, P. J. 1999. The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain* 122 (Pt 4), pp. 593-624.

Harrison, P. J. and Tunbridge, E. M. 2007. Catechol-O-Methyltransferase (COMT): A Gene Contributing to Sex Differences in Brain Function, and to Sexual Dimorphism in the Predisposition to Psychiatric Disorders. *Neuropsychopharmacology*.

Hashimoto, R. et al. 2004. Expression analysis of neuregulin-1 in the dorsolateral prefrontal cortex in schizophrenia. *Mol Psychiatry* 9(3), pp. 299-307.

Hawkins, J. R. et al. 1992. Mutational analysis of SRY: nonsense and missense mutations in XY sex reversal. *Hum Genet* 88(4), pp. 471-474.

Hebebrand, J. et al. 2006. A genome-wide scan for attention-deficit/hyperactivity disorder in 155 German sib-pairs. *Mol Psychiatry* 11(2), pp. 196-205.

Hensbroek, R. A. et al. 1995. Y chromosomal effects on hippocampal mossy fiber distributions in mice selected for aggression. *Brain Res* 682(1-2), pp. 203-206.

Hill, S. K. et al. 2008. Neurocognitive allied phenotypes for schizophrenia and bipolar disorder. *Schizophr Bull* 34(4), pp. 743-759.

Hirschhorn, J. N. and Daly, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6(2), pp. 95-108.

Holden, C. 2005. Sex and the suffering brain. *Science* 308(5728), p. 1574.

Holmans, P. A. et al. 2009. Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. *Mol Psychiatry.*

Holmes, J. et al. 2004. The Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI): reliability and validity. *Br J Psychiatry* 184, pp. 74-78.

Hubbard, T. J. et al. 2009. Ensembl 2009. *Nucleic Acids Res* 37(Database issue), pp. D690-697.

Hunter, S. et al. 1993. Y chromosome loss in esophageal carcinoma: an in situ hybridization study. *Genes Chromosomes Cancer* 8(3), pp. 172-177.

Ingason, A. et al. 2009. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry.*

International HapMap Consortium 2003. The International HapMap Project. *Nature* 426(6968), pp. 789-796.

International HapMap Consortium. [no date]. *International HapMap Project* [Online]. Available at: http://www.hapmap.org/index.html.en [Accessed: 25 June 2009]

Ishiguro, H. et al. 1998. Systematic search for variations in the tyrosine hydroxylase gene and their associations with schizophrenia, affective disorders, and alcoholism. *Am J Med Genet* 81(5), pp. 388-396.

Jacewicz, R. et al. 2006. Will genetic polymorphism of tetranucleotide sequences help in the diagnostics of major psychiatric disorders? *Forensic Sci Int* 162(1-3), pp. 24-27.

Jamain, S. et al. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 34(1), pp. 27-29.

Jamain, S. et al. 2002. Y chromosome haplogroups in autistic subjects. *Mol Psychiatry* 7(2), pp. 217-219.

Jobling, M. A. and Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8), pp. 598-612.

Jobling, M. A. et al. 1998. A selective difference between human Y-chromosomal DNA haplotypes. *Curr Biol* 8(25), pp. 1391-1394.

Jonsson, E. G. et al. 1998. Failure to replicate an association between a rare allele of a tyrosine hydroxylase gene microsatellite and schizophrenia. *Eur Arch Psychiatry Clin Neurosci* 248(2), pp. 61-63.

Justice, A. J. and De Wit, H. 2000. Acute effects of d-amphetamine during the early and late follicular phases of the menstrual cycle in women. *Pharmacol Biochem Behav* 66(3), pp. 509-515.

Kahn, R. S. et al. 2003. Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors. *J Pediatr* 143(1), pp. 104-110.

Karafet, T. M. et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18(5), pp. 830-838.

Katona, C. et al. 2008. *Psychiatry at a glance*. 4th ed. Oxford: Wiley-Blackwell.

Kendler, K. S. et al. 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. *Arch Gen Psychiatry* 51(6), pp. 456-468.

Kent, L. et al. 2008. X linked ichthyosis (steroid sulphatase deficiency) is associated with increased risk of attention deficit hyperactivity disorder, autism and social communication deficits. *J Med Genet.*

Kiecolt-Glaser, J. K. and Newton, T. L. 2001. Marriage and health: his and hers. *Psychol Bull* 127(4), pp. 472-503.

Killgore, W. D. and Yurgelun-Todd, D. A. 2004. Sex-related developmental differences in the lateralized activation of the prefrontal cortex and amygdala during perception of facial affect. *Percept Mot Skills* 99(2), pp. 371-391.

Kinney, D. K. et al. 1994. Season of birth and obstetrical complications in schizophrenics. *J Psychiatr Res* 28(6), pp. 499-509.

Kirov, G. et al. 2009. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18(8), pp. 1497-1503.

Kirov, G. et al. 2008a. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* 17(3), pp. 458-465.

Kirov, G. et al. 2004. Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry* 55(10), pp. 971-975.

Kirov, G. et al. 2005. Finding schizophrenia genes. *The journal of clinical investigation* 115(6), pp. 1440-1448.

Kirov, G. et al. 2008b. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry.*

Kittles, R. A. et al. 1999. Cladistic association analysis of Y chromosome effects on alcohol dependence and related personality traits. *Proc Natl Acad Sci U S A* 96(7), pp. 4204-4209.

Klein, R. G. and Mannuzza, S. 1991. Long-term outcome of hyperactive children: a review. *J Am Acad Child Adolesc Psychiatry* 30(3), pp. 383-387.

Kopsida, E. 2009 Personal communication. *Cognitive effects of Sry-knockdown in the brain.*

Krausz, C. et al. 2004. Y chromosome polymorphisms in medicine. *Ann Med* 36(8), pp. 573-583.

Krausz, C. et al. 2001. Identification of a Y chromosome haplogroup associated with reduced sperm counts. *Hum Mol Genet* 10(18), pp. 1873-1877.

