GENETIC DIVERSITY OF 'BRAIN GENES' ACROSS WORLDWIDE POPULATIONS

DOCTORAL THESIS

Michelle Gardner

Barcelona, May 2007

Departament de Ciències Experimentals i de la Salut

International PhD program in the Health and Life Sciences

Universitat Pompeu Fabra

GENETIC DIVERSITY OF 'BRAIN GENES' ACROSS WORLDWIDE POPULATIONS.

Dissertation presented by Michelle Gardner to obtain the PhD degree from the Pompeu Fabra University. This work was carried out under the supervision of David Comas in the Unit of Evolutionary Biology of the Department of Experimental Health and Life Sciences, Pompeu Fabra University.

David Comas Martínez **Michelle Gardner**

Barcelona, May 2007

INDEX

ACKNOWLEDGEMENTS

INTRODUCTION

1. Human Genetic Variation

One of the biggest achievements in modern genetics since DNA was first discovered over fifty years ago has been the sequencing of the human genome, the so-called 'Human Genome Project' (HGP) (Venter et al. 2001). The completion of this immense undertaking carried out by an international group of top scientists was eagerly awaited by the scientific and medical communities and indeed the world as it had been heralded as having the capacity to unlock the secrets of our genomes, and importantly not least for the funding bodies was the hope that it would be the key to rapid advances in medical genetics, the elucidation of the genes underlying many of the most common ailments of our species.

Along with the more practical, clinical aspirations of the project, what really kept many scientists waiting with baited breath for the completion date was the belief, whether directly or indirectly promised by the HGP scientists or not, that this would prove a turning point in our understanding of our species, the diversity among human populations, along with our divergence from our closest relatives, species from the primate order. Basically we awaited the answers to the primordial questions of 'Who are we and what makes us human?'

And so, what of the outcome? After the initial euphoria at the success of the project had worn off we were left with some rather disconcerting questions, yes we had the human genome sequence but what had this achievement done for advancement of our knowledge of human disease. Most of the genes for Mendelian, simple, monogenetic disorders such as Cystic Fibrosis, Huntington's Chorea among

others had already been pinpointed in advance of the publication of any of the project data. However these single gene highly heritable diseases are by and large rare so what was really of interest was the untangling of the genetic factors involved in complex diseases, by far accounting for the vast majority of the health burden in the Western world (Botstein and Risch 2003). In this case the results so far have been quite disappointing, despite a flurry of publication in the post genomic sequence era, corresponding tangible results in the form of the identification of genes unequivocally causing disease susceptibility so far evading us (discussed further in Section 2.2, 'Complex Diseases').

In addition, somewhat humbling findings of the HGP for us humans and with the potential to deflate our collective ego, include the conclusion on the number of genes contained in the Human Genome (rather less than we expected), the amount of difference between ourselves and the nearest non-human primate species (very little, we share about 99% of our DNA with Chimpanzees) and our relative homogeneity as a species (Bertranpetit 2004). What then explains the vast differences between us and our nearest relative, the Chimpanzee for example, and what explains the phenotypic differences we can see between human populations scattered around the globe?

So it is clear where the challenges in the genetics field lie, yet what has the HGP done to advance our knowledge in these arenas? One of the major contributions of the project has been in the identification of a wealth of human genetic variation, providing an immense aid to advancement of studies of the origin and evolution of humans and medical genetics alike. This in turn has spurred on the undertaking of large genotyping projects of this genetic variation, such as Perlegen (Hinds et al. 2005) and the International HapMap project (Altshuler 2005).

These ‘follow-on projects’ and other advances encouraged by the HGP in turn will take us to the next step of the ‘post genome era’, that of a realisation of the greater complexity of ‘big questions’ that we are asking than previously thought, hopefully a more humble approach to our work towards obtaining those answers, and yet at the same time facing the future with hope, knowing that we are heading in the right direction and that these primordial questions will some day be answered.

1.1. DNA Sequence variation

'Polymorphism' (from Greek: *poly* "many", *morph* "form") in the context of Human Genetics refers to variation in the DNA sequence among individuals that may or may not affect its phenotype. Polymorphisms which have a population frequency of 1% or more are considered to be common polymorphisms. Human Genetic polymorphism ranges from single base changes in the DNA (SNPs), small insertions and deletions of a number of bases, through expansions or contractions in the number of tandemly repeated DNA motifs, insertions of transposable elements, insertions, deletions (together referred to as 'indels'), duplications and insertions of megabase segments of DNA, to translocation of chromosomal segments and even changes in chromsomal number.

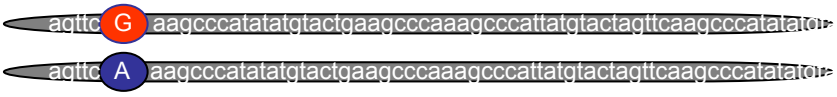
As mentioned previously the advent of the fully sequenced human genome has aided in the identification of an abundance of new DNA sequence variation in the human genome. These differences account for heritable variation among individuals including susceptibility to disease. By far the most common form that genetic variation takes is that of a substitution of at a single base pair of DNA, dubbed Single Nucleotide polymorphisms (SNPs).

SNPs

The human nuclear genome contains 3000 million base pairs (bp) of DNA. SNPs are single base pair changes from one of the four bases making up human DNA to another, for example a G to an A (Figure 1). SNPs arise due to mutation, normally due to a misincorporation of a nucleotide during replication, or by chemical or physical mutagenesis. They are biallelic that is two alleles or forms of the SNP

exist in human populations with a certain percentage in a population carrying one of the alleles and the rest carrying the other. The allele found at highest frequency is referred to as the 'Major Allele' and the lower frequency allele is referred to as the 'Minor Allele'.

Figure 1: Example of a SNP. The two chromosomses in an individual with a G/A SNP.



It is estimated that there are around 11 million common, >1% Minor Alele Frequency (MAF) frequency SNPs in the human genome and ~7 million with MAF >5% (Kruglyak and Nickerson 2001). The frequency of SNPs may vary according to population due to demographic influences undergone by the population such as migration, genetic drift and also population specific selection, discussed in more detail in Section 1.2, Forces shaping Genetic Variation.

SNPs may be found in non-genic regions of the genome in addition to being found in genes. SNPs that are located in coding regions of a gene are called coding SNPs (cSNPs). It is estimated that the average human gene contains 126 SNPs, 5 of which are found in coding regions (Crawford et al. 2005). Non-synonymous SNPs (nsSNPs) have an effect on protein structure and/or function by causing an amino acid substitution. Non-synonymous coding SNPs comprise a group of SNPs that, together with SNPs in regulatory regions, are believed to have the highest impact on phenotype.

SNP markers have become one of the tools of choice for many different types of genetic studies. Their popularity has much to do with their abundance and density

in the human genome. As well as testing SNPs which are potentially the disease causing mutation for direct implication in a disease, SNPs are also one of the most commonly used markers for disease gene mapping studies, see section 2.4 for more indepth discussion, replacing other classes of markers used up until now (various types of tandem repeat markers). They are also used in evolutionary studies, for example in detection of natural selection (See 'Natural Selection' section).

Haplotypes

Haplotypes are particular combinations of alleles observed together on a chromosome that are inherited together as a unit (Figure 2).

Figure 2: A 3 SNP Haplotype in a diploid individual.

When a new variant arises, it does so on a particular chromosomal haplotype. Haplotypes are only disrupted by mutation and recombination in subsequent generations. Haplotypes therefore can also be used as markers for tracking a variant allele in a population. Using haplotypes in disease gene mapping has the advantage that it allows most of the genetic variation across a sizeable region to be captured, using just a few markers (SNPs) to identify the haplotype.

It has been demonstrated that the Human Genome has a particular haplotype structure: that of blocks of varying length with a reduced number of common haplotypes separated by areas of recombination (Gabriel et al. 2002; Kauppi et al. 2004; Sachidanandam et al. 2001). Within these blocks three to five common

haplotypes capture ≈ 90% of all chromosomes in a population (Gabriel et al. 2002). Non-African populations exhibit a reduced haplotype diversity compared to African populations due to the known bottleneck that occurred in these populations according to the 'Out of Africa' theory of human origin (Cann et al. 1984; Stringer and Andrews 1988). This theory holds that African populations have the greatest diversity because they have had the largest population size over a greater amount of time, and that the populations that migrated out of Africa contain only a subset of that diversity. The more generations that have passed since the common ancestor the greater the opportunity for recombination to have broken up a haplotype. This means that African populations have both shorter 'Haplotype blocks' and an increased diversity of haplotypes within these blocks compared to other populations such as Asian and European. This has been found to be the case in numerous studies in recent years (Crawford et al. 2004; Gonzalez-Neira et al. 2004; Stead et al. 2003).

Apart from gene mapping the distribution of haplotypes and haplotype lengths is of importance in the study of evolution and the search for traces of natural selection.

Linkage Disequilibrium

Linkage Disequilibrium or LD refers to correlations among neighbouring alleles reflecting haplotypes descended from single ancestral chromosomes (Reich et al. 2001). LD has been found to be variable both within and among loci and populations (Gabriel et al. 2002; Goldstein and Weale 2001; Pritchard and Przeworski 2001).

Studies of haplotype diversity first indicated that the human genome was composed of stretches of high LD (corresponding with blocks of reduced Haplotype diversity mentioned before), punctuated by recombination hotspots or points of extremely low LD (Gabriel et al. 2002; Goldstein 2001; Reich et al. 2001). As with Haplotype diversity, the extent and amount of Linkage disequilibrium is found to be lower in Africans, with shorter stretches of LD in African populations, conversely Non-African populations have greater amounts of LD extending greater distances reflecting known human evolutionary history (Bertranpetit et al. 2003; Gabriel et al. 2002; Gonzalez-Neira et al. 2004; Reich et al. 2001).

Knowledge of the LD pattern of the human genome forms the basis of many gene mapping strategies, as most association studies rely on LD between the markers analysed for association with disease (most commonly SNPs) and the actual disease locus, as the disease locus is unknown and there is often no *a priori* candidate disease causing variant (Goldstein and Weale 2001). Moreover, as many SNPs show correlated genotypes due to LD (Risch and Merikangas 1996; Sachidanandam et al. 2001), this results in redundancy and means that fewer SNPs need to be typed resulting in a great reduction of cost and time (Goldstein and Cavalleri 2005; Howie et al. 2006).

The International HapMap project

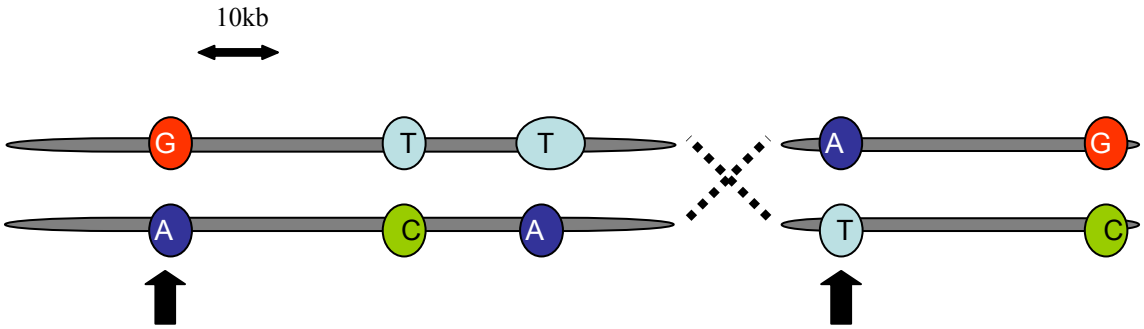
The International HapMap project (HapMap) (Altshuler 2005) is a large scale multi-centre SNP genotyping project launched in the wake of the Human Genome Project and with the aid of the new wave of SNP polymorphism data resulting from the HGP project and compiled by the International SNP Map working group (Sachidanandam et al. 2001). The release of the whole human genome sequence,

which was in effect the sequence of just one mosaic individual, led to the realisation that what was now necessary was a comprehensive study of human genetic variation in as large a number of individuals as possible and in populations of distinct ancestry. The HapMap project aimed to provide key information on genome-wide diversity and the extent and pattern of LD in four worldwide populations: Ceph European (of Central North European Ancestry), Yoruba (Nigeria), Han Chinese and Japanese. It was envisaged that such a resource, providing a reference map of haplotype regions across the genome and with information on the LD structure in these four populations, would greatly facilitate gene mapping studies (Adam 2001; Altshuler 2005) as LD forms the basis for many gene mapping approaches (Ardlie et al. 2002; Wall and Pritchard 2003).

tagSNPs

The redundancies among SNPs due to Linkage Disequilibrium is of central importance in the design and analysis of genetic association studies. In genetic association studies ideally all putative causative alleles for association with disease should be tested and, as often candidate variants are unknown, so these too must be tested for in some manner, usually via surrogate markers, therefore a comprehensive association study might require an unfeasibly large number of SNPs to be genotyped. Knowledge of the Haplotype structure of the genomic region of interest allows selection of a reduced number of SNPs which ‘tag’ the common haplotypes of a region (Figure 3). This increases the chances that at least one typed SNP will be in LD with the disease causing variant, thus the typing of redundant SNPs is avoided saving time and cost, and with minimum loss of information (Carlson et al. 2004; Halldorsson et al. 2004; Johnson et al. 2001).

Figure 3: Two chromosomes with tagSNPs indicated by arrows



One of the most direct applications of HapMap to genetic association studies therefore is that of facilitating the selection of tagSNPs (Altshuler 2005). As previous surveys have shown that haplotype composition and structure and quantity of LD is heterogenous across global populations (Bertranpetit et al. 2003; Mateu et al. 2001; Pritchard and Przeworski 2001) a valid question for researchers is whether tagSNPs selected in either of the four HapMap populations are applicable to other worldwide populations. A number of studies have aimed to address this issue, and on the whole initial results have found in favour of the HapMap project that tagSNPs selected in a HapMap population are applicable to similar populations or other populations from the same continent (Gu et al. 2007; Ramirez-Soriano et al. 2005; Ribas et al. 2006) and even to populations from other continental regions (Gonzalez-Neira et al. 2006).

Once a clear answer from initial studies of applicability of tagSNPs to 'real life' association studies emerges, an escalation in discovery of the genes underlying common complex disease is envisaged (Hirschhorn et al. 2002), with already numerous tagSNP selecting algorithms available (Howie et al. 2006) (Carlson et al. 2004) incorporated into user friendly programs and even a number of genome wide tagSNP sets having become recently comercially available for the carrying out of whole genome scans (Barrett and Cardon 2006).

1.2. Forces shaping Genetic Variation

Genetic variation initially arises by genomic based mechanisms such as mutation, recombination and chromosomal rearrangements. The array of extant human genetic variation that can be observed today has been shaped by forces from our evolutionary past including demographic factors and selection.

Demographic forces

Human genetic variation in the form of Allele Frequencies, Haplotype Diversity and Linkage Disequilibrium has been shaped by a number of population based forces over the course of a population's history. These forces include genetic drift, which is the tendency of allele frequencies to fluctuate randomly over time due to statistical variation (Wright 1931, 1938). Eventually, after a certain number of generations the drift will carry an allele to fixation or eliminate it. Genetic drift is dependent on the size of a population, in particular the Effective population size, N_e, that is the size of an idealized population that experiences the same amount of drift as the population under study, with small populations being more susceptible to the effects of genetic drift.

Genetic drift also plays a role in other demographic forces that shape the genetic variation of a population: 'Population Bottlenecks', where a population undergoes a sudden constriction in size, such as in the 'Out of Africa' theory of human evolution (Cann et al. 1984; Stringer and Andrews 1988) and the 'Founder effect' (Mayr 1963) where a subset of a population carrying a fraction of the original diversity establishes a new separate population such as in the peopling of the Americas (Hey 2005). The reduction in genetic diversity in each of these cases

corresponding to a lower N_e leaves the population more vulnerable to the effects of genetic drift.

Another demographic factor affecting genetic diversity and genetic differences between populations is gene flow. Gene flow between human populations occurs due to migration of individuals into (immigration) or out of (emigration) a population. Immigration logically increases genetic diversity and emigration results in a loss of diversity.

As global populations have their individual demographic histories, so differences in populations may be observed in the current patterns of genetic variation between human populations. It is therefore possible to recognise and distinguish the different population processes by the distinctive patterns they have left in extant genetic variation in a population.

Natural Selection

One of the basic premises upon which modern day population genetic thought is based is the so called 'Neutral theory of molecular evolution' (Kimura 1968). This theory stipulates that genetic drift is one of the main driving forces in evolution and that most genetic variation is neutral with regard to selection. Studies of selection therefore must distinguish those traces left on the pattern of genetic diversity by selective forces from those due to demographic factors mentioned above including genetic drift (Akey et al. 2004). This 'neutral theory of evolution' can also be used as the null hypothesis against which empirical data can be compared to determine the case for or against selection (Akey et al. 2004).

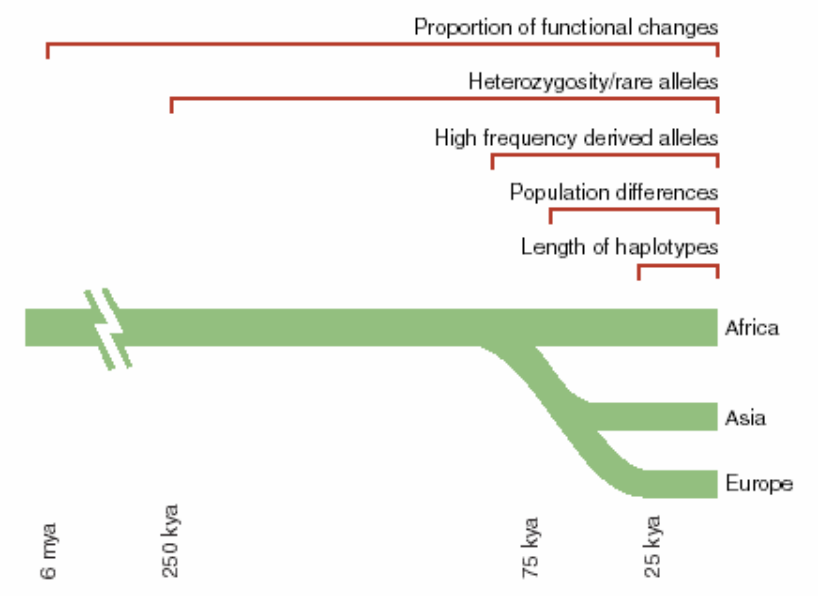
The advent of the human genome sequence and the flood of new polymorphism data that has come with it has provided valuable new tools in the search for traces of Natural Selection in the human genome. Natural Selection is one of the forces that has most shaped our species (Sabeti et al. 2006). Positive selection is the tendency of beneficial traits to become more frequent in a population over time (Darwin 1858) and is likely to have aided in our ability to adapt to new and diverse environments. Negative selection (also known as purifying selection or stabilising selection) is conversely the selective removal of rare deleterious alleles from a population and Balancing selection refers to when the advantage is in having a number of polymorphisms at relatively high frequency in a population such as in the case of Heterozygote advantage (Heterozygous state confers resistance to disease), for example heterozygotes of the sicle cell haemoglobin gene and resistance to malaria (Aidoo et al. 2002). Each of these types of selective forces will leave their characteristic signals on the shape of genetic diversity around the gene or allele under selection.

Some of the most easily distinguished traces left by the forces of selection are those left by selective sweeps (Nurminsky 2001). Selective sweeps occur when an allele becomes more frequent in a population as a result of positive selection. As the positively selected allele increases in frequency so too will linked nearby alleles, a phenomenon known as genetic hitchhiking. A strong selective sweep will result in a region of the genome where the positively selected haplotype (of the selected variant and linked neighbouring alleles) is at high frequency, thus leading to a reduced haplotype diversity in the region. Thus the occurrence of a selective sweep can be investigated by measuring LD or by observing if a haplotype is overrepresented in a

population. As a selective sweep carries an allele on a specific haplotype to high frequency faster than the rate at which it is broken down by recombination then high frequency haplotypes will be observed longer than expected under neutrality (Sabeti et al. 2002). This phenomenon has been exploited in the 'Extended Haplotype Homozygosity' algorithm for detecting recent positive selection by Sabeti et al (Sabeti et al. 2002) and may be useful in detecting more recent positive selection (Figure 4).

Local selection refers to natural selection acting at the local level, that is, population specific selection. Population specific selection allows for adaptation to new environments, exploiting geographic differences in available foodstuffs for example in the case of the Lactase gene (Bersaglieri et al. 2004), for increasing fitness in the form of resistance to endemic infectious disease for example the CCR5-Delta32 gene (Sabeti et al. 2005) and the Duffy blood group locus and resistance to malaria (Hamblin et al. 2002). Population specific selection has therefore occurred since the migration of human populations out of Africa and into new environments (Figure 4). The altered allele frequency patterns arising from such local adaptation can be detected as a greater differentiation in allele frequencies between populations than that expected under neutrality and this is commonly measured with the F_{ST} statistic. F_{ST} is simply a measure of population differentiation based on allele frequencies (Wright 1969). Balancing selection or species-wide positive selection would give rise to more uniform allele frequencies compared to that at neutral loci (Hamblin et al. 2002).

Figure 4: The timescales for detecting signatures of selection**.**



From (Sabeti et al. 2006)

In addition a selective sweep acting on a gene should also cause a detectable disproportionate number of SNPs with high frequency of rare alleles (Minor Allele Frequency), while recent balancing selection may be detected by observing an excess of SNPs with high MAF (Walsh et al. 2005). An excess of derived alleles may be indicative of a genetic hitchhiking event (Fay and Wu 2000) (Figure 4).

Background selection is the converse of the case of a selective sweep. A reduction in linked variation occurs as a deleterious site is selected against leading to a region of the genome with reduced genetic diversity.

Knowledge of the past selective events undergone may provide new insights into the aetiology of human disease past and present and also shed light on the events that have shaped our species so it is exciting that such knowledge is within our grasp due to the technological and methodological advancements in recent years.

2. *The Genetics of Complex Disease*

2.1. Monogenic Diseases

Monogenic diseases are those showing the most straightforward mode of genetic transmission, and therefore there has been most success in the identification of genes underlying these type of diseases. A Monogenic trait is one which is controlled by a single locus and is passed on from parents to offspring in Mendelian fashion. They are normally fully penetrant, that is presence of the allele corresponds with presence of the trait, and no environmental factors are required for expression of the trait. Sickle-cell anaemia, Tay-Sachs disease, cystic fibrosis and xeroderma pigmentosa are all examples of Monogenic diseases. Initial successes in discovering the genes underlying diseases with a heritable component were almost entirely centred on simple monogenic disorders following a Medelian pattern of inheritance (Ardlie et al. 2002). However these disorders are relatively rare, the vast majority of the common illnesses instead have a complex mode of inheritance.

2.2. Complex Diseases

“Human genetics is now at a critical juncture. The molecular methods used successfully to identify the genes underlying rare mendelian syndromes are failing to find the numerous genes causing more common, familial, non-mendelian diseases.” (Risch 2000).

Application of the gene mapping strategies which drew such success in the field of Mendelian disease to the field of complex disease is plagued by a marked

lack of success (Cardon and Bell 2001) (Terwilliger and Weiss 1998) with just a few confirmed cases of a gene associated with disease (Hirschhorn et al. 2002). However it is precisely with complex diseases that the greatest challenge lies as complex diseases such as cancer, cardiovascular disease and psychiatric disorders reperesent some of the most common illnesses in humans of our times.

Complex diseases are caused by one or more genes in conjunction with environmental factors, with some estimates putting the genetic component of susceptibility at around 40%-70% (Goldstein and Cavalleri 2005). Complications of the mode of inheritance such as gene-gene interactions (epistasis), gene-environment interactions, and non-additive gene effects abound. In complex disease there may be multiple genes of small effect. Complicating factors may also include genetic or allelic heterogeneity, where different genes or different alleles within the same gene give rise to increased susceptibility to the disease phenotype, this may be the case in particular in different populations. There may also be distinct environmental factors interacting with the genetic component in different populations. Penetrance, that is, lack of a one to one transmission between the genotype and the phenotype (Merikangas and Risch 2003) may also be a factor, as may phenocopy (phenotype observed but not due to genotype).

The widely held ‘Common variant common disease’ hypothesis (Lohmueller et al. 2003) holds that common variants have an important role in common diseases (Altshuler 2005). These common variants may have small effect sizes, incomplete penetrance, additive effects or unclear interactions with environmental factors thus hindering attempts at their elucidation (Pritchard and Cox 2002).

Most psychiatric disorders show a complex mode of transmission, therefore the problems plaguing complex disease genetics are also common to the psychiatric genetics field with a general air of dissappointment that despite the recent technological advances there has been a distinct lack of success in pinpointing genes underlying the common psychiatric and behavioural disorders (Owen et al. 2000) (Merikangas and Risch 2003).

However all hope may not be lost, as after initial disappointment the attitude seems to be 'regroup and come back stronger' and so new strategies for approaching complex disease gene mapping are emerging and this together with the falling cost of genotyping and the new influx of polymorphism data from the HapMap Project have brought genome-wide association studies (GWAS) of complex diseases to within our reach (Barrett and Cardon 2006), and it is exactly this approach which seems to hold the most promise for shedding light on the genes for complex disease (See more detail in 'Association Studies').

2.3. Psychiatric Disorders

Most Psychiatric disorders are complex traits with both genetic and environmental factors influencing an individual's susceptibility to suffering disease. The heritability estimate for some of the commonly studied psychiatric disorders ranges from 0.28 for depression to 0.90 for autism (Merikangas and Risch 2003), (Table 1).

The recent advances in molecular genetics including the Human Genome Project promised to greatly increase our understanding of psychiatric disorders and human behaviour, however these expectations have yet to be fulfilled (Owen et al.

2000). This failure is due to many reasons, the difficulty of identifying genes for complex disease in general (discussed in depth in Sections 2.2 and 2.5) and due to characteristics specific to psychiatric disorders such as the difficulty of phenotype defintion.

Table 1: Heritability estimates and risk ratios (proportion of affected first degree relatives of affected probands versus the proportion of affected relatives of nonaffected control subjects) for a range of commonly studied psychiatric disorders.

Disorder	Risk Ratio	Heritability Estimate
Mood disorders		
Bipolar disorder	7–10	0.60–0.70
Major depression	2–3	0.28–0.40
Anxiety		
All	4–6	0.30–0.40
Panic disorder	3–8	0.50–0.60
Autism	50–100	0.90
Schizophrenia	8–10	0.80–0.84
Substance dependence	4–8	0.30–0.50

From (Merikangas and Risch 2003)

The promise offered by shedding light on the underlying genetic component of susceptibility to psychiatric disease is enormous. Discovery of genes would shed light on the molecular pathways underlying these diseases, many of which are poorly understood, thereby giving hope for new treatment options. It could also help by providing new information on the environmental and other risk factors which interact with the susceptibility genotype predisposing an individual to disease, thereby allowing informed lifestyle choices to be made so as to avoid those risk factors. A greater understanding of the biological underpinnings of psychiatric disease could also help with the problem of psychiatric disorder classification, which is hindered by problems such as disease heterogeneity (diseases with distinct biological basis grouped together due to gross clinical observations of disease), variable expressivity,

difficulties with diagnostic spectra such as delimiting where one disorder defintion ends and another begins (Merikangas and Risch 2003).

Recently interest has grown in defining so called ‘endo-phenotypes’ defined based on phenotypic traits or markers with a more sound biological foundation, which may provide a more realistic phenotype for testing for genetic association (Gould and Gottesman 2006; Merikangas and Risch 2003). This new approach together with the flood of new polymorphism data in the form of SNPs available for genotyping, new data on the Linkage Disequilibium in the human genome in different populations allowing selection of SNPs of optimal efficiency (tagSNPs) and new strategies with more power to detect the genetic component of complex disease including psychiatric disease such as ‘whole genome scans’ mean that the identification of the genes predisposing to psychiatric disease may be imminent.

The Schizophrenia example

“If the human race survives, future men will, I suspect, look back on our enlightened epoch as a veritable Age of Darkness… They will see that what was considered 'schizophrenic' was one of the forms in which, often through quite ordinary people, the light began to break into our all-too-closed minds.”

- R.D.Laing

Schizophrenia from the Greek word σχιζοφρένεια, or schizophreneia, meaning "split mind" is one of the most devastating syndromes to beset humans causing a ‘state change’ in the person affected and while medication exists to treat the symptoms, the illness cannot be cured nor the affected individual returned to

premorbid state. It has been suggested that schizophrenia represents not just one disease but in fact a spectrum of disorders (Jablensky 1997).