Kren, V. et al. 2001. Y-chromosome transfer induces changes in blood pressure and blood lipids in SHR. *Hypertension* 37(4), pp. 1147-1152.

Kuroki, Y. et al. 1999. Spermatogenic ability is different among males in different Y chromosome lineage. *J Hum Genet* 44(5), pp. 289-292.

Kurumaji, A. et al. 2001. An association of the polymorphic repeat of tetranucleotide (TCAT) in the first intron of the human tyrosine hydroxylase gene with schizophrenia in a Japanese sample. *J Neural Transm* 108(4), pp. 489-495.

Lachman, H. M. et al. 1996. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics* 6(3), pp. 243-250.

Lahr, G. et al. 1995. Transcription of the Y chromosomal gene, Sry, in adult mouse brain. *Brain Res Mol Brain Res* 33(1), pp. 179-182.

Langley, K. et al. 2005. Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour. A review. *Minerva Pediatr* 57(6), pp. 359-371.

Langley, K. et al. 2008. Testing for gene x environment interaction effects in attention deficit hyperactivity disorder and associated antisocial behavior. *Am J Med Genet B Neuropsychiatr Genet* 147B(1), pp. 49-53.

Larsson, H. et al. 2006. Genetic contributions to the development of ADHD subtypes from childhood to adolescence. *J Am Acad Child Adolesc Psychiatry* 45(8), pp. 973-981.

Lasky-Su, J. et al. 2008. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1345-1354.

Laumonnier, F. et al. 2004. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 74(3), pp. 552-557.

Lawrie, S. M. and Abukmeil, S. S. 1998. Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. *Br J Psychiatry* 172, pp. 110-120.

Lencz, T. et al. 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 12(6), pp. 572-580.

Lesch, K. P. et al. Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. *Mol Psychiatry.*

Lesch, K. P. et al. 2008. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm* 115(11), pp. 1573-1585.

Lewis, C. M. et al. 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 73(1), pp. 34-48.

Li, D. et al. 2006. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). *Hum Mol Genet* 15(14), pp. 2276-2284.

Li, X. M. et al. 1996. Characterization of the promoter region of human steroid sulfatase: a gene which escapes X inactivation. *Somat Cell Mol Genet* 22(2), pp. 105-117.

Libioulle, C. et al. 2007. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 3(4), p. e58.

Lichtenstein, P. et al. 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 373(9659), pp. 234-239.

Linnet, K. M. et al. 2003. Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. *Am J Psychiatry* 160(6), pp. 1028-1040.

Liu, P. Y. et al. 2005. A survey of haplotype variants at several disease candidate genes: the importance of rare variants for complex diseases. *J Med Genet* 42(3), pp. 221-227.

Lockshin, M. D. 2006. Sex differences in autoimmune disease. *Lupus* 15(11), pp. 753-756.

Lowe, N. et al. 2004. Joint analysis of the DRD5 marker concludes association with attention-deficit/hyperactivity disorder confined to the predominantly inattentive and combined subtypes. *Am J Hum Genet* 74(2), pp. 348-356.

Luo, X. et al. 1995. A cell-specific nuclear receptor plays essential roles in adrenal and gonadal development. *Endocr Res* 21(1-2), pp. 517-524.

Madeira, M. D. and Lieberman, A. R. 1995. Sexual dimorphism in the mammalian limbic system. *Prog Neurobiol* 45(4), pp. 275-333.

Mah, S. et al. 2006. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry* 11(5), pp. 471-478.

Maher, B. S. et al. 2002. Dopamine system genes and attention deficit hyperactivity disorder: a meta-analysis. *Psychiatr Genet* 12(4), pp. 207-215.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet* 24(3), pp. 133-141.

Maxson, S. C. 1996. Searching for candidate genes with effects on an agonistic behavior, offense, in mice. *Behav Genet* 26(5), pp. 471-476.

Maxson, S. C. et al. 1979. Interaction of Y-chromosomal and autosomal gene(s) in the development of intermale aggression in mice. *Behav Genet* 9(3), pp. 219-226.

Mayer, A. et al. 1998. The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. *Neurogenetics* 1(4), pp. 281-288.

Mayer, A. et al. 2000. Developmental profile of Sry transcripts in mouse brain. *Neurogenetics* 3(1), pp. 25-30.

McCarthy, M. I. and Hirschhorn, J. N. 2008. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 17(R2), pp. R156-165.

McCarthy, M. I. and Zeggini, E. 2009. Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9(2), pp. 164-171.

McCarthy, S. E. et al. 2009. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 41(11), pp. 1223-1227.

McCarty, B. M. et al. 2006. Medical and psychosexual outcome in women affected by complete gonadal dysgenesis. *J Pediatr Endocrinol Metab* 19(7), pp. 873-877.

McEwen, B. S. 2000. The neurobiology of stress: from serendipity to clinical relevance. *Brain Res* 886(1-2), pp. 172-189.

McFadyen-Ketchum, S. A. et al. 1996. Patterns of change in early childhood aggressive-disruptive behavior: gender differences in predictions from early coercive and affectionate mother-child interactions. *Child Dev* 67(5), pp. 2417-2433.

McGaugh, J. L. 2004. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu Rev Neurosci* 27, pp. 1-28.

McGrath, J. et al. 2004. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* 2, p. 13.

McLoughlin, G. et al. 2007. Genetic support for the dual nature of attention deficit hyperactivity disorder: substantial genetic overlap between the inattentive and hyperactive-impulsive components. *J Abnorm Child Psychol* 35(6), pp. 999-1008.

Mefford, H. C. et al. 2008. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359(16), pp. 1685-1699.

Meloni, R. et al. 2002. Post-genomic era and gene discovery for psychiatric diseases: there is a new art of the trade? The example of the HUMTH01 microsatellite in the Tyrosine Hydroxylase gene. *Mol Neurobiol* 26(2-3), pp. 389-403.

Mill, J. and Petronis, A. 2008. Pre- and peri-natal environmental risks for attention-deficit hyperactivity disorder (ADHD): the potential role of epigenetic processes in mediating susceptibility. *J Child Psychol Psychiatry* 49(10), pp. 1020-1030.

Mill, J. et al. 2005. Quantitative trait locus analysis of candidate gene alleles associated with attention deficit hyperactivity disorder (ADHD) in five genes: DRD4, DAT1, DRD5, SNAP-25, and 5HT1B. *Am J Med Genet B Neuropsychiatr Genet* 133B(1), pp. 68-73.

Millstein, R. et al. 1997. Presenting ADHD symptoms and subtypes in clinically referred adults with ADHD. *J. Attent. Disord* (2), pp. 159-166.

Milsted, A. et al. 2004. Regulation of tyrosine hydroxylase gene transcription by Sry. *Neurosci Lett* 369(3), pp. 203-207.

Mittwoch, U. 2000. Genetics of sex determination: exceptions that prove the rule. *Mol Genet Metab* 71(1-2), pp. 405-410.

Moffitt, T. E. 1993. Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100(4), pp. 674-701.

Mors, O. et al. 2001. No evidence of increased risk for schizophrenia or bipolar affective disorder in persons with aneuploidies of the sex chromosomes. *Psychol Med* 31(3), pp. 425-430.

Mueller, J. C. 2004. Linkage disequilibrium for different scales and applications. *Brief Bioinform* 5(4), pp. 355-364.

Mulligan, A. et al. 2008. A Case of ADHD and a Major Y Chromosome Abnormality. *J Atten Disord.*

Murphy, K. C. and Owen, M. J. 2001. Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. *Br J Psychiatry* 179, pp. 397-402.

Myers, A. J. et al. 2007. A survey of genetic human cortical gene expression. *Nat Genet* 39(12), pp. 1494-1499.

Nagatsu, T. and Ichinose, H. 1991. Comparative studies on the structure of human tyrosine hydroxylase with those of the enzyme of various mammals. *Comp Biochem Physiol* 98, pp. 203-210.

Naoi, M. and Parvez, S. 1993. *Tyrosine Hydroxylase: From Discovery to Cloning.* Utrecht: VSR BP.

Neale, B. 2009 personal communication. *GWAS of ADHD using IMAGE II.*

Neale, B. M. et al. 2008. Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1337-1344.

Need, A. C. et al. 2009. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5(2), p. e1000373.

Negrin, C. D. et al. 2001. Reciprocal consomic strains to evaluate y chromosome effects. *Hypertension* 37(2 Part 2), pp. 391-397.

Neuman, R. J. et al. 2007. Prenatal smoking exposure and dopaminergic genotypes interact to cause a severe ADHD subtype. *Biol Psychiatry* 61(12), pp. 1320-1328.

Ng, M. Y. et al. 2008. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry.*

Nguyen, D. K. and Disteche, C. M. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet* 38(1), pp. 47-53.

Nigg, J. T. et al. 2005. Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? *Biol Psychiatry* 57(11), pp. 1224-1230.

Nishizawa, S. et al. 1997. Differences between males and females in rates of serotonin synthesis in human brain. *Proc Natl Acad Sci U S A* 94(10), pp. 5308-5313.

Nopoulos, P. et al. 1997. Sex differences in brain morphology in schizophrenia. *Am J Psychiatry* 154(12), pp. 1648-1654.

O'Donovan, M. C. et al. 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet.*

Ober, C. et al. 2008. Sex-specific genetic architecture of human disease. *Nat Rev Genet* 9(12), pp. 911-922.

Office of Population Censures and Surveys 1995. *Standard Occupational Classification.* 2nd ed. London: HMSO.

Ogdie, M. N. et al. 2006. Pooled genome-wide linkage data on 424 ADHD ASPs suggests genetic heterogeneity and a common risk locus at 5p13. *Mol Psychiatry* 11(1), pp. 5-8.

Ogdie, M. N. et al. 2004. Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. *Am J Hum Genet* 75(4), pp. 661-668.

Ogdie, M. N. et al. 2003. A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: suggestive linkage on 17p11. *Am J Hum Genet* 72(5), pp. 1268-1279.

Oota, H. et al. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat Genet* 29(1), pp. 20-21.

Ota, M. et al. 2001. Exon 3 of tyrosine hydroxylase gene: lack of association with Japanese schizophrenic patients. *Mol Psychiatry* 6(3), pp. 315-319.

Owen, M. J. et al. 2005. Schizophrenia: genes at last? *Trends Genet* 21(9), pp. 518-525.

Owen, M. J. et al. 2002. Schizophrenia. In: McGuffin, P. et al. eds. *Psychiatric Genetics and Genomics*. New York: Oxford University Press.

Owen, M. J. et al. 2009. Schizophrenia genetics: advancing on two fronts. *Curr Opin Genet Dev.*

Packard, M. G. et al. 1996. Posttraining intrahippocampal estradiol injections enhance spatial memory in male rats: interaction with cholinergic systems. *Behav Neurosci* 110(3), pp. 626-632.

Pae, C. U. et al. 2003. VNTR polymorphism of tyrosine hydroxylase gene and schizophrenia in the Korean population. *Neuropsychobiology* 47(3), pp. 131-136.

Paracchini, S. et al. 2003. A Y chromosomal influence on prostate cancer risk: the multi-ethnic cohort study. *J Med Genet* 40(11), pp. 815-819.

Payton, A. et al. 2001. Examining for association between candidate gene polymorphisms in the dopamine pathway and attention-deficit hyperactivity disorder: a family-based study. *Am J Med Genet* 105(5), pp. 464-470.

Pe'er, I. et al. 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 32(4), pp. 381-385.

Pearlson, G. D. 2000. Neurobiology of schizophrenia. *Ann Neurol* 48(4), pp. 556-566.

Pearlson, G. D. and Marsh, L. 1999. Structural brain imaging in schizophrenia: a selective review. *Biol Psychiatry* 46(5), pp. 627-649.