The new approaches emerging for the search for risk genes for complex disease including psychiatric disease are particularly relevant to schizophrenia. Schizophrenia has a similar incidence across cultures, a prevalence of 1%, with the exception of some 'outlier' populations (Jablensky 1997). Although diagnoses of other psychiatric disorders vary across cultures, schizophrenia has a consistent prevalance between and among cultures and ethnicities (Siever and Davis 2004). There are some voices of dissent however, indicating that when the various syndromes of schizophrenia spectrum disorders are seperated out difference in prevelance is found in different populations (Goldner et al. 2002).

The single most important known risk factor for schizophrenia is genetic risk with numerous family, twin and adoption studies confirming the role of inheritance, and with various sources putting the heritability of the disease at around 80% (Merikangas and Risch 2003; Rapoport et al. 2005). Schizophrenia shares the characteristics of other complex disorders of incomplete penetrance, non-Mendelian inheritance, heterogeneity (genetic and allelic) and probably phenocopies (Cardno et al. 1999; Kendler and Gardner 1997). In addition it is thought that there is a substantial genetic influence on the age of onset of schizophrenia (Kendler et al. 1987).

Obviously there is great interest in elucidating the gene behind this disorder (or disorders), with a plethora of studies to date aiming to do just this. There have been numerous association studies focussing on candidate genes from functional pathways implicated in the disorder such as those involved in brain development,

synaptic connectivity and neurotransmission, for example, the so-called 'Dopamine Hypothesis' (Winterer and Weinberger 2004), the 'Serotonin Hypothesis' of schizophrenia (Aghajanian and Marek 2000) or the glutamate theory of schizophrenia (Collier and Li 2003), with varying success (Berry et al. 2003; Harrison and Owen 2003; Owen et al. 2000; Riley and Kendler 2006). One of the most exciting candidate genes to emerge from genetic studies to date is the Neuregulin 1 gene (*NRG1*) (Stefansson et al. 2002), a candidate gene that agrees with the neurodevelopmental model of Schizophrenia (Rapoport et al. 2005) , and the hypothesized role of glutamate, as discussed further in Chapter 3 'Brain Genes'.

Schizophrenia has been referred to as an 'evolutionary enigma' (Brune 2004), that is, despite a widely accepted reduction in fecundity (male fecundity reduced by about 50% compared to healthy males) it persists in virtually all human populations at an incidence rate of roughly 1% exceeding a mere chance effect due to high mutation rates (Brune 2004). This is now known as the 'schizophrenia paradox'. Schizophrenia can either be viewed as an evolutionary advantageous condition (for example advantage to kin) or as a disadvantageous byproduct of normal brain evolution (the development of language (Crow 2000)) (Polimeni and Reiss 2003).

2.4. Complex disease Gene Mapping

Linkage Analysis

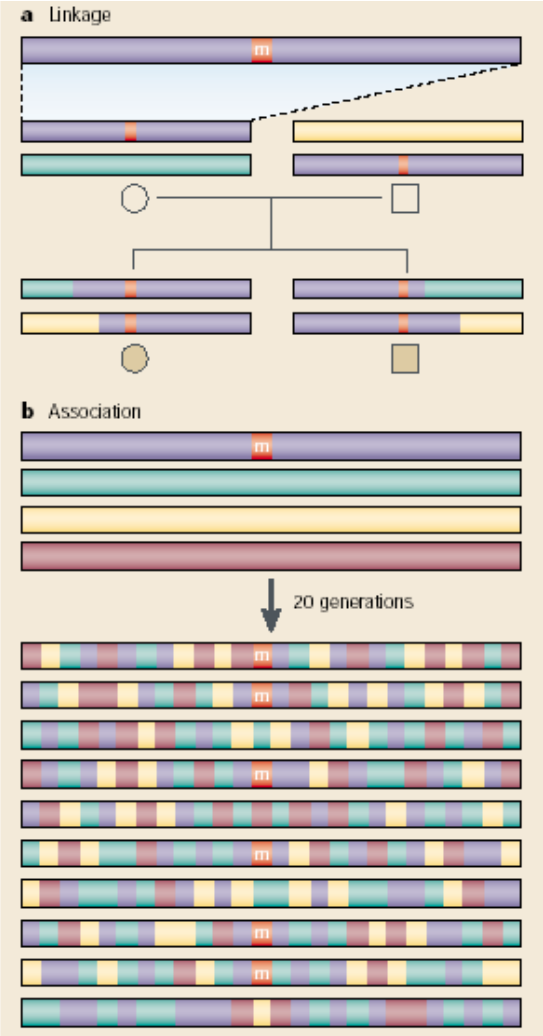
Success in Mendelian disorder gene mapping to date owes much to the Linkage Analysis method of Gene Mapping (Ardlie et al. 2002). In Linkage Analysis regions of the genome which cosegregate with the disease in large multi generational affected families or in a number of independent affected families are zoned in on as

potential disease gene harbouring regions (Figure 5). The disease locus, that is if there is just one disease locus, will lie in the region of the genome that is shared by all affected members of a pedigree. One limitation is that the area the potential disease locus is localised to may be still quite large, prohibiting the exact pinpointing of the gene, never mind the disease causing variant. Success using this approach in the context of complex disease has been limited however due to the low power in detecting genes of small effect (often the case in complex diseases) (Pritchard and Cox 2002) and the number of families required to correctly identify linkage would be prohibitively large (Owen et al. 2000; Risch and Merikangas 1996)

Association Studies

Genetic association studies in contrast with Linkage analysis assess correlations between genetic variants and trait differences on a population scale, see Figure 5. Association studies have been proposed as a solution to the low power which poses a problem when attempting to identify the genes involved in complex disease which individually are often of modest affect (Risch and Merikangas 1996).

Figure 5: Linkage Analysis and Association Study approaches to gene mapping.



From (Cardon and Bell 2001)

Until recently association studies most frequently took the form of case-control studies, where a difference in allele frequency between affected individuals and unrelated controls was sought. However this type of study design has the serious drawback of confounding due to population stratification, where allele frequency differences exist between the cases and controls for reasons other than the presence or absence of the disease phenotype under study, e.g. ethnicity (Risch 2000). This problem exists even when cases and controls are taken from the same population as subtle substratification may exist. For this reason many researchers now favour family based association studies, examples being the transmission/disequilibrium test

(TDT) (Spielman et al. 1993) and Sibpair studies. The TDT tests transmission of marker alleles from parents heterozygous for the marker to affected offspring. Family based studies however have the drawbacks of being sensitive to the effects of non-random mating and have less potential for examining interactions with environmental risk factors (Owen et al. 2000).

Association studies may take the candidate gene approach, that is investigating association of the disease with a gene or genes which have been tentatively identified as being involved in disease susceptibility. This requires previous notions of the biological basis of the disease, which often is unclear in the case of complex diseases, along with some knowledge of the function of the gene or genes involved.

The whole genome association approach (WGA), is a new approach which has been made feasible by the recent surge in SNP polymorphism data along with the technological advances of recent times. Whole genome studies of genotype-phenotype associations without the necessity of an hypothesis *a priori* as to the disease mechanism have the potential to reveal the contributors to many complex traits and diseases thus leading to new insights into the aetiology of complex disease. The WGA involves the screening of hundreds of thousands of markers on large samples of individuals with the phenotype in question i.e. the psychiatric disorder or a continuous endophenotype. In direct WGA studies polymorphisms likely to alter structure, function or expression of the gene product are studied whereas indirect approaches to WGA use a dense enough set of markers to allow LD between at least one marker and the susceptibility locus (Lawrence et al. 2005)

2.5. Factors Influencing outcome of Association Studies

While association studies hold great hope for the identification of genetic variants influencing susceptibility to common complex diseases, they are plagued by the impression that they are not consistently reproducible (Cardon and Bell 2001; Lohmueller et al. 2003). The reasons behind the apparent lack of success has been hotly debated and papers putting forward ideas on the motives behind this are almost as numerous as the publications reporting associations or lack of it, replication of an association, or more frequently the failure to replicate a previous finding of association. To confirm the role of a gene or variant in a disease it is necessary to replicate the initial finding of association, with some suggestions that the only true way to confirm a gene or variant's implication in a disease is through Meta-Analyses of all the available data (Ioannidis 2003).

The problems encountered most certainly are due at least in part to the complex patterns of transmission characterising complex diseases (see section 2.2). Limitations of the molecular methods used with success in the field of Mendelian disease as applied to the field of complex disease genetics have also become apparent (Risch 2000).

All of this is compounded in the case of psychiatric disorders by the problems of defining accurately the clinical phenotype to be tested (Section 2.3). Differences in disease incidence among populations could play a role in these failures of replication, in the case of psychiatric disease where incidences are shown to vary in different populations (Siever and Davis 2004). In the case of Schizophrenia, although incidence is quite consistent across populations, this has been suggested to be a

problem of phenotype definition, that is, that schizophrenia represents a heterogeneous group of disorders with distinct biological bases grouped together based on observed clinical characteristics (Jablensky 1997), each of these distinct disorders having a different incidence in different populations (Goldner et al. 2002).

Importantly and often overlooked is the fact that population based differences may also play a role in the varying outcomes of disease association studies (Gardner et al. 2007). True variation in the presence or size of an association between an allele and the disease phenotype in different populations could underlie failure in replication of associations (Colhoun et al. 2003). This variation could arise if different disease causing genes or different disease causing alleles within those genes predominate in different study populations. Although as common variants are more likely to be found globally, a causal association between a candidate SNP and trait should be reproducible in ethnically diverse populations (Risch 2000).

Crucially, population differences in linkage disequilibrium (LD) and allele frequencies (of marker or disease causing variant), the result of the distinct population history of the populations under study including demographic and selective forces (See 1.2) may play a role (Gardner et al. 2007; Zondervan and Cardon 2004).

LD varies within and among loci and populations (Gabriel et al. 2002; Pritchard and Przeworski 2001), a potential cause of discepencies in LD between causal variant and marker SNP in different populations. As most association studies are ‘indirect association studies’ relying on the LD between marker and disease causing variant to detect association, this represents one very strong possible contributing factor to failure of replication in different study populations. Variability

in disease or marker allele frequencies between populations is another possible confounding factor, as mismatches between marker and disease allele frequencies may reduce the statistical power to detect associations between the complex trait and the candidate gene (Zondervan and Cardon 2004). Taking all of the above into account, it is clear that association studies must be interpreted within the context of the genetic structure of the populations being studied (Lohmueller et al. 2006)

Although population based differences can pose obstacles in genetic association studies they can also be used to advantage, with the varying LD between ethnicities used to resolve true association. High LD populations can be used to initially detect SNP associations and low LD populations may resolve which SNP effect is primary (Risch 2000).

3. ‘Brain Genes’

“In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary aquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history” *Origin of the Species* (Darwin 1859).

The size and complexity of our brains sets us apart from all other species on earth. Determining the genetic basis of our brain size and complexity might go some way towards answering that primordial question ‘just what makes us human?’. In the last few years at least some genes have been identified which have been demonstrated to have had an influence on the increase in cerebral size and complexity in the human lineage, examples of note include the Microcephalin gene and *ASPM* in Brain size (Evans et al. 2004; Mekel-Bobrov et al. 2005; Mekel-Bobrov et al. 2007)and *FOXP2* in speech and language (Enard et al. 2002).

Understanding variation globally in genes related to brain size, function and complexity allows an insight into their evolution over time, may allow a deeper understanding of our divergence from primates and permits observation of whether or not selection may have acted on these genes. It may also shed light on the link between human brain gene variation and differences in brain related phenotypes observed in humans.

Brain genes are of particular interest from a medical genetics point of view. The incidence of psychiatric disorders worldwide is high and causes a great deal of suffering for both patients and their families, and although a genetic component is

implicated in a great proportion of psychiatric disorders the aetiology and underlying genes in most cases remain unknown. Identification of the genes involved may shed light on the biological pathways involved in disease offering hope for new treatment options. The identification of genes may even help in psychiatric nosology.

The 'Brain genes' examined in the body of work constituting this PhD thesis code for proteins with essential roles in the human neurological system. Dopamine and Serotonin genes belonging to the Dopamine and Serotonin neurotransmitter pathways respectively and Neuregulin 1 involved in glutamate neurotransmission among other roles (Table 2).

Table 2: Genes studied in this thesis.

Gene		System
COMT	Catechol-O-methyltransferase	Dopamine
DBH	Dopamine beta-hydroxylase	Dopamine
DDC	Dopa decarboxylase	Dopamine
DRD1	Dopamine receptor D1	Dopamine
DRD2	Dopamine receptor D2	Dopamine
DRD3	Dopamine receptor D3	Dopamine
DRD4	Dopamine receptor D4	Dopamine
DRD5	Dopamine receptor D5	Dopamine
HTR1A	5-hydroxytryptamine (serotonin) receptor 1A	Serotonin
HTR1B	5-hydroxytryptamine (serotonin) receptor 1B	Serotonin
HTR2A	5-hydroxytryptamine (serotonin) receptor 2A	Serotonin
HTR2C	5-hydroxytryptamine (serotonin) receptor 2C	Serotonin
HTR4	5-hydroxytryptamine (serotonin) receptor 4	Serotonin
MAOA	Monoamine oxidase A	Dopamine
MAOB	Monoamine oxidase B	Dopamine
NRG1	Neuregulin 1	Glutamate
PPP1R1B	Protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32)	Dopamine
SLC6A3	Solute carrier family 6 (neurotransmitter transporter, dopamine), member 3	Dopamine
SLC6A4	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	Serotonin
TH	Tyrosine Hydroxylase	Dopamine
TPH1	Tryptophan hydroxylase 1	Serotonin
TPH2	Tryptophan hydroxylase 2	Serotonin

Human Neurotransmitter Systems

Neurotransmitters are chemicals that are used to relay, amplify and modulate electrical signals between neurons and other cells. Some of the criteria by which neurotransmitters have been defined classically include that: They are synthesized within presynaptic neurons and are present in sufficient quantity in the pre-synaptic neurons so as to exert an effect on the postsynaptic neuron, and that a biochemical mechanism for inactivation of their action exists.

Neurotransmitters can be divided into amino acids, peptides, and monoamines. Dopamine (DA) & serotonin (5-HT) are two of the three types of monoamine neurotransmitters, and Glutamate (Glu) is an amino acid neurotransmitter. Dopamine and Serotonin are phylogenetically ancient neurotransmitters intrinsic to brain function and behavior (Cravchik and Goldman 2000). Glutamate is considered the principle excitatory amino acid in the Central Nervous System (CNS). Its role in cellular metabolism is well known and it is distributed widely throughout the neuroaxis.