Pecon-Slattery, J. et al. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *Proc Natl Acad Sci U S A* 97(10), pp. 5307-5312.

Pliszka, S. R. 2000. Patterns of psychiatric comorbidity with attention-deficit/hyperactivity disorder. *Child Adolesc Psychiatr Clin N Am* 9(3), pp. 525-540, vii.

Plomin, R. et al. 2008. *Behavioural Genetics*. 5th ed. Worth Publishers.

Polanco, J. C. and Koopman, P. 2006. Sry and the hesitant beginnings of male development. *Dev Biol.*

Pooley, E. C. et al. 2007. The met(158) allele of catechol-O-methyltransferase (COMT) is associated with obsessive-compulsive disorder in men: case-control study and meta-analysis. *Mol Psychiatry* 12(6), pp. 556-561.

Postma, D. S. 2007. Gender differences in asthma development and progression. *Gend Med* 4 Suppl B, pp. S133-146.

Poulat, F. et al. 1995. Nuclear localization of the testis determining gene product SRY. *J Cell Biol* 128(5), pp. 737-748.

Preston, N. J. et al. 2002. Gender differences in premorbid adjustment of patients with first episode psychosis. *Schizophr Res* 55(3), pp. 285-290.

Previdere, C. et al. 1999. Y-chromosomal DNA haplotype differences in control and infertile Italian subpopulations. *Eur J Hum Genet* 7(6), pp. 733-736.

Purcell, S. et al. 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19(1), pp. 149-150.

Purcell, S. et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3), pp. 559-575.

Purper-Ouakil, D. et al. 2005. Meta-analysis of family-based association studies between the dopamine transporter gene and attention deficit hyperactivity disorder. *Psychiatr Genet* 15(1), pp. 53-59.

Qian, Q. et al. 2003. Family-based and case-control association studies of catechol-O-methyltransferase in attention deficit hyperactivity disorder suggest genetic sexual dimorphism. *Am J Med Genet B Neuropsychiatr Genet* 118B(1), pp. 103-109.

Quintana-Murci, L. et al. 2003. Y chromosome haplotypes and testicular cancer in the English population. *J Med Genet* 40(3), p. e20.

Raab, H. et al. 1995. Effects of sex and estrogen on tyrosine hydroxylase mRNA in cultured embryonic rat mesencephalon. *Brain Res Mol Brain Res* 33(1), pp. 157-164.

Reich, D. E. and Lander, E. S. 2001. On the allelic spectrum of human disease. *Trends Genet* 17(9), pp. 502-510.

Repping, S. et al. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 38(4), pp. 463-467.

Risch, N. 1990. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46(2), pp. 229-241.

Rodriguez, S. et al. 2005. Non-recombining chromosome Y haplogroups and centromeric HindIII RFLP in relation to blood pressure in 2,743 middle-aged Caucasian men from the UK. *Hum Genet* 116(4), pp. 311-318.

Rodriguez, S. et al. 2004. Haplotypic analyses of the IGF2-INS-TH gene cluster in relation to cardiovascular risk traits. *Hum Mol Genet* 13(7), pp. 715-725.

Rodriguez, S. et al. 2006. A study of TH01 and IGF2-INS-TH haplotypes in relation to smoking initiation in three independent surveys. *Pharmacogenet Genomics* 16(1), pp. 15-23.

Romanos, M. et al. 2008. Genome-wide linkage analysis of ADHD using high-density SNP arrays: novel loci at 5q13.1 and 14q12. *Mol Psychiatry* 13(5), pp. 522-530.

Rommelse, N. N. et al. 2008. Differential association between MAOA, ADHD and neuropsychological functioning in boys and girls. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1524-1530.

Ronald, A. et al. 2008. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *J Child Psychol Psychiatry* 49(5), pp. 535-542.

Ropers, H. H. and Hamel, B. C. 2005. X-linked mental retardation. *Nat Rev Genet* 6(1), pp. 46-57.

Rosler, M. et al. 2004. Prevalence of attention deficit-/hyperactivity disorder (ADHD) and comorbid disorders in young male prison inmates. *Eur Arch Psychiatry Clin Neurosci* 254(6), pp. 365-371.

Ross, J. et al. 2006. Cognition and the sex chromosomes: studies in Turner syndrome. *Horm Res* 65(1), pp. 47-56.

Ross, M. T. et al. 2005. The DNA sequence of the human X chromosome. *Nature* 434(7031), pp. 325-337.

Rosser, Z. H. et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67(6), pp. 1526-1543.

Rozen, S. et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423(6942), pp. 873-876.

Rucker, B. et al. 2004. Inhibitory avoidance task reveals differences in ectonucleotidase activities between male and female rats. *Neurochem Res* 29(12), pp. 2231-2237.

Rujescu, D. et al. 2009. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 18(5), pp. 988-996.

Russell, H. F. et al. 2006. Increased prevalence of ADHD in Turner syndrome with no evidence of imprinting effects. *J Pediatr Psychol* 31(9), pp. 945-955.

Russell, V. A. 2007. Neurobiology of animal models of attention-deficit hyperactivity disorder. *J Neurosci Methods* 161(2), pp. 185-198.

Russo, P. et al. 2008. Genetic variants of Y chromosome are associated with a protective lipid profile in black men. *Arterioscler Thromb Vasc Biol* 28(8), pp. 1569-1574.

Rutter, M. 2007. Gene-environment interdependence. *Dev Sci* 10(1), pp. 12-18.

Rutter, M. et al. 2008. *Rutter's Child and Adolescent Psychiatry*. 5th ed. Wiley-Blackwell, p. 1248.

Rutter, M. et al. 2003. Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *J Child Psychol Psychiatry* 44(8), pp. 1092-1115.

Sauter, G. et al. 1995. Y chromosome loss detected by FISH in bladder cancer. *Cancer Genet Cytogenet* 82(2), pp. 163-169.

Scherer, S. W. et al. 2007. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 39(7 Suppl), pp. S7-15.