Neuregulin 1

The Neuregulins are a family of four structurally related proteins that are part of the EGF family of proteins. These proteins have been shown to have diverse functions in the development of the nervous system.

The Neuregulin 1 (*NRG1*) gene produces numerous *NRG1* isoforms through alternative splicing (Michailov et al. 2004). *NRG1* is expressed at a high level in brain (Corfas et al. 1995) playing important roles in development and plasticity (Steinthorsdottir et al. 2004). It has been implicated in neurodevelopmental processes

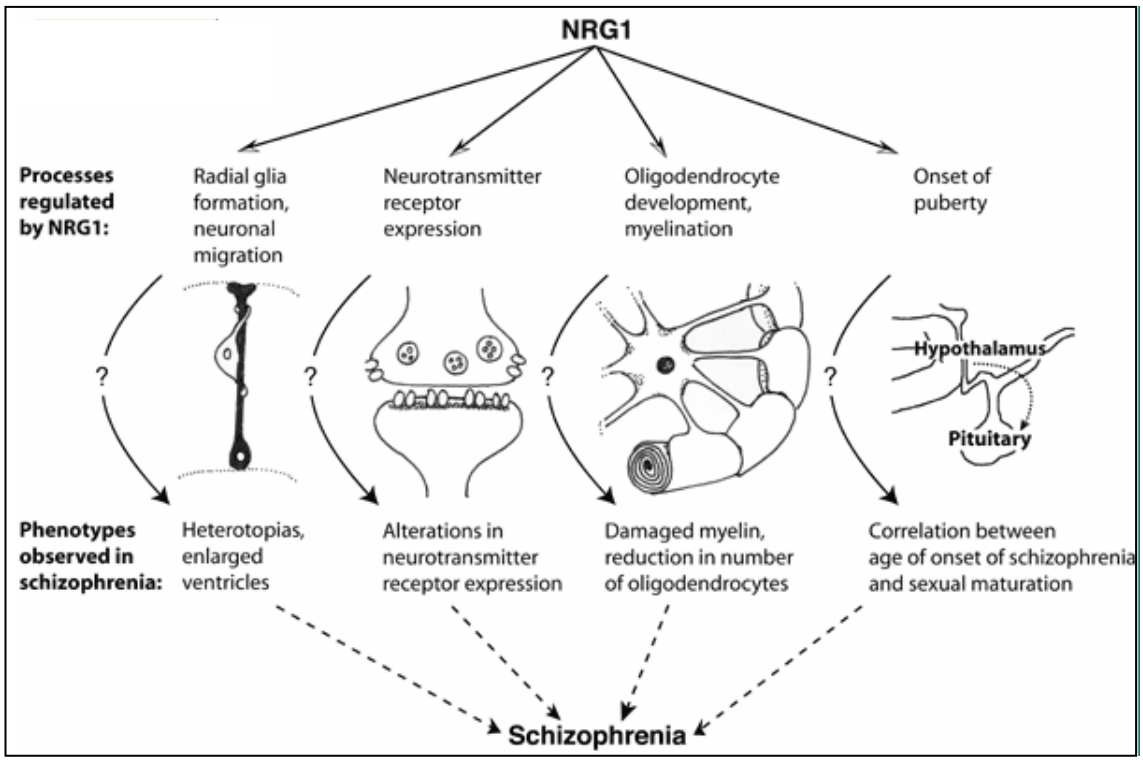
such as neuronal migration and specification and Schwann cell development and proliferation (Corfas et al. 2004b). It is also expressed at central nervous system synapses and has a clear role in the expression and activation of neurotransmitter receptors, including glutamate receptors (also $GABA_A$ and acetylcholine receptors) (Falls 2003).

NRG1 was first associated with Schizophrenia in an Icelandic population by whole genome linkage analysis in large families with multiply affected members followed by an association study to narrow down the region identified (Stefansson et al. 2002), a region which had previously been identified as being linked to schizophrenia in other studies, but this represented the first study to pinpoint the actual gene involved (Blaveri et al. 2001; Kendler et al. 1996; Kendler et al. 2000). Subsequent replications studies in other populations confirmed the association (Corvin et al. 2004; Stefansson et al. 2003; Williams et al. 2003; Yang et al. 2003) although in some ethnically distinct populations there has been a failure to reproduce the original findings (Ingason et al. 2006; Iwata et al. 2004; Tang et al. 2004; Thiselton et al. 2004; Venter et al. 2001) or indeed association with a different region, haplotype or allele within the *NRG1* gene than those of the original study has been described (Collier and Li 2003; Li et al. 2004). This may reflect underlying differences in allele frequencies and LD due to poulation based factors (Gardner et al. 2007).

Functional studies support the hypothesized role of *NRG1* in schizophrenia pathophysiology (Law 2003; Stefansson et al. 2004; Steinthorsdottir et al. 2004). In addition the known roles of *NRG1* agree with some of the most important theories of

Schizophrenia aetilogy, the Neurodevelopmetal theory (Rapoport et al. 2005) and the Glutamate theory of Schizophrenia (Collier and Li 2003).

Figure 6: *NRG1*, its functions and possible correspondances with Schizophrenia phenotypes



From (Corfas et al. 2004a)

The Dopamine System Genes

Dopamine is formed in brain (in the substantia nigra, in the midbrain), adrenal chromaffin cells in the adrenal gland, and sympathetic nerves from phenylalanine and tyrosine precursors. In the Central Nervous system there are specific dopamine containing neurons (dopaminergic neurons) (Seeman et al. 1978). The dopaminergic sysem includes nigrostriatal, meocortilimbic, and

tuberoinfundibular pathways. Dopamine modulates both the brain's reward mechanism and the motor system (Lewis et al. 2003; Sachidanandam et al. 2001). It also has important roles in cognition and endocrinological function (Sachidanandam et al. 2001).

Genetic association studies continue to explore the contribution of gene variants in the dopaminergic system to various psychiatric, neurological and behavioural disorders. Notions of a role for dopamine in these disorders first came from clinical observations made during drug treatment which allowed an insight into the relevance of the dopaminergic system in brain function and disease. Antipsychotic medication was observed to act on dopamine receptors (Seeman et al. 1975), dopamine replacement therapy was observed to alleviate the symptoms of Parkinson's disease while its depletion to lead to depression (Cravchik and Goldman 2000) and studies of drug addiction also implicated dopamine (Lewis et al. 2003).

The understanding of the implications of the dopaminergic system in both health and pathological states has also been greatly increased in recent years thanks to the development of new technologies for genetic analysis and the insight into the genetic sequences of dopaminergic system permitted by the availability of the human genome sequence. The identification of genetic variants that may represent risk or protection factors for a variety of psychiatric disorders has therefore also been allowed (Hoenicka et al. 2007).

Dopamine systems genes have been implicated in a wide range of psychiatric conditions, neurological disease and behavioural disorders from genetic studies. The disorders range from Anorexia (Bergen et al. 2005), Attention Deficit Hyperactivity Disorder (ADHD) (Payton et al. 2001), Bipolar disorder (Abdolmaleky et al. 2005;

Gutierrez et al. 1997), Huntingtons disease (Backman and Farde 2001) to Parkinson's disease (Kang et al. 2006) and Schizophrenia (Winterer and Weinberger 2004) among others. It would impossible to compile a definitive or inclusive list of each dopamine gene and the diseases it has been associated with, as there are literally thousands of publications reporting or refuting associations and numerous diseases associated with each gene and conversely numerous genes with each disease and considerable overlap.

The Dopamine system genes involved in the present study include the genes encoding for the five dopamine receptors, *DRD1, DRD2, DRD3, DRD4* and *DRD5* which have been demonstrated to mediate the diverse physiological functions of dopamine. These genes have various types of polymorphisms that can produce changes in the genetic product or expression levels. The *MAOA* and *MAOB* genes, encoding two distinct forms of the enzyme Monoamine oxidase (MAO), which are regulators of metabolism of neurotransmitters such as dopamine (also Serotonin) are studied as is *COMT* which inactivates neurotransmitters including dopamine, also genes coding for proteins involved in Dopamine biosynthesis such as *DBH*, *DDC* and *TH*, a dopamine neurotransporter *SLC6A3* and *PPP1R1B* a pivotal integrator of information in dopaminoceptive neurons, regulating the response to neuroleptics, psychotomimetics, and drugs of abuse, and affecting striatal function and plasticity.

The Serotonin System Genes

Serotonin is synthesized in serotonergic neurons in the central nervous system (CNS). It is released specifically by cells in the brainstem, in an area called the raphe nuclei, but travels around the brain along the medial forebrain bundle activating the cortex, hippocampus, thalamus, hypothalamus and cerebellum. Also, it is released in

the Caudal serotonin nuclei, so as to have effect on the spinal cord. In the peripherial nervous system (such as in the gut wall) serotonin regulates vascular tone.

Serotonin is involved in temperature regulation, sensory perception and mood control, however it plays a major role in emotional disorders such as depression, suicide, impulsive behaviour and aggression. It is involved in the pharmacology of depression and psychosis (Cravchik and Goldman 2000). Genes from the Serotonin system have also been associated with suicidal behaviour, Bipolar disorder and various behavioural disorders (Hirschhorn et al. 2002). One of the chief theories of the aetiology of schizophrenia is the so-called 'serotonin hypothesis' (Aghajanian and Marek 2000).

Serotonergic and dopaminergic systems are largely interconnected in the brain. The serotonergic projections inhibit dopamine function in the midbrain via inhibition of the firing of the substantia nigra dopamine cells, while in the striatum and in the cortex they inhibit synaptic release and probably synthesis of dopamine (Kapur and Remington 1996).

The serotonin pathway genes studied in this body of work include five of the genes encoding for serotonin transporters and most commonly associated with psychiatric diseases : *HTR1A, HTR1B, HTR2A, HTR2C, HTR4*, the serotonin neurotransmitter transporter *SLC6A4* a neurotransmitter of serotonin in the central and peripheral nervous systems, and a target of an important class of antidepressant drugs, the serotonin selective reuptake inhibitors, and the genes *TPH1* and *TPH2* encoding rate-limiting enzymes in the serotonin biosynthesis pathway.

OBJECTIVES

The work presented in this thesis is a study of the genetic variation in a set of genes related to neurological function ('Brain genes'). Twenty two genes are examined, all of which are involved in either the Dopaminergic, Serotonergic or the Glutamatergic systems of neurotransmission.

The Objective of the study has two aspects: on the one hand the analysis of genetic variation in a set of genes which are implicated in human disease, in this case psychiatric disease, across global human populations, towards the end of providing some new insight for gene mapping efforts, and on the other hand the study of genetic variation in this set of genes may reveal traces of the population history events undergone, including possible evidence for selection.

Many of the common psychiatric and behavioural disorders burdening society in the 21st century have been shown to have a genetic component, and this genetic component is thought in many cases to have a complex mode of transmission. This complexity has hindered attempts to pinpoint the genes involved in disorders such as Schizophrenia, Bipolar disorder and Autism. The set of genes of the current study have all been implicated in psychiatric disease, in many cases being associated with more than one disorder. By elucidating the genetic variation of these genes we hoped to shed light on a major potential confounding factor of genetic association studies for psychiatric disorders involving these genes: that of the variation between studies due to population based factors. Describing population specific patterns of variation can provide a guide for future studies in similar populations.

In addition we consider the evolutionary history of these genes and look for traces of natural selection via analysis of the genetic variation. Shedding light on possible selective events and the gene and evolutionary history of genes involved in

human neurological function is obviously of great interest, as is the study of population based variation in such genes. Also discovering evidence for selection may uncover environmental factors which interact with the genetic component underlying psychiatric and behavioural disorder susceptibility.

SAMPLES

The Human Genome Diversity Panel

The Human Genome Diversity Panel (HGDP) (Cann et al. 2002) contains 1064 individuals representing 51 worldwide populations Figure 7. This set of populations covers most of the complete human genetic diversity, as reported by Rosenberg et al (Rosenberg et al. 2002). Samples were regrouped into 39 populations based on geographic and ethnic criteria to avoid small sample size. Tuscans and North Italians were grouped as Continental Italians; Dai, Lahu, Miaozu, Naxi, She, Tujia and Yiku populations were combined as South Chinese (SCH); Tu Uygur and Xibo populations as North West Chinese (NWC) and Daur, Hezhen, Mongolian and Orogen as North East Chinese (NEC). The total number of populations was thus 39.

The panel used was the H1048 according to Rosenberg *et al* (Rosenberg 2006) which includes 1,048 individuals and omits a number of sample duplications and errors from the original panel. The 39 populations were grouped into 7 continental regions for the purpose of analysis, see Table 3.

The population and continental groupings were the same for all the analyses involving the HGDP, with one exception: For the first Neuregulin 1 study (Gardner et al. 2007) an additional European population was included: the Catalan population (CAT) and the seven North Chinese populations were grouped together as one combined population: North Chinese (NCH), instead of two as in all of the other analysis, thus in the first NRG1 study the total populations was also 39.

Figure 7

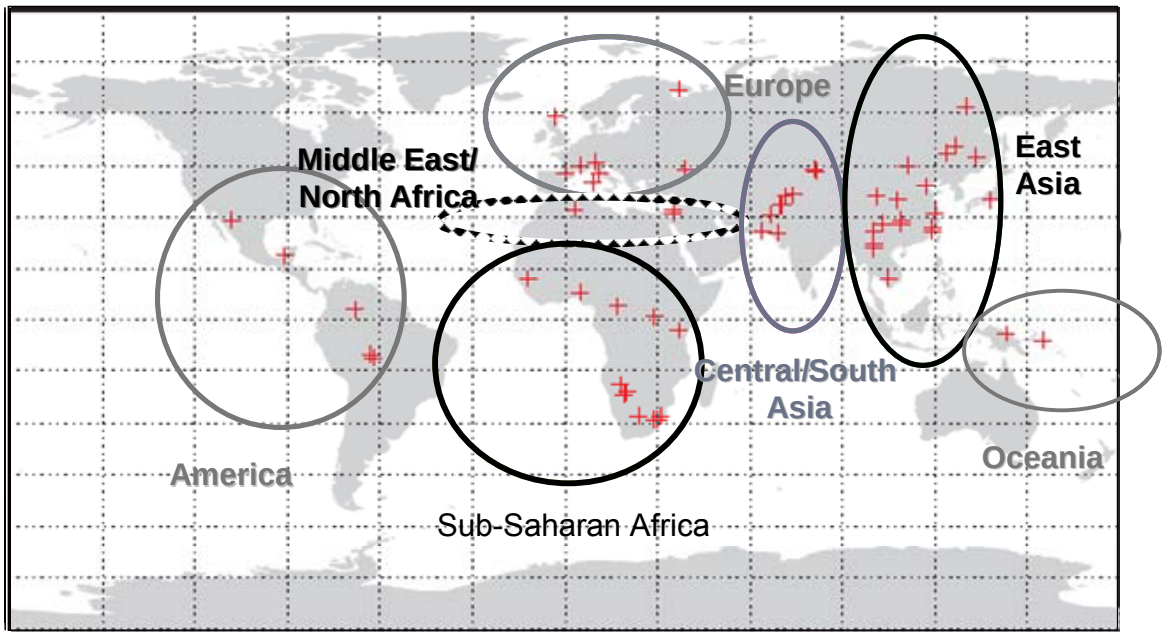

Europe
East
Asia
Middle East/
North Africa
Central/South
Asia
Oceania
America
Sub-Saharan Africa

Table 3

Continental Region	Population	N
America		
	Colombian	13
	Karitiana	24
	Maya	25
	Pima	25
	Surui	21
Central/South Asia		
	Balochi	25
	Brahui	25
	Burusho	25
	Hazara	24
	Kalash	25
	Makrani	25
	North West China	29
	Pathan	25
	Sindhi	25
East Asia		
	Cambodian	11
	Han	44
	Japanese	29
	North East China	39
	South China	70
	Yakut	25
Europe		
	Adygei	17
	Basque	24
	French	28
	North Italy	21
	Orcadian	16
	Russian	25
	Sardinian	28
Middle East/North Africa		
	Bedouin	48
	Druze	47
	Mozabite	30
	Palestinian	51
Oceania		
	NAN Melanesian	19
	Papuan	17
Sub Saharan Africa		
	Bantu	20
	Biaka Pygmies	32
	Mandenka	24
	Mbuti Pygmies	15
	San	7
	Yoruba	24

Spanish Schizophrenia/Psychosis Family Sample Set

The Sample set for the Schizophrenia/Psychosis association study comprised of a total of 575 Spanish Individuals from 151 nuclear families. Of these families, 29 were triads: an affected individual and both parents, see Figure 8a and 122 were quadruplets: an affected individual with both parents and one healthy sibling Figure 8b. The 151 affected individuals (113 males and 38 females with mean age= 24.4 years, SD= 6.5) were all affected by DSM-IV schizophrenia and schizophrenia spectrum disorders.

Figure 8

(a) **(b)**

The families were ascertained through 151 patients affected by schizophrenia spectrum disorders, recruited from:

- Psychiatric Unit, Virgen del Camino Hospital, Pamplona (95 quadruplets).
- Mental Health Center of Psychiatry and Drug Addiction Service of the Hospital Santa Maria in Lleida (27 quadruplets).
- Adolescents Area of Benito Menni Assistance Complex in Mental Health in Sant Boi de Llobregat. Barcelona (29 trios).

Information for diagnoses was derived from all available sources of information during the index admission, information provided by close relatives and

a comprehensive review of medical records. Based on information from these sources, DSM-IV diagnoses were generated. Clinical symptoms were assessed using CASH (The comprehensive assessment of symptoms and history, (Andreasen et al. 1992)). CASH was also used in siblings and parents to confirm lifetime absence of any psychotic disorder.

The mean duration of the illness in patients was 5 years (5.48 years). All of the subjects were invited to participate in the study and provided with written informed consent after procedures had been fully explained.

RESULTS

Chapter 1: Extreme Population differences across Neuregulin 1 gene with Implications for association studies

M Gardner, A Gonzalez-Neira, O Lao, F Calafell, J Bertranpetit and D Comas.

Molecular Psychiatry (2005), 1–10

Gardner M, Gonzalez-Neira A, Lao O, Calafell F, Bertranpetit J, Comas D.
Extreme population differences across Neuregulin 1 gene, with implications for association studies.
Molecular Psychiatry. 2006 Jan;11(1):66-75.

Chapter 2: Family-Based Association Study of Neuregulin-1 Gene and Psychosis in a Spanish Sample.

Araceli Rosa, **M. Gardner**, M.J. Cuesta, V. Peralta, M. Fatjo-Vilas, S. Miret, E. Navarro, D. Comas, and L. Fanañas.

Am J Med Genet (Neuropsychiat. Genet.), In Press.

Rosa A, Gardner M, Cuesta MJ, Peralta V, Fatjo-Vilas M, Miret S, Navarro ME, Comas D, Fananas L.
Family-based association study of neuregulin-1 gene and psychosis in a Spanish sample.
American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics. 2007 May 14; [Epub ahead of print]

Chapter 3: Extreme individual marker F_{ST} values do not imply population-specific selection in humans: the NRG1 example.

Michelle Gardner, Scott Williamson, Ferran Casals, Elena Bosch, Arcadi Navarro, Francesc Calafell, Jaume Bertranpetit, David Comas.

Human Genetics (April 2007).

Gardner M, Williamson S, Casals F, Bosch E, Navarro A, Calafell F, Bertranpetit J, Comas D.
Extreme individual marker F(ST)values do not imply population-specific selection in humans: the NRG1 example.
Human Genetics. 2007 Jul;121(6):759-62. Epub 2007 Apr 25.

Chapter 4: Worldwide Genetic Variation and tagSNP Transferability in Dopamine and Serotonin pathway genes.

Michelle Gardner and David Comas.

Manuscript in preparation.

Worldwide Genetic Variation and tagSNP Transferability in Dopamine and Serotonin pathway genes.

Michelle Gardner and David Comas.

Unitat de Biologia Evolutiva, Universitat Pompeu Fabra,

Doctor Aiguader 88, 08003 Barcelona, Spain.

Abstract

The dopamine and serotonin systems are two of the most important neurotransmitter pathways in the human nervous system with their role in controlling behavior and mental status well accepted from studies showing that these functional pathways are the target of antidepressant and psychoactive drugs. Genes from both systems have been widely implicated in psychiatric and behavioural disorders, with numerous reports of associations and almost equally as numerous reports of failure of replication. We investigate whether this failure of replication in the context of the twenty one dopamine and serotonin genes presented may have as its basis population based differences in allele frequencies and linkage diseqilibrium in a set of worldwide populations. Our findings suggest that reasons other than ethnic differences in genetic variation may explain the discrepancies. We also examine the transferability of tagSNPs defined in the HapMap populations to distinct worldwide populations thereby providing a guide for future association studies with these genes.

Introduction

The dopamine and serotonin systems are two of the major neurotransmitter systems in humans. Genes from both have been implicated and associated with many neurological and psychiatric disorders, ranging from addiction to Tourette syndrome including many others in between, notably Parkinson's disease and schizophrenia [1]. Dopamine affects brain processes that control both motor and emotional behaviour. Its role in the response and the ability to experience pleasure and pain is well documented and this has made many genes involved in the dopamine pathway candidates for genetic association studies of addiction [2]. In addition, the so-called 'dopamine hypothesis' remains one of the chief theories in the search for the molecular basis of schizophrenia. Brain dopamine receptors are also the primary targets in the treatment of Parkinson's disease and Huntington's chorea [3-5]. Serotonin has a critical role in temperature regulation, sensory perception, locomotion, and sleep and mood control. It is involved in the pharmacology of depression and psychosis and is in a way the 'second in line' as being involved in the causation of schizophrenia, the 'serotonin hypothesis' [6]. Genes within the serotonin pathway have also been associated with suicidal behaviour [7], bipolar disorder [8] and various other behavioural disorders.

Most psychiatric diseases and behavioural disorders are complex genetic traits with both genetic and environmental factors influencing an individual's risk of being affected. The heritability estimate for some of the commonly studied psychiatric disorders ranges from 0.28 for depression to 0.90 for autism [9]. This complex pattern of inheritance complicates efforts to identify the genes contributing to the genetic component of an individual's susceptibility to the disease. Despite the existence of, in some cases, hundreds of studies for an individual dopamine or

serotonin system gene purported as a candidate for one of the various diseases and disorders that the gene has been suggested as a candidate for, there has been trouble reaching consensus and in the replication of positive findings in an association study. This reflects the problem of pinpointing genes involved in complex diseases in general and it is also complicated by specific characteristics of psychiatric disease such as the definition of clinical phenotypes. In fact, from the many studies done for each of the many genes for each of the many diseases for which they are candidates very few examples of a confirmed association stand out [10] [11].

One motive for inconsistencies in replication studies may be variability in association among different populations [12]. This could be due to real variance in disease incidence among populations (although for example in the case of schizophrenia, similar incidence across cultures, a prevalence of 1%, with the exception of some 'outlier' populations [13]), or it could also be due to genetic heterogeneity, that is different disease-causing alleles predominating in different populations. Crucially, population differences in linkage disequilibrium (LD) and allele frequencies (of marker or disease causing variant), the result of the distinct demographic history of the populations under study (Genetic drift, migration and bottle necks) may play a role [13], [14], [15].

LD has been found to be variable within and among loci and populations [16,17], and this in the context of association studies potentially means differences in LD between causal variant and marker SNP in different populations. Since most association studies rely on the LD between marker and disease causing variant to detect association, this is a possible contributing factor to failure of replication in different study populations. Variability in disease or marker allele frequencies between populations is another possible confounding factor, as mismatches between

marker and disease allele frequencies may reduce the statistical power to detect associations between the complex trait and the candidate gene [15].

Analysis of common SNPs now forms the basis of current approaches to identify the genetic basis of complex disease. The widely accepted 'common variant, common disease'[18] hypothesis holds that high frequency (>1%) low penetrant alleles underlie the genetic component of complex diseases such as psychiatric disorders. There are estimated to be more than 10 million (common) SNPs in the human genome [19]. Systematic studies of these genetic variants for association with disease are facilitated by a certain redundancy in SNP genotyping due to LD, that is the tendency of certain alleles to be inherited together. Recent large scale human genetics projects, such as HapMap, aimed at quantifying LD across the genome and in different populations have as one of their main aims the aiding of the reduction of the number of SNPs necessary for typing in association studies [20]. These so called 'tagSNPs' can reduce greatly the cost and time involved in genetic association studies.

In this study we aim to describe such population based genetic variability in allele frequencies and linkage disequilibrium across a set of world wide populations in 21 Dopamine and Serotonin genes commonly tested for association to psychiatric and behavioural disorders. In addition, we investigate the applicability of the tagSNPs selected in HapMap populations to the more diverse set of human populations of the present study. In this way, we provide a guide for future association studies based on tagSNP selection from HapMap populations applied to other worldwide populations. The LD quantity of the genes in worldwide populations along with allele frequency differences are also considered and discussed.

Materials and Methods

Samples

The sample set consists of the Human Genome Diversity Panel (HGDP) [21], a set of 1,064 purified DNA samples from worldwide populations covering most of the whole human genetic diversity as reported by Rosenberg *et al* [22]. From the original panel, several duplicated samples were used as internal controls and some atypical individuals were omitted. The panel used is the H1048 according to Rosenberg *et al* [23] which includes 1,048 individuals. Samples were regrouped into 39 populations based on geographic and ethnic criteria to avoid small sample size. Populations were also categorised into seven broad continental regions: Sub Saharan Africa (SSAFR), Middle East/North Africa (MENA), Europe (EUR), Central/South Asia (CSASIA), East Asia (EASIA), Oceania (OCE) and America (AME), see Gardner *et al* [14] for the classification of populations into continental groupings.

Gene choice, SNP selection and genotyping

The criteria for inclusion in the study of a gene from one of the two neurotransmitter pathways was a PubMed published positive association of the gene with a psychiatric disease/behavioural trait, in most cases multiple reports of positive associations with sometimes more than one psychiatric disease or behavioural trait. Therefore, we present a study which includes although not exhaustively, all of the important dopamine and serotonin system genes that have been associated with psychiatric disease/behavioural traits, a total of 21 genes (Table 4). SNPs were chosen at a density of every 5-10kb within the genes. Additionally, a number of SNPs were selected up to 30kb flanking 3' and 5' of each gene, if possible at 30kb, 20kb, 10kb and 5 kb flanking either side. All the SNPs were chosen on the basis of

being involved in the HapMap project, having a high Illumina score (at least over 0.6) and also a minimum minor allele frequency of at least 0.1 in one of the HapMap populations. Preference also was for SNPs 'goldengate validated' as opposed to '2 hit validated'. In addition to the above, any coding SNPs with a reasonable Illumina score were chosen regardless of the other criteria applied to non-coding SNPs.

Genotyping was carried out using the Beadarray® Platform (Illumina Inc., San Diego, USA.) [24] according to the manufacturer's protocol.

Data Analysis

The F_{ST} (a measure of population differentiation based on allele frequencies) was calculated using the Arlequin [25] program for each of the 21 genes at the population and continental level. The average of the F_{ST} values for the 18 autosomal genes (excluding the genes in the X chromosome in order to avoid genome biases) was compared to averages for other F_{ST} value distributions (all autosomal genes).The F_{ST} value distributions were from the following datasets: the Marshfield Indel set (210 biallelic anonymous markers), the 427 gene based SNPs in the ALFRED database [26] (a similar set of 38 worldwide populations as the ones used in the present analysis), and a set of 121 SNPs from a 'neutral' gene free region of Chromosome 22 [27]. In order to provide a graphical view of the population differences, F_{ST} genetic distances [28] between individual SNPs were calculated and represented in a multidimensional scaling (MDS) plot.

SNPator, the SNP data management program, (Morcillo et al, unpublished data) was used for SNP data handling and basic data analysis such as calculation and filtering of genotype failures and allele frequency and related calculations including Hardy-Weinberg equilibrium. The phasing of the genotypes using the Bayesian

algorithm based PHASE program [29] for haplotype reconstruction was carried out within the framework of the SNPator program. The default settings for PHASE were used and the number of iterations was set to 1,000. In addition, independently of SNPator the Haploview program [30] was used for visualisation of linkage disequilibrium (LD) and haplotype block structure, loading the phased haplotypes. The quantity of LD based on the LD measurement r^2 was calculated. The r^2 values for marker pairs were obtained from Haploview. The quantity of LD was compared between continental regions for the adjacent marker pair r^2 values, using the Friedman test (between all continental groups together) and the Wilcoxin test (pair wise comparisons) within the SPSS package.