Schizophrenia Research Forum. 2005. *SchizophreniaGene (SZGene) Published Candidate Genes for Schizophrenia* [Online]. Providence: Available at: http://www.schizophreniaforum.org/res/sczgene/default.asp [Accessed: 25 July 2009]

Scholz, B. et al. 2006. Sex-dependent gene expression in early brain development of chicken embryos. *BMC Neurosci* 7, p. 12.

Schwab, S. G. et al. 2003. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* 72(1), pp. 185-190.

Schweitzer, J. B. et al. 2003. A positron emission tomography study of methylphenidate in adults with ADHD: alterations in resting blood flow and predicting treatment response. *Neuropsychopharmacology* 28(5), pp. 967-973.

Sebat, J. et al. 2009. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet* 25(12), pp. 528-535.

Seidman, L. J. et al. 2005. Structural brain imaging of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57(11), pp. 1263-1272.

Sekido, R. and Lovell-Badge, R. 2009. Sex determination and SRY: down to a wink and a nudge? *Trends Genet* 25(1), pp. 19-29.

Shah, S. S. et al. 2009. Y haplogroups and aggressive behavior in a Pakistani ethnic group. *Aggress Behav* 35(1), pp. 68-74.

Sham, P. and McGuffin, P. 2002. Linkage and Association. In: McGuffin, P. et al. eds. *Psychiatric Genetics and Genomics*. Oxford: Oxford University Press, pp. 55-76.

Shaw, P. et al. 2007. Polymorphisms of the dopamine D4 receptor, clinical outcome, and cortical structure in attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 64(8), pp. 921-931.

Shen, P. et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci U S A* 97(13), pp. 7354-7359.

Shi, J. et al. 2009. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*.

Shifman, S. et al. 2002. A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet* 71(6), pp. 1296-1302.

Shifman, S. et al. 2008. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 4(2), p. e28.

Sinclair, A. H. et al. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346(6281), pp. 240-244.

Siva, N. 2008. 1000 Genomes project. *Nat Biotechnol* 26(3), p. 256.

Skaletsky, H. et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942), pp. 825-837.

Skuse, D. H. 1999. Genomic imprinting of the X chromosome: a novel mechanism for the evolution of sexual dimorphism. *J Lab Clin Med* 133(1), pp. 23-32.

Skuse, D. H. 2000. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatr Res* 47(1), pp. 9-16.

Skuse, D. H. 2005. X-linked genes and mental functioning. *Hum Mol Genet* 14 Spec No 1, pp. R27-32.

Skuse, D. H. et al. 1997. Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* 387(6634), pp. 705-708.

Skuse, D. H. et al. 2005. Functional dissociation of amygdala-modulated arousal and cognitive appraisal, in Turner syndrome. *Brain* 128(Pt 9), pp. 2084-2096.

Sluyter, F. et al. 1995. Autosomal and Y chromosomal effects on the stereotyped response to apomorphine in wild house mice. *Pharmacol Biochem Behav* 52(1), pp. 17-22.

Smalley, S. L. et al. 2002. Genetic linkage of attention-deficit/hyperactivity disorder on chromosome 16p13, in a region implicated in autism. *Am J Hum Genet* 71(4), pp. 959-963.

Solari, A. J. and Rey Valzacchi, G. 1997. The prevalence of a YY synaptonemal complex over XY synapsis in an XYY man with exclusive XYY spermatocytes. *Chromosome Res* 5(7), pp. 467-474.

Sonuga-Barke, E. J. et al. 2008. Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1359-1368.

Spencer, T. J. et al. 2005. In vivo neuroreceptor imaging in attention-deficit/hyperactivity disorder: a focus on the dopamine transporter. *Biol Psychiatry* 57(11), pp. 1293-1300.

Sprich, S. et al. 2000. Adoptive and biological families of children and adolescents with ADHD. *J Am Acad Child Adolesc Psychiatry* 39(11), pp. 1432-1437.

St Clair, D. et al. 2005. Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. *JAMA* 294(5), pp. 557-562.

Stefansson, H. et al. 2009. Common variants conferring risk of schizophrenia. *Nature*.

Stefansson, H. et al. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature*.

Stefansson, H. et al. 2002. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 71(4), pp. 877-892.

Stein, M. B. et al. 2005. COMT polymorphisms and anxiety-related personality traits. *Neuropsychopharmacology* 30(11), pp. 2092-2102.

Strachan, T. and Reed, A. 2003. *Human Molecular Genetics*. New Delhi: Garland Publishing.

Stranger, B. E. et al. 2007. Population genomics of human gene expression. *Nat Genet* 39(10), pp. 1217-1224.

Straub, R. E. et al. 2002. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet* 71(2), pp. 337-348.

Sullivan, P. F. et al. 2008. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 13(6), pp. 570-584.

Susser, E. et al. 1996. Schizophrenia after prenatal famine. Further evidence. *Arch Gen Psychiatry* 53(1), pp. 25-31.

Swain, A. et al. 1998. Dax1 antagonizes Sry action in mammalian sex determination. *Nature* 391(6669), pp. 761-767.

Talbot, K. et al. 2004. Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest* 113(9), pp. 1353-1363.

Talebizadeh, Z. et al. 2004. Do known mutations in neuroligin genes (NLGN3 and NLGN4) cause autism? *J Autism Dev Disord* 34(6), pp. 735-736.

Taylor, E. et al. 1996. Hyperactivity and conduct problems as risk factors for adolescent development. *J Am Acad Child Adolesc Psychiatry* 35(9), pp. 1213-1226.

Taylor, E. and Sonuga-Barke, E. 2008. Disorders of attention and activity. In: Rutter, M. et al. eds. *Rutter's Child and Adolescent Psychiatry*. 5th ed. Malden: Wiley-Blackwell.

Temple, C. M. and Sanfilippo, P. M. 2003. Executive skills in Klinefelter's syndrome. *Neuropsychologia* 41(11), pp. 1547-1559.

Thapar, A. et al. 2007a. Do intrauterine or genetic influences explain the foetal origins of chronic disease? A novel experimental method for disentangling effects. *BMC Med Res Methodol* 7(1), p. 25.

Thapar, A. et al. 2007b. The contribution of gene-environment interaction to psychopathology. *Dev Psychopathol* 19(4), pp. 989-1004.

Thapar, A. et al. 2001. Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. *Br J Psychiatry* 179, pp. 224-229.

Thapar, A. et al. 1999. Genetic basis of attention deficit and hyperactivity. *Br J Psychiatry* 174, pp. 105-111.

Thapar, A. et al. 2007c. Gene-environment interplay in attention-deficit hyperactivity disorder and the importance of a developmental perspective. *Br J Psychiatry* 190, pp. 1-3.

Thapar, A. et al. 2005. Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 62(11), pp. 1275-1278.

Thapar, A. et al. 2006. Refining the attention deficit hyperactivity disorder phenotype for molecular genetic studies. *Mol Psychiatry* 11(8), pp. 714-720.

Thapar, A. et al. 2007d. Advances in genetic findings on attention deficit hyperactivity disorder. *Psychol Med*, pp. 1-12.

Thapar, A. and Rutter, M. 2009. Do prenatal risk factors cause psychiatric disorder? Be wary of causal claims. *Br J Psychiatry* 195(2), pp. 100-101.

Thapar, A. and Stergiakouli, E. 2008. Genetic influences on the development of childhood psychiatric diosrders. *Psychiatry* 7(7), pp. 277-281.

The International Schizophrenia Consortium 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455(7210), pp. 237-241.

The International Schizophrenia Consortium 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*.

The Y Chromosome Consortium 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12(2), pp. 339-348.

Thomson, R. et al. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* 97(13), pp. 7360-7365.

Todd, R. D. et al. 2008. Predictors of stability of attention-deficit/hyperactivity disorder subtypes from childhood to young adulthood. *J Am Acad Child Adolesc Psychiatry* 47(1), pp. 76-85.

Toot, J. et al. 2004. The SHR Y-chromosome increases testosterone and aggression, but decreases serotonin as compared to the WKY Y-chromosome in the rat model. *Behav Genet* 34(5), pp. 515-524.

Tordjman, S. et al. 1995. Linkage between brain serotonin concentration and the sex-specific part of the Y-chromosome in mice. *Neurosci Lett* 183(3), pp. 190-192.

Tranel, D. et al. 2005. Does gender play a role in functional asymmetry of ventromedial prefrontal cortex? *Brain* 128(Pt 12), pp. 2872-2881.

Turner, M. E. et al. 2007. Genomic and expression analysis of multiple Sry loci from a single Rattus norvegicus Y chromosome. *BMC Genet* 8, p. 11.

Uehara, Y. et al. 1998. A hypertensive father, but not hypertensive mother, determines blood pressure in normotensive male offspring through body mass index. *J Hum Hypertens* 12(7), pp. 441-445.

Uhl, G. R. et al. 2008. Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Ann N Y Acad Sci* 1141, pp. 318-381.

Underhill, P. A. et al. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7(10), pp. 996-1005.

Underhill, P. A. et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* 26(3), pp. 358-361.

Vaidya, C. J. et al. 2005. Altered neural substrates of cognitive control in childhood ADHD: evidence from functional magnetic resonance imaging. *Am J Psychiatry* 162(9), pp. 1605-1613.

Van Snellenberg, J. X. and de Candia, T. 2009. Meta-analytic evidence for familial coaggregation of schizophrenia and bipolar disorder. *Arch Gen Psychiatry* 66(7), pp. 748-755.

Veijola, J. et al. 1998. Sex differences in the association between childhood experiences and adult depression. *Psychol Med* 28(1), pp. 21-27.

Verbeek, M. M. et al. 2007. Mutations in the cyclic adenosine monophosphate response element of the tyrosine hydroxylase gene. *Ann Neurol* 62(4), pp. 422-426.

Veyrunes, F. et al. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* 18(6), pp. 965-973.

Vidal, V. P. et al. 2001. Sox9 induces testis development in XX transgenic mice. *Nat Genet* 28(3), pp. 216-217.

Vincent, J. B. et al. 2004. Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. *Am J Med Genet B Neuropsychiatr Genet* 129B(1), pp. 82-84.

Volkow, N. D. and Swanson, J. M. 2003. Variables that affect the clinical use and abuse of methylphenidate in the treatment of ADHD. *Am J Psychiatry* 160(11), pp. 1909-1918.

Wade, J. and Arnold, A. P. 1996. Functional testicular tissue does not masculinize development of the zebra finch song system. *Proc Natl Acad Sci U S A* 93(11), pp. 5264-5268.

Walsh, T. et al. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320(5875), pp. 539-543.

Walther, D. J. et al. 2003. Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* 299(5603), p. 76.

Wang, W. et al. 2005. Genome-wide association studies: theoritical and practical concerns. *Nature reviews genetics* 6, pp. 109-118.

Weale, M. E. et al. 2002. Y chromosome evidence for Anglo-Saxon mass migration. *Mol Biol Evol* 19(7), pp. 1008-1021.

Wechsler, D. 1992. *Wechsler Intelligence scale for Children: UK 3rd edition.* London: Psychological Corporation.

Weickert, C. S. et al. 2004. Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain. *Arch Gen Psychiatry* 61(6), pp. 544-555.

Weiss, L. A. 2009. Autism genetics: emerging data from genome-wide copy-number and single nucleotide polymorphism scans. *Expert Rev Mol Diagn* 9(8), pp. 795-803.

Weiss, L. A. et al. 2006. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 38(2), pp. 218-222.

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.

Willcutt, E. G. et al. 2005. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol Psychiatry* 57(11), pp. 1336-1346.

Willcutt, E. G. et al. 2001. A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *J Abnorm Psychol* 110(1), pp. 157-172.