TagSNP Transferability

We compared the efficiency of the set of tagSNPs defined in each of the three HapMap populations (CEPH, Europe; Japanese and Chinese analysed together as Asian; and Yoruban, Africa) as applied to the corresponding subset of continental groups of the 39 populations of the Human Genome Diversity Panel. The set of tagSNPs defined in the HapMap CEPH population was applied to the European (EUR), Middle East/North Africa (MENA) and Central/South Asia (CSASIA) continental grouping of the HGDP; HapMap Asian tagSNPs were applied to East Asian (EASIA), Oceanic (OCE) and American (AME) groups, and the Yoruban HapMap tagSNP set was applied to the Sub-Saharan African (SSAFR) group.

The Tagger program [31] as implemented in the Haploview program [30] was used for tagSNP selection. Tagger selects tagSNPs according to the Carlson et al [32] algorithm. For each of the HapMap populations, the phased HapMap data was uploaded to Haploview, with the exception of the three X chromosome genes, for

which the unphased HapMap data was uploaded. The genotyped SNPs in each gene were included as the pool of SNPs from which the tagSNP set would be defined. A minimum minor allele frequency of 0.001 was applied (the default value) and the r^2 threshold was set to 0.8. The obtained tagSNP set for each of the HapMap populations was applied to the comparable HGDP populations (see above) thus: each of the 39 populations for a gene was uploaded in Haploview in turn. The HapMap tagSNP sets obtained were applied in turn to each HGDP population. Thus, we obtained the mean r^2 value of the HapMap tagSNPs applied to each population, the percentage of SNPs with an $r^2 \geq 0.8$, and the number of tagSNPs.

Results

A total of 303 SNPs were successfully typed with more than 50% success in all populations (overall success rate of 90%). From these, 17 coding SNPs, including eleven non-synonymous, were genotyped. The final mean spacing across the 21 genes was approximately 8.36 kb (Table 1). No particular SNP or population showed significant deviation from Hardy-Weinberg after correcting for multiple testing.

Differentiation based on SNP allele frequencies at the population and continental level were tested with the F_{ST} statistic. In general, the F_{ST} values across the dataset were low as observed by a mean F_{ST} value of 0.106 (18 autosomal genes only), indicating low differentiation between populations. This value represents the lowest mean F_{ST} for any of the three F_{ST} distributions used for comparison, comprising genes involved in other genomic pathways. The average F_{ST} for X chromosome genes was higher than for autosomes (0.146), reflecting the lower effective population size (two copies in females and one in males) of this chromosome [20]. We also observed a higher F_{ST} in coding SNPs (mean F_{ST} value of 0.135) compared to non-coding SNPs, that is coding SNPs in genes exhibited a marked increase in population differentiation across populations, which concurs with previous reports [33].

To view graphically the genetic relationships between populations for the serotonin and dopamine autosomal genes, a multidimensional scaling (MDS) plot based on genetic distances was drawn (Figure 1). The plot shows populations grouped according to continental region, with higher intra-continental differences between American, Oceanic and African populations, whereas Europeans and Asians present high genetic homogeneity for serotonin and dopamine genes.

In the comparisons of LD between continental regions, a significant ($p<0.05$) result comparing all continental groups (by means of the Friedman test) or performing pair wise comparisons (Wilcoxin test) would imply a difference in the amount of LD. Eight out of 21 of the genes showed significant differences in quantity of LD between continental regions ($p<0.05$, Friedman test) after correction for multiple testing. This was mainly explained by the differences exhibited by the *HTR4* and *HTR2C* genes in LD quantity between a number of continental groups as seen with the Wilcoxin pair wise analysis (significant after correction for multiple testing). In addition the *COMT* gene also showed significant differences between Sub Saharan African populations and the rest of the world populations. The contribution of the Sub Saharan Africans to these differences was confirmed by looking at the LD structure, where there was a lower amount of LD than in other continents. The *HTR2C* gene, located in the X chromosome, is almost a contiguous stretch of LD in most continental groups except for Sub-Saharan African. This result should be interpreted with caution however, as the X chromosome in general is known to have higher LD [34] due to its lower effective population size.

TagSNP transferability

Of the 303 SNPs genotyped, the number of tagSNPs for each HapMap sample was 173 for Europeans (CEPH sample), 167 for Asians (Chinese and Japanese) and 213 for Africans (Nigerian Yoruba). This greater number of tagSNPs in Africans is in agreement with previous results [32,35,36] due to the lower LD in Africa. Thus, the average "tagging efficiency" (total number of SNPs genotyped divided by the number of tagSNPs in each sample) is lower in African samples (average 1.4), compared to Europeans and Asians (average ~1.8) (Table 2). The tagging efficiency

differs substantially depending on the gene analysed, ranging from 1 (the lowest efficiency) in the case of *TH* gene to 9.75 in the case of *HTR2C* in Europeans, where only four SNPs would be needed to capture the 39 SNPs genotyped in the gene with an $r^2 \geq 0.8$.

Table 3 shows a summary of the results of the tagSNP transferability analysis for the three HapMap populations against the comparable HGDP samples. More detailed results by population available on request. The dopamine and serotonin tagSNPs defined in the HapMap populations (CEPH-Europeans, Chinese/Japanese, and Yoruba) applied to HGDP populations grouped into three equivalent groups (see Gardner *et al* [14] for the grouping categories) capture, as an average, more than 80% of the SNPs genotyped: 85.6% of the SNPs in the European group, 82.9% in the Asian group, and 83.4% in the African group. Furthermore, the results obtained for populations within each continental region were extremely similar (data not shown) indicating that the tagSNPs defined in the HapMap populations can be applied to any of their corresponding population samples.

Discussion

One of the chief aims of the present study was to shed light on the population related reasons underlying the lack of consistent replication of association findings in the context of the dopamine and serotonin genes with psychiatric disease. The results found analysing 21 genes in 39 worldwide samples demonstrate that the dopamine and serotonin SNP frequencies across populations are no more differentiated across populations than other regions of the genome, as shown by the average F_{ST} values. However, coding SNPs showed higher variance in allele frequencies across populations and continents, in agreement with previous reports [33], as did SNPs on the X chromosome. In addition, none of the populations studied was an outlier of note with regard to allele frequencies, and populations showed quite some homogeneity within continental groups as seen in the MDS plot of the average F_{ST} values per population.

Another remarkable result of the present study is the population homogeneity in dopamine and serotonin linkage disequilibrium within geographical regions, with LD being lower in Sub Saharan African populations in agreement with previous reports [27, 33,37]. One of the possible causes behind the lack of replication of association studies is the variation of LD among the different populations studied. The LD approach to gene mapping assumes that the associated variant is not causal but is in high LD with such a disease causing variant. Our results show that the failure to replicate a finding of association involving dopamine and serotonin genes between populations of the same geographical region to be due to other factors (such as missclassification of phenotypes, lack of power, or chance) other than population differences in LD between the markers analysed and the causal variant. This LD

homogeneity in serotonin and dopamine genes within continental regions is in agreement with previous results in other genomic regions [27,35].

The International HapMap project has as one of its main aims the characterisation of linkage disequilibrium across different genomic regions towards the end of facilitating genetic association studies [20]. A definitive answer on the applicability of the findings from the three HapMap populations (a European, African and Asian population) to other worldwide populations has not been yet reached although preliminary analyses suggest it is promising [38, 35,36]. The utilisation of tagSNPs in association studies takes advantage of LD to reduce the number of SNPs typed, allowing a reduction of cost and time. We addressed this issue in the context of a set of dopamine and serotonin genes commonly examined for association to psychiatric disease and with many reported associations including a few confirmed associations. Our results show differences in tagging efficiency depending on the amount of LD. African populations as compared to European or Asians show a lower tagging efficiency in the present set of genes, and therefore, more SNPs should be typed for these populations when performing an association study with dopamine and serotonin genes. Furthermore, the LD differences across genomic regions cause tag efficiency differences depending on the gene analysed: the genotyping effort in *HTR2C* can be reduced in all populations due to the presence of high LD, whereas other genes such as *TH* exhibit a low efficiency and therefore more SNPs would be nessecary to ensure sufficient coverage.

The transferability of tagSNPs defined in the HapMap samples to other populations not represented in HapMap is a crucial issue. Initial studies of transferability of tagSNPs defined in HapMap to the HGDP worldwide population dataset for other genomic regions have indicated that HapMap data could be used as

a good proxy to the diversity found in other populations [35,36]. Our results for serotonin and dopamine genes reinforces this transferability of HapMap data to other population samples. Association studies containing European, Middle Eastern, North African, and South/Central Asian samples may therefore use the HapMap CEPH-European tagSNPs in dopamine and serotonin pathway genes since our present results show a clear homogeneity of these samples in allele frequencies and LD. In the same way, association studies involving East Asian, Amerindian, and Oceanic samples may use the HapMap Asian tagSNPs, and finally HapMap Yoruban tagSNPs may be used in association studies containing African samples.

Websites

SNPator: http://bioinformatica.cegen.upf.es

HapMap: www.hapmap.org

Haploview: http://www.broad.mit.edu/mpg/haploview/

Acknowledgements

We are grateful to Anna González-Neira in the Centro Nacional de Genotipación, Madrid for help with the SNP genotyping; and Howard Cann for his effort in providing the CEPH samples. We would like to thank Francesc Calafell for suggestions on the analysis. We thank Ferran Casals, Elena Bosch, Tomàs Marquès, Andrés Moreno and Anna Ferrer for technical help. The present study was supported by the Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain (BFU2006-01235/BMC), Direcció General de Recerca, Generalitat de Catalunya (2005SGR/00608) and "Fundación Genoma España" (proyectos piloto CEGEN 2004-2005). M.G. received a fellowship (2004FI-IQUC2/0047) from the International Graduate School of Catalunya (Generalitat de Catalunya).

References

1. Ueno, S. Genetic polymorphisms of serotonin and dopamine transporters in mental disorders. *J Med Invest* **50**, 25-31 (2003).
2. Kidd, K. K. et al. A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. *Hum Genet* **103**, 211-27 (1998).
3. Ichikawa, J. & Meltzer, H. Y. Relationship between dopaminergic and serotonergic neuronal activity in the frontal cortex and the action of typical and atypical antipsychotic drugs. *Eur Arch Psychiatry Clin Neurosci* **249 Suppl 4**, 90-8 (1999).
4. Nieoullon, A. Dopamine and the regulation of cognition and attention. *Prog Neurobiol* **67**, 53-83 (2002).
5. Backman, L. & Farde, L. Dopamine and cognitive functioning: brain imaging findings in Huntington's disease and normal aging. *Scand J Psychol* **42**, 287-96 (2001).
6. Aghajanian, G. K. & Marek, G. J. Serotonin model of schizophrenia: emerging role of glutamate mechanisms. *Brain Res Brain Res Rev* **31**, 302-12 (2000).
7. Bondy, B., Buettner, A. & Zill, P. Genetics of suicide. *Mol Psychiatry* **11**, 336-51 (2006).
8. Craddock, N., Dave, S. & Greening, J. Association studies of bipolar disorder. *Bipolar Disord* **3**, 284-98 (2001).
9. Merikangas, K. R. & Risch, N. Will the genomics revolution revolutionize psychiatry? *Am J Psychiatry* **160**, 625-35 (2003).
10. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet Med* **4**, 45-61 (2002).
11. Owen, M. J., Cardno, A. G. & O'Donovan, M. C. Psychiatric genetics: back to the future. *Mol Psychiatry* **5**, 22-31 (2000).
12. Colhoun, H. M., McKeigue, P. M. & Davey Smith, G. Problems of reporting genetic associations with complex outcomes. *Lancet* **361**, 865-72 (2003).
13. Jablensky, A. The 100-year epidemiology of schizophrenia. *Schizophr Res* **28**, 111-25 (1997).

14. Gardner, M. et al. Extreme population differences across Neuregulin 1 gene, with implications for association studies. *Mol Psychiatry* **11**, 66-75 (2006).
15. Zondervan, K. T. & Cardon, L. R. The complex interplay among factors that influence allelic association. *Nat Rev Genet* **5**, 89-100 (2004).
16. Pritchard, J. K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* **69**, 1-14 (2001).
17. Gabriel, S. B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
18. Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
19. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nat Genet* **27**, 234-6 (2001).
20. Consortium, I. H. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
21. Cann, H. M. et al. A human genome diversity cell line panel. *Science* **296**, 261-2 (2002).
22. Rosenberg, N. A. et al. Genetic structure of human populations. *Science* **298**, 2381-5 (2002).
23. Rosenberg, N. A. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* **70**, 841-7 (2006).
24. Oliphant, A., Barker, D. L., Stuelpnagel, J. R. & Chee, M. S. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* **Suppl**, 56-8, 60-1 (2002).
25. Schneider, S., Roessli, D., Excoffier, L. . Arlequin ver. 2.0: A software for population genetic data analysis. *Genetics and Biometry Laboratory, University of Geneva, Switzerland.* (2000).
26. Cheung, K. H. et al. ALFRED: a Web-accessible allele frequency database. *Pac Symp Biocomput*, 639-50 (2000).

27. Gonzalez-Neira, A. et al. Geographic stratification of linkage disequilibrium: a worldwide population study in a region of chromosome 22. *Hum Genomics* **1**, 399-409 (2004).
28. Reynolds J, W. B., Cockerham C Genetics Estimation for the coancestry coefficient: basis for a short-term genetic distance., 767-779 (1983).
29. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).
30. Brookes, K. et al. 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**, 934-53 (2006).
31. de Bakker, P. I. et al. Efficiency and power in genetic association studies. *Nat Genet* **37**, 1217-23 (2005).
32. Carlson, C. S. et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* **74**, 106-20 (2004).
33. Hinds, D. A. et al. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072-9 (2005).
34. Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, A.S., Daly, M.J., and Donnelly, P. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
35. Gonzalez-Neira, A. et al. The portability of tagSNPs across populations: a worldwide survey. *Genome Res* **16**, 323-30 (2006).
36. Gu, S. et al. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet* (2007).
37. Reich, D. E. et al. Linkage disequilibrium in the human genome. *Nature* **411**, 199-204 (2001).
38. Ribas, G. et al. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum Genet* **118**, 669-79 (2006).

Legends

Figure 1. Multidimensional scaling plot of the F_{ST} population values for the serotonin and dopamine genes. Each point corresponds to the average F_{ST} values for a population.

Table 4. Genes in dopamine and serotonin pathways analysed in worldwide populations

Gene	Pathway	Chromosome	Gene length (kb)	No. SNPs	SNP spacing (kb)	No. Coding SNP (nonsyn)
COMT	Dopamine	22	27254	16	4.89	1 (1)
DBH	Dopamine	9	22969	16	4.81	4 (3)
DDC	Dopamine	7	102615	21	7.74	2 (1)
DRD1	Dopamine	5	3126	8	8.40	1 (1)
DRD2	Dopamine	11	65575	16	7.40	-
DRD3	Dopamine	3	50199	14	8.00	1 (1)
DRD4	Dopamine	11	3398	6	8.36	-
DRD5	Dopamine	4	2029	7	8.09	-
HTR1A	Serotonin	5	1268	6	10.49	-
HTR1B	Serotonin	6	1259	10	17.25	2 (2)
HTR2A	Serotonin	13	62665	18	9.91	-
HTR2C	Serotonin	X	326073	39	6.14	1 (1)
HTR4	Serotonin	5	203037	32	7.23	-
MAOA	Dopamine	X	90601	12	7.86	1
MAOB	Dopamine	X	115835	10	9.91	-
PPP1R1B	Dopamine	17	9696	7	6.36	-
SLC6A3	Dopamine	5	52636	16	6.93	2
SLC6A4	Serotonin	17	37799	14	9.72	-
TH	Dopamine	11	7875	9	8.43	1 (1)
TPH1	Serotonin	11	21113	9	10.01	-
TPH2	Serotonin	12	93595	17	7.66	1

Table 2. Average tagging efficiency of HapMap populations in dopamine and serotonin genes.

	EUROPEAN	ASIAN	AFRICAN
COMT	1.45	1.45	1.23
DBH	1.33	1.78	1.33
DDC	1.31	1.91	1.4
DRD1	1.6	1.33	1.33
DRD2	1.45	1.6	1.23
DRD3	1.14	1.56	1
DRD4	1.2	1.2	1.2
DRD5	2.33	1.17	1.4
HTR1A	6	2	2
HTR1B	1.67	1.25	2
HTR2A	1.16	1.38	1.29
HTR2C	9.75	4.88	2.05
HTR4	2.46	1.68	1.52
MAOA	4	6	2.4
MAOB	1.43	1.67	1.43
PPP1R1B	1.4	7	1.4
SLC6A3	1.45	2.29	1.33
SLC6A4	2	2.33	1.27
TH	1	1	1
TPH1	1.5	1.29	1.29
TPH2	1.7	1.55	1.42

Table 3. tagSNP parameters between continental groups for HGDP samples.

'Number of tagSNPs' is the number of tagSNPs of the total genotyped SNPs defined in the HapMap population as tagging all other genotyped SNPs with $r^2 \geq 0.8$,

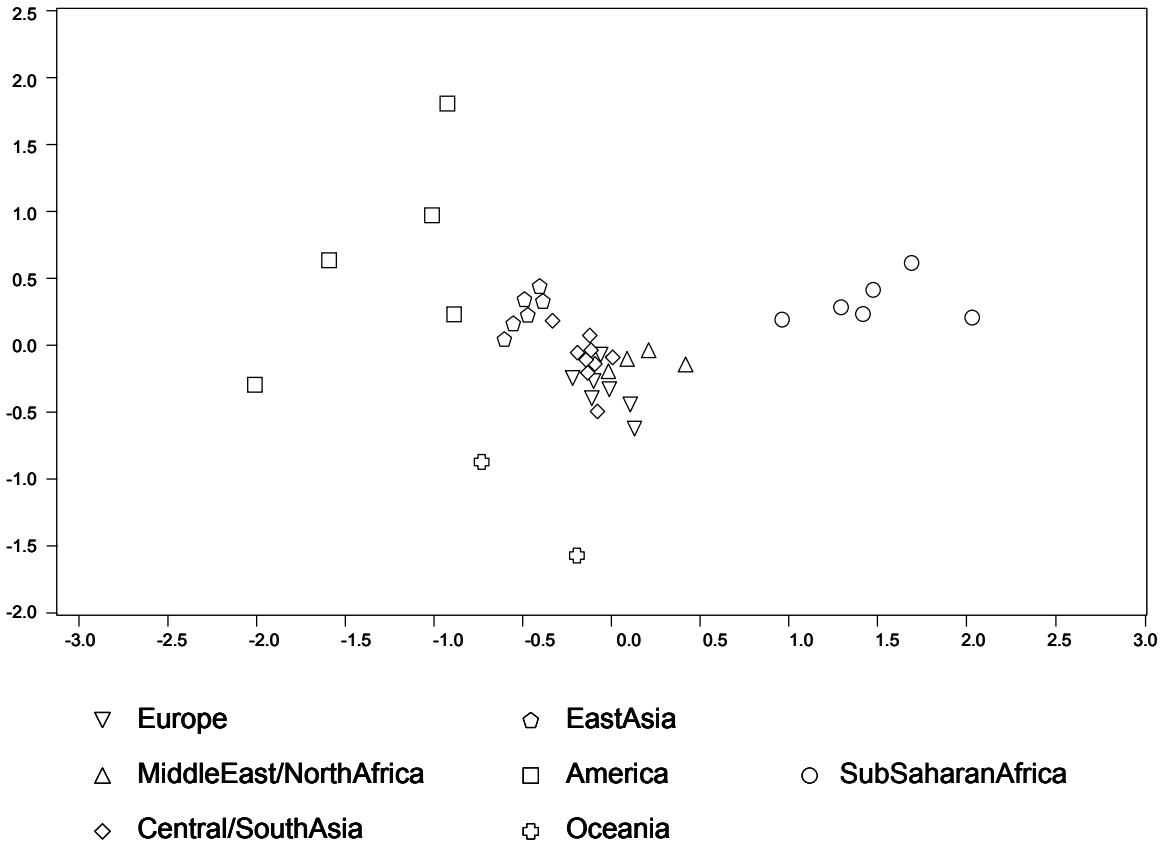
'average r^2' is the average r^2 when the HapMap defined tagSNPs are applied to all genotyped SNPs.

'% SNPs captured' is the percent of SNPs captured of all SNPs genotyped with $r^2 \geq 0.8$**.**

		EUR	MENA	CSASIA	EASIA	AME	OCE	SSAFR
COMT	Number of tagSNPs	11	11	11	11	11	11	13
	average r^2	0.91	0.90	0.89	0.90	0.86	0.92	0.90
	% SNPs captured	79.71	72.00	72.44	86.83	86.60	84.50	84.50
DBH	Number of tagSNPs	12	12	12	9	9	9	12
	average r^2	0.92	0.88	0.88	0.70	0.75	0.63	0.77
	% SNPs captured	90.57	79.75	84.89	67.83	72.60	62.00	75.00
DDC	Number of tagSNPs	16	16	16	11	11	11	15
	average r^2	0.98	0.98	0.97	0.87	0.85	0.88	0.86
	% SNPs captured	93.86	95.00	93.56	83.50	81.00	86.00	80.83
DRD1	Number of tagSNPs	5	5	5	6	6	6	6
	average r^2	0.76	0.73	0.75	0.77	0.80	0.79	0.78
	% SNPs captured	69.43	62.00	70.67	75.00	77.60	75.00	75.00
DRD2	Number of tagSNPs	11	11	11	10	10	10	13
	average r^2	0.96	0.94	0.96	0.94	0.95	0.97	0.91
	% SNPs captured	93.14	91.00	93.33	92.00	92.80	97.00	88.83
DRD3	Number of tagSNPs	12	12	12	9	9	9	14
	average r^2	0.96	0.95	0.97	0.93	0.88	0.90	1
	% SNPs captured	93.00	93.00	95.33	86.00	80.40	75.00	100
DRD4	Number of tagSNPs	5	5	5	5	5	5	5
	average r^2	0.95	0.97	0.95	0.99	0.87	0.98	0.91
	% SNPs captured	92.71	91.50	88.67	100	86.40	91.50	85.83
DRD5	Number of tagSNPs	3	3	3	6	6	6	5
	average r^2	0.91	0.84	0.88	0.98	1	0.93	0.91
	% SNPs captured	88.00	71.25	74.56	95.33	100	93.00	88.33
HTR1A	Number of tagSNPs	1	1	1	3	3	3	3
	average r^2	0.98	0.96	0.96	1	1	1	0.87
	% SNPs captured	100	95.75	90.56	100	100	100	83.33
HTR1B	Number of tagSNPs	6	6	6	8	8	8	5
	average r^2	0.80	0.80	0.80	1	0.98	1	0.70
	% SNPs captured	80.00	80.00	80.00	100	98.00	100	68.33
HTR2A	Number of tagSNPs	16	16	16	13	13	13	14
	average r^2	0.95	0.93	0.94	0.83	0.83	0.83	0.82
	% SNPs captured	91.14	86.25	89.00	74.83	79.80	80.50	78.83
HTR2C	Number of tagSNPs	4	4	4	8	8	8	19
	average r^2	0.87	0.77	0.77	0.59	0.39	0.58	0.89
	% SNPs captured	84.00	66.00	67.67	57.33	36.80	57.50	84.50
HTR4	Number of tagSNPs	13	13	13	19	19	19	21
	average r^2	0.91	0.87	0.85	0.92	0.92	0.87	0.85

	% SNPs captured	88.14	72.00	70.56	89.83	92.20	81.50	78.50
MAOA	Number of tagSNPs	3	3	3	2	2	2	5
	average r^2	0.96	0.90	0.93	0.88	0.83	0.94	0.95
	% SNPs captured	86.86	83.25	85.11	79.00	73.60	100	88.83
MAOB	Number of tagSNPs	7	7	7	6	6	6	7
	average r^2	0.93	0.89	0.91	0.78	0.70	0.72	0.76
	% SNPs captured	91.43	85.00	86.67	75.00	70.00	70.00	75.00
PPP1R1B	Number of tagSNPs	5	5	5	1	1	1	5
	average r^2	0.93	0.92	0.96	0.64	0.70	0.67	0.93
	% SNPs captured	87.86	82.25	93.78	57.00	68.20	57.00	83.50
SLC6A3	Number of tagSNPs	11	11	11	7	7	7	12
	average r^2	0.87	0.86	0.88	0.68	0.71	0.68	0.88
	% SNPs captured	79.29	78.00	79.00	62.17	62.20	62.00	80.00
SLC6A4	Number of tagSNPs	7	7	7	6	6	6	11
	average r^2	0.93	0.88	0.89	0.90	0.92	0.92	0.88
	% SNPs captured	90.00	84.25	83.67	86.00	91.60	93.00	86.00
TH	Number of tagSNPs	9	9	9	9	9	9	9
	average r^2	1	1	1	1	1	1	1
	% SNPs captured	100	100	100	100	100	100	100
TPH1	Number of tagSNPs	6	6	6	7	7	7	7
	average r^2	0.98	0.97	0.97	1	1	0.89	0.91
	% SNPs captured	100	94.50	95.11	100	100	89.00	90.83
TPH2	Number of tagSNPs	10	10	10	11	11	11	12
	average r^2	0.96	0.93	0.91	0.91	0.87	0.88	0.84
	% SNPs captured	93.14	83.50	78.44	87.00	84.40	82.00	76.17

Figure 1


2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Europe
EastAsia
MiddleEast/NorthAfrica
America
SubSaharanAfrica
Central/SouthAsia
Oceania

Chapter 5: The search for traces of Natural Selection in 'Brain Genes', a set of 21 Dopamine and Serotonin pathway genes, across worldwide populations.

Michelle Gardner, Ferran Casals, Elena Bosch, Francesc Calafell, Jaume Bertranpetit, David Comas.

Manuscript in preparation.

The search for traces of Natural Selection in 'Brain Genes', a set of 21 Dopamine and Serotonin pathway genes, across Worldwide populations.

Michelle Gardner, Ferran Casals, Elena Bosch, Francesc Calafell, Jaume Bertranpetit, David Comas.

Universitat Pompeu Fabra, Barcelona, Spain.

Abstract

The dopamine and serotonin systems are involved in both emotion and behaviour. As these functions show very human specific traits and individual differences in these genes are of extreme interest both for understanding species and individual uniqueness, discovering the evolutionary history including investigating for evidence of selection in these genes is of great value.

We present a study of twenty one dopamine and serotonin system genes, all playing vital roles in neurotransmission in the human nervous system. We investigate evidence of selection in these genes in a set of worldwide populations representing total human genetic diversity.

In general we find that these genes regulating human brain function are less differentiated across geographic regions than other large gene sets. However we report some indicators of the genetic variation in some of the genes having been shaped by population history forces which may include selection.

Introduction

The Dopamine and Serotonin systems are two of the major Neurotransmitter systems in humans. These two phylogenetically ancient neurotransmitters are intrinsic to human brain function and behaviour. Dopamine affects brain processes that control both motor and emotional behaviour. Its role in the response and the ability to experience pleasure and pain (the brains' rewards cicuitry) is well documented (Wise 2002), therefore making Dopamine genes candidates for genetic association studies of Addiction. Brain dopamine receptors are also the primary targets in the treatment of Parkinson's disease and Huntington's chorea (Cravchik and Goldman 2000). Serotonin has a critical role in temperature regulation, sensory perception, locomotion, and sleep and mood control. It is involved in the pharmacology of depression and psychosis (Cravchik and Goldman 2000). Genes from the Serotonin system have also been associated with suicidal behaviour, Bipolar disorder and various behavioural disorders (Hirschhorn et al. 2002). Both systems have been implicated in the aetiology of schizophrenia, the so called 'dopamine hypothesis' (Winterer and Weinberger 2004) or the 'serotonin hypothesis' (Aghajanian and Marek 2000).

Elucidation of the genes involved in human brain function and affecting human behaviour are obviously of great interest, both from the medical genetics point of view as many are implicated in psychiatric disease and behavioural disorders but also from an evolutionary genetics point of view. A greater understanding of the genes underlying many neurological and cognitive processes may lead to an understanding of human specific features as well as explaining some of the phenotypic differences observed between individuals.

Selective processes undergone during a gene's evolutionary past represent important factors influencing the genetic variation observable in the gene today, thus examining the extant genetic variation may allow us gain some insight into the genes evolutionary history including possible selective events (Sabeti et al. 2006). Recent advances in genomic sequencing and genotyping technologies have led to an explosion of publications on the topic of natural selection in humans. This is not at all surprising as interest in our evolution and the genetic differences that distinguish us from our nearest relatives, the apes, is high and has the potential to provide answers on the primordial question of just what makes us humans human. Moreover understanding which gene or portion of the genome is being driven to fixation by natural selection can tell us about speciation and the nature of adaptation. Understanding the genes underlying local selection, that is, those that may have facilitated the adaptation to new environments can reveal environmental factors that have been the focus of adaptive local selection. In the case of genes which are also implicated in disease causation, this may provide important information on environmental factors contributing to disease susceptibility.