Willcutt, E. G. et al. 2000. Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet* 96(3), pp. 293-301.

Williams, H. J. et al. 2009a. New findings from genetic association studies of schizophrenia. *J Hum Genet* 54(1), pp. 9-14.

Williams, H. J. et al. 2009b. Schizophrenia genetics: new insights from new approaches. *Br Med Bull.*

Williams, N. 2009 Personal Communication. Rare chromosomal deletions and duplications are associated with attention deficit hyperactivity disorder and overlap with those conferring susceptibility to autism and schizophrenia.

Williams, N. M. et al. 2005. Is the dysbindin gene (DTNBP1) a susceptibility gene for schizophrenia? *Schizophr Bull* 31(4), pp. 800-805.

Wilson, J. F. et al. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci U S A* 98(9), pp. 5078-5083.

Wilson, N. D. et al. 2006. PCDH11 is X/Y homologous in Homo sapiens but not in Gorilla gorilla and Pan troglodytes. *Cytogenet Genome Res* 114(2), pp. 137-139.

Wing, J. K. et al. 1990. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 47(6), pp. 589-593.

WJ, K. et al. 2002. *The human genome browser at UCSC* [Online]. Available at: http://genome.ucsc.edu/ [Accessed: 23 July 2009]

Wolfs, M. G. et al. 2009. Type 2 Diabetes Mellitus: New Genetic Insights will Lead to New Therapeutics. *Curr Genomics* 10(2), pp. 110-118.

Woods, S. C. et al. 2003. Gender differences in the control of energy homeostasis. *Exp Biol Med (Maywood)* 228(10), pp. 1175-1180.

World Health Organisation 1993. *International Statistical Classification of Diseases and Related Health Problems*
10th Revision Geneva: World Health Organisation.

Xu, B. et al. 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40(7), pp. 880-885.

Xu, J. et al. 2002. Sex differences in sex chromosome gene expression in mouse brain. *Hum Mol Genet* 11(12), pp. 1409-1419.

Xu, J. and Disteche, C. M. 2006. Sex differences in brain expression of X- and Y-linked genes. *Brain Res* 1126(1), pp. 50-55.

Xu, J. et al. 2006. Sexually dimorphic expression of the X-linked gene Eif2s3x mRNA but not protein in mouse brain. *Gene Expr Patterns* 6(2), pp. 146-155.

Yang, B. et al. 2007. A meta-analysis of association studies between the 10-repeat allele of a VNTR polymorphism in the 3'-UTR of dopamine transporter gene and attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 144B(4), pp. 541-550.

Yang, X. et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res* 16(8), pp. 995-1004.

Yao, H. H. et al. 2002. The battle of the sexes: opposing pathways in sex determination. *Novartis Found Symp* 244, pp. 187-198; discussion 198-206, 253-187.

Yen, P. H. et al. 1988. The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. *Cell* 55(6), pp. 1123-1135.

Ylisaukko-oja, T. et al. 2005. Analysis of four neuroligin genes as candidates for autism. *Eur J Hum Genet* 13(12), pp. 1285-1292.

Yoshitsugu, K. et al. 2003. Fine mapping of an isodicentric Y chromosomal breakpoint from a schizophrenic patient. *Am J Med Genet B Neuropsychiatr Genet* 116B(1), pp. 27-31.

Zechner, U. et al. 2001. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet* 17(12), pp. 697-701.

Zhou, K. et al. 2008. Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 147B(8), pp. 1392-1398.

Zuo, L. et al. 2007. CNR1 variation modulates risk for drug and alcohol dependence. *Biol Psychiatry* 62(6), pp. 616-626.

Appendix I – Quality Control information

Standard quality control procedures were followed in all samples and SNPs genotyped. First, SNPs with a call rate <95% were excluded from further analysis. In addition, the difference in the genotyping rate for each SNP between cases and controls was assessed to exclude possible sources of bias. Individuals were only included in analysis if they had genotypes for over 85% of the SNPs analysed. Finally, Hardy-Weinberg (HW) equilibrium was tested for all markers in both cases and controls. Y chromosome SNPs were not tested for HW equilibrium, since they are inherited only from one parent. All quality control tests were performed using PLINK 1.06 (Purcell et al. 2007).

Y chromosome SNPs

14 SNPs across the Y chromosome were chosen to genotype in the ADHD and schizophrenia cases and 1958 Birth Cohort control samples. Two SNPs completely failed to provide genotypes in the genotyping platform and we excluded them from any further analysis. Three SNPs (M45, M74 and M231) were excluded from further analysis for both the ADHD and the schizophrenia case-control association study due to low genotyping rates. The genotyping rates for these SNPs were also significantly different from the controls. In addition, 6 individuals from the ADHD sample and 13 individuals from the schizophrenia sample were removed prior to association analysis due to having genotypes for <85% of the SNPs analysed.

Table 1: Quality control measures for the ADHD case-control association study using 12 Y chromosome SNPs

Y chromosome SNPs	Call rate (cases) %	Call rate (controls) %	p value for difference in call rates
M9	99.5	99.9	0.26
M45	75.7	68.1	0.0058*
M74	85.6	92.9	0.00061**
M89	100.0	99.4	0.34
M153	100.0	99.9	1
M170	98.8	99.1	0.57
M173	99.1	99.3	0.649
M201	98.4	99.1	0.33
M231	93.2	97.7	0.00042**
M304	99.3	99.1	0.65
M405	99.5	99.9	0.26
M410	100.0	100.0	1

*: significant at the $p \leq 0.01$ level; **: significant at the $p \leq 0.001$ level

Table 2: Quality control measures for the schizophrenia case-control association study using 12 Y chromosome SNPs

Y chromosome SNPs	Call rate (cases) %	Call rate (controls) %	p value for difference in call rates
M9	99.9	99.9	1
M45	76.7	68.1	0.0064*
M74	87.6	92.9	0.0008**
M89	99.5	99.4	0.16
M153	99.9	99.9	1
M170	99.9	99.1	0.57
M173	99.6	99.3	0.71
M201	98.2	99.1	0.22
M231	94.1	97.7	0.00065**
M304	99.8	99.1	0.59
M405	100.0	99.9	1
M410	100.0	100.0	1

*: significant at the $p \leq 0.01$ level; **: significant at the $p \leq 0.001$ level

TH SNPs

14 SNPs across the TH gene and surrounding region were chosen to genotype in the ADHD and schizophrenia cases and 1958 Birth Cohort control samples. Two SNPs (rs6356 and rs7950050) completely failed to provide genotypes in the genotyping platform and we excluded them from any further analysis. Two SNPs (rs11042982 and rs11043003) were excluded from further analysis for both the ADHD and the schizophrenia case-control association study due to low genotyping rates and deviations from HW equilibrium. The genotyping rates for these SNPs were also significantly different from the controls. In addition, 10 individuals from the ADHD sample and 19 individuals from the schizophrenia sample were removed prior to association analysis due to having genotypes for <85% of the SNPs analysed.

Table 3: Quality control measures for the ADHD case-control association study using 12 TH SNPs

SNP rs number	Call rate (cases) %	Call rate (controls) %	p value for difference in call rates	HW p value in cases	HW p value in controls
rs10770140	99.2	99.8	0.45	0.91	0.85
rs10840489	99.1	100	0.08	1	0.9
rs10840490	100	100	1	0.81	0.38
rs10840491	99	99.7	0.09	0.46	1
rs11042978	99.4	99.9	0.13	1	0.39
rs11042982	91	81	0.001	0.003*	0.09
rs11043003	71	62	0.0009	0.002*	0.05
rs11564703	99.1	99.7	0.1	1	0.41
rs11564709	99.4	100	0.1	0.78	0.26
rs2070762	99.4	99.7	0.34	1	0.71
rs4320932	98.5	98.8	0.32	0.057	0.06
rs7924316	99.7	99.8	0.55	0.75	0.5

*: significant at the $p \leq 0.01$ level; **: significant at the $p \leq 0.001$ level

Table 4: Quality control measures for the schizophrenia case-control association study using 12 TH SNPs

SNP rs number	Call rate (cases) %	Call rate (controls) %	p value for difference in call rates	HW p value in cases	HW p value in controls
rs10770140	99.7	99.8	1	0.06	0.85
rs10840489	99.3	100	0.08	0.55	0.9
rs10840490	100	100	1	0.24	0.38
rs10840491	99.4	99.7	0.34	0.16	1
rs11042978	99.7	99.9	0.62	1	0.39
rs11042982	62	81	0.00002	0.001**	0.09
rs11043003	53	62	0.0004	0.00033**	0.05
rs11564703	99.8	99.7	0.6	0.85	0.41
rs11564709	100	100	1	0.1	0.26
rs2070762	100	99.7	0.62	0.72	0.71
rs4320932	98.2	98.8	0.27	1	0.06
rs7924316	99.8	99.8	1	0.65	0.5

*: significant at the $p \leq 0.01$ level; **: significant at the $p \leq 0.001$ level

Appendix II

Table 1: Combination of SNP alleles for each Y chromosome haplogroup. 0 represents the ancestral state of each SNP and 1 the derived state. YAP has not been genotyped and is present for completion, since it is the 'root' of the Y chromosome haplogroup tree. Y^* is the most ancient haplogroup where all men belong

	YAP	M89	M201	M170	M304	M410	M9	M173	M405	M153
Y*	0	0	0	0	0	0	0	0	0	0
DE	1	0	0	0	0	0	0	0	0	0
F*	0	1	0	0	0	0	0	0	0	0
G	0	1	1	0	0	0	0	0	0	0
I*	0	1	0	1	0	0	0	0	0	0
J*	0	1	0	0	1	0	0	0	0	0
J2	0	1	0	0	1	1	0	0	0	0
K*	0	1	0	0	0	0	1	0	0	0
R1*	0	1	0	0	0	0	1	1	0	0
R1b3M405	0	1	0	0	0	0	1	1	1	0
R1b3d	0	1	0	0	0	0	1	1	0	1

Appendix III – Interactions

Table 1: Interactions between 3 TH SNPs and 9 Y chromosome SNPs in 312 male schizophrenia cases and 646 male controls. It was not possible to test for interaction with Y chromosome SNP M153 because it was not polymorphic in the schizophrenia cases

TH SNP	Y chromosome SNP	OR for interaction	χ^2	p value
rs10770140	M153	n/a	n/a	n/a
rs10770140	M170	0.99	0.002	0.96
rs10770140	M173	1.09	0.56	0.46
rs10770140	M201	1.66	0.76	0.38
rs10770140	M304	1.15	0.17	0.68
rs10770140	M405	1.07	0.26	0.61
rs10770140	M410	1.66	1.07	0.3
rs10770140	M89	1.19	0.41	0.52
rs10770140	M9	1.08	0.4	0.53
rs11042978	M153	n/a	n/a	n/a
rs11042978	M170	1.03	0.04	0.83
rs11042978	M173	0.94	0.27	0.61
rs11042978	M201	1.04	0.01	0.92
rs11042978	M304	1.07	0.02	0.89
rs11042978	M405	1.10	0.53	0.47
rs11042978	M410	0.91	0.02	0.9
rs11042978	M89	0.79	0.5	0.48
rs11042978	M9	0.98	0.02	0.9
rs11564709	M153	n/a	n/a	n/a
rs11564709	M170	1.27	1.39	0.24
rs11564709	M173	1.22	1.17	0.28
rs11564709	M201	1.09	0.02	0.89
rs11564709	M304	0.99	0.001	0.98
rs11564709	M405	1.18	0.59	0.44
rs11564709	M410	1.58	0.8	0.37
rs11564709	M89	1.08	0.02	0.89
rs11564709	M9	1.24	1.36	0.24

TH: Tyrosine Hydroxylase; OR: Odds Ratio; n/a: non-applicable