Nowhere has the search for traces of selection more potential to reveal potentially human specific changes than in genes related to brain function. Phenotypic differences in human individuals with regard to behaviour, cognition and psychiatric disease, may also potentially be better understood as may the adaptations undergone and the environmental factors adapted to from the point of view of the functions carried out by the human brain, including cognition, social behaviour among others.

Here we present a study searching for traces of selection in a set of twenty one dopamine and serotonin neurotransmitter pathway genes. Shedding light on possible selective forces having acted on these genes which have been associated with many psychiatric and behavioural disorders is of great interest as it may shed some new light on the evolution of our species and the adaptation of different populations to new environments from the point of view of genes involved in brain function, specifically neurotransmission.

Materials and Methods

Samples

The sample set consists of the Human Genome Diversity Panel (HGDP) (Cann et al. 2002), a set of 1062 purified DNA samples covering most of the whole human genetic diversity (Rosenberg et al. 2002). The panel used was the H1048 according to Rosenberg *et al* (Rosenberg 2006) which includes 1,048 individuals and omits a number of sample duplications and errors from the original panel. The 39 worldwide populations of the HGDP were grouped into 7 continental regions for the purpose of analysis see Table 1. Furthermore a number of primate samples were included to allow the ancestral state of the SNPs to be determined. A total of 2 Chimpanzee (*Pan Troglodytes*) samples and 1 Gorilla (*Gorilla gorilla*) sample were typed with a high degree of success. Almost all of the SNPs were typed for at least one of the primates; therefore it was possible to determine with certainty the ancestral allele in almost all cases. In those cases where it was not possible to determine ancestral state by genotype, the information was gleaned from one of the following additional sources; dbSNP, USCS Genome browser or the Ancestral Allele database associated with the SWEEP^{TM} program. This consequently allowed Derived Allele Frequency (DAF) to be calculated.

Gene choice, SNP selection and genotyping.

Selection of genes for inclusion in the study was carried out via Entrez Gene. Search terms used were 'Dopamine, *Homo Sapiens*' and 'Serotonin, *Homo Sapiens*' to obtain the gene-lists for these terms as of mid April 2005 (14/04/2005). Thereafter the criteria for inclusion in the study of a gene from one of the two Neurotransmitter

systems was a Pub Med published positive association of the gene with a psychiatric disease/behavioural trait, in most cases multiple reports of positive associations with sometimes more than one psychiatric disease or behavioural trait (the exception being the *PPP1R1B* gene, which had such a functional role within the system as to warrant it of interest for inclusion in the study despite no positive reported association with disease as of the selection date). Therefore we present a study which includes although not exhaustively, all of the important dopamine and serotonin system genes that have been associated with psychiatric disease/behavioural traits, a total of 21 genes. See Table 2 for the list of genes included in the present study.

SNPs were chosen at a density of every 5-10kb within the genes. Additionally a number of SNPs were selected up to 30kb flanking 3' and 5' of each gene, if possible at 30kb, 20kb, 10kb and 5 kb flanking either side. All the SNPs were chosen on the basis of being involved in the HapMap project, having a high Illumina score (at least over 0.6) and also a minimum Minor Allele Frequency of at least 0.1 in one of the HapMap populations (frequencies from Illumina, based on HapMap pops). Preference also was for SNPs 'goldengate validated' as opposed to '2 hit validated'. In addition to the above any coding SNPs with a reasonable Illumina score were chosen regardless of the other criteria applied to non-coding SNPs.

Genotyping was carried out using the Beadarray® Platform (Illumina Inc., San Diego, USA.) (Oliphant et al. 2002). Genotyping was carried out according to the manufacturer's protocol.

SNPator, the SNP data management program, (Morcillo *et al*, unpublished data) was used for SNP data handling and basic data analysis such as calculation and

filtering of genotype failures and allele frequency and related calculations including Hardy Weinberg equilibrium.

Search for Traces of Positive Selection

A number of tests were carried out to test for evidence of positive selection based on allele frequencies, calculation of F_{ST} values and their subsequent comparison to other F_{ST} value distributions and a method based on assessing extended LD.

The Minor Allele Frequencies (MAF) and Derived Allele Frequencies (DAF) for the 21 Dopamine and Serotonin genes set were calculated in order to carry out the MAF and DAF selection tests. 'Minor Allele' was assigned in each population as the allele which was lower in frequency in that population. The proportion of the SNPs per gene with a MAF of >40% was plotted against the proportion of SNPs with MAF <10%. A similar analysis was carried out with the proportion of SNPs per gene with Derived allele frequency (DAF) >80% plotted against the proportion of SNPs with DAF <20% for the gene set. The results for each test for the twenty one dopamine and serotonin genes was plotted against the MAF and DAF values for a large gene set of 168 immune function related genes and simulated data from the same study (Walsh et al. 2005) for the purpose of comparison. The Walsh *et al* immune function gene study included three population samples: Ceph European (Ceu), Han Chinese (Han) and Yoruba African (Yri).

F_{ST} was calculated using the Arlequin (Schneider 2000) program for each of the 21 genes at the population and continental level. The distribution of the F_{ST} values for the 18 autosomal genes was compared to other F_{ST} value distributions (all autosomal genes). The F_{ST} value distributions were from the following datasets: A

set of 26 genes involved in Glycosylation (175 SNPs) (Ferrer-Admetlla *et al* in preparation), a set of 19 High Divergence genes (302 SNPs) (Moreno-Estrada *et al* in preparation), the Marshfield Indel set (210 biallelic anonymous markers), the ALFRED database (Cheung et al. 2000) 427 Gene based SNPs in a similar set of 38 worldwide populations, and a set of 121 SNPs from a 'neutral' gene free region of Chromosome 22 (Gonzalez-Neira et al. 2004). Comparisons were carried out within the SPSS package using parametric tests such as the ANOVA (analysis of variance) pair wise and for the complete set of the six distributions together, the non-parametric tests Krushkal-Wallis for *k* unrelated samples, and the Mann-Whitney and Kolmongorov-Smirnov for 2 samples.

The SWEEP^{TM} program (Sabeti et al. 2002) was used for investigation of possible signatures of positive selection. We set an arbitrary cutoff of 10kb for inclusion of genes in the analysis, as detection of long range extension of haplotypes requires a sufficient distance of typed SNPs. Taking all autosome genes, and those with SNPs typed over a distance of at least 10kb gave a total of 11 genes (*COMT, DBH, DDC, DRD2, DRD3, HTR2A, HTR4, SLC6A3, SLC6A4, TPH1* and *TPH2*). Phased haplotypes generated with PHASE (Stephens et al. 2001) for each gene and for each population served as input for the program, which is based on the Extended Haplotype method, EHH (Sabeti et al. 2002). To provide a larger data set from which to calculate empirical significance, the data from the present study was compared to the large dataset available from the 168 Immune function related gene study (Walsh et al. 2005). To further investigate possible traces of positive selection and selective sweeps for the genes of the present study, this time with the HapMap data, the Haplotter program which implements the integrated Haplotype score (iHS) (Voight

et al. 2006), a derivative of the EHH test of Sabeti et al, and which is applied to HapMap data only, was used to observe if signatures of selection in the genes of the present study had been detected with the HapMap data (with 1Mb flanking each side, in the three HapMap populations.

Results

Genotyping

A total of 303 SNPs were successfully typed with more than 50% success in all populations (An overall success rate of 90%) in 1047 individuals (Sample failure rate of 0.19%). The final mean spacing across the 21 genes was approximately 8.36kb; see Table 2 for number of SNPs and SNP density per gene. No particular SNP or population showed significant deviation from Hardy-Weinberg after correcting for multiple testing.

Search for Traces of Positive Selection

The MAF test may detect selective events such as a selective sweep acting on a gene and should cause a disproportionate number of SNPs with rare MAFs, while recent balancing selection would cause an excess of SNPs with high MAFs (Walsh et al. 2005). The DAF test may detect the signature of genetic hitchhiking as observation of an excess of derived alleles may be indicative of a genetic hitchhiking event (Fay and Wu 2000). The MAF and DAF tests for dopamine and serotonin genes were carried out and the results plotted with the experimental and simulated data from a large study on immune function related genes (Walsh et al. 2005) for comparison. See Figures 1-3 for graphs of the MAF tests and Figures 4-6 for the DAF tests.

In the MAF test, *HTR1A* was seen to show an excess of high frequency SNPs in many populations from all continents with the exception of SSAFR, although this gene contains just 6 typed SNPs. *MAOA* also showed this trend in some populations, and *HTR2C* showed a disproportionate number of rare MAFs in many populations from all the continents again this time with the exception of SSAFR, although

caution should be taken in drawing conclusions from the findings of this test in the case of X chromosome genes, as they are more susceptible to the influence of demographic effects such as genetic drift because of lower effective population size (Ne) (Akey et al. 2002). In the DAF test the *TH* and *PPP1R1B* genes showed a trend towards an excess of derived alleles although in both cases the number of SNPs typed was less than 10.

Population and regional differentiation at the allele frequency level as measured by the F_{ST} statistic can also be used to search for evidence of local selection. A high F_{ST} value, indicating elevated allele frequency differences between populations or continents can indicate selection acting locally. In general the F_{ST} values for this set of genes was low, as can be seen by the comparison with the other F_{ST} distributions as observed by a mean F_{ST} value for the dataset of 0.106 (Autosomal genes only), indicating low differentiation between populations (see Fig 7). A number of high individual marker F_{ST} values were noted however, in particular a contiguous stretch of elevated F_{ST} values in the *HTR2C* gene including a value of 0.41, mostly explained by differences between SSAFR and other continental groups. Also of note were the top values of the distribution, a value of 0.423 in *PPP1R1B*, and 2 values for *TH* (0.362 and 0.354) also within the top 5% of the distribution. In the pair wise parametric comparison of all six F_{ST} distributions described in the Materials and Methods section, the dopamine and serotonin gene F_{ST} distribution exhibited a significant difference with the ALFRED distribution and borderline significant differences when compared to the Glycosylation and the Marshfield distributions. For the non-parametric Krushkal-Wallis test significant differences between the six distributions taken together were observed and in the Mann-Whitney and Kolmogorov-Smirnov non-parametric tests for two samples, there were significant

differences between ALFRED and Glycosylation distributions and the dopamine serotonin gene F_{ST} distribution. These differences highlighted the overall lower genetic differentiation in the as measured by the F_{ST} statistic in the set of dopamine and serotonin genes.

The EHH test as implemented in the SWEEP program is based on the premise that a variant or SNP that is positively selected for will be found on an unusually long haplotype due it rising in frequency rapidly without the corresponding time frame needed for breakdown of LD in the surrounding area (Sabeti et al. 2002).The results from the SWEEP, EHH based analysis of the 11 gene regions showed the *HTR4* gene as having some high frequency (>30%) haplotypes, significant empirically for the EHH test in a number of populations (data not shown). Also the *SLC6A3* gene showed a high EHH score for a single core haplotype in two Asian populations. These two genes did not have signals of selection in any of the HapMap populations when looking using the iH score based Haplotter program. However three genes out of the rest of the twenty one genes in the study did have some signals of selection, namely *COMT* (in Yorubans), *SLC6A4* (in Europeans and Asians) and *TPH2* (in Europeans) (Voight et al. 2006). The differences in the findings from the two related algorithms could be due the different populations examined, none of the significant core haplotypes found in the SWEEP analysis were in any of the HapMap populations, or it could also be due to the different size range of the analysis, and lastly due to the different density of the SNPs in each case.

Discussion

Interest in discovering the genetic basis underlying individual differences in brain function related phenotype explaining distinct behaviour, cognition and psychiatric disorder susceptibility is at an all time high. The shaping of our genome has been heavily influenced by historical selective forces acting on it. Discovering which genes have undergone selection may hold the key to understanding both our uniqueness as a species, i.e. the 'human' genes, and the genes underlying the individual differences mentioned before. The search for traces of selection in genes related to brain function therefore, is of great interest, as it is in 'brain traits' that we show most differences from our nearest primate relatives, and in these traits also that our individual differences in cognition, personality, behaviour, social functioning lie.

The Dopamine and Serotonin systems are two of the most important neurotransmitter pathways in the human nervous system. Genes from both have become the focus of studies investigating individual susceptibility to behavioural and psychiatric disease. These neurotransmitter pathways control a number of important behavioural and emotional traits (Cravchik and Goldman 2000). Therefore nowhere is the search for evidence of selection of more interest than in genes from these two neurotransmitter pathways.

A number of the genes in the current data set have been previously reported as having undergone positive selection, *MAOA* (Andres et al. 2004), *SLC6A4* (Voight et al. 2006) and *DRD4* (Ding et al. 2002; Wang et al. 2004) both in the human lineage and population specific selection. We present a study where we investigated the evidence for traces of selection in this set of twenty one dopamine and serotonin genes in the set of worldwide populations constituting the HGDP.

Our findings indicate that the genes are generally more conserved across populations, with a lower between continental region differentiation based on allele frequencies as measured by the F_{ST} value, as seen by comparisons with five sets of genome wide F_{ST} distributions for distinct gene classes. In particular the dopamine and serotonin genes showed significant differences with a set of Glycosylation genes (Ferrer-Admetlla *et al* in preparation), a set genes related to infectious disease immunity which presumably would display more local adaptive selection. Some individual marker high values were noted, in particular a stretch of relatively high F_{ST} values in the *HTR2C* gene, although this gene being on the X chromosome and therefore having a lower N_e may have been more susceptible to the effects of genetic drift.

In the MAF and DAF tests, there was no strong outlier with regard to proportion of high frequency MAFs or rare MAFs when genes with less than ten typed SNPs were ruled out of the analysis, although the X chromosome gene *MAOA* and *HTR2C* may have shown a trend towards an excess of high frequency SNPs (an indicator of recent balancing selection) and rare SNPs (an indicator of a selective sweep) respectively. This finding taken together with the results of the F_{ST} suggest that *HTR2C* may be worth investigating further although caution should be taken as it is an X chromosome gene and therefore more susceptible to demographic affects other than selection. A trend towards an excess of derived alleles was observed in both *TH* and *PPP1R1B*, both of which also displayed high individual marker F_{ST} values and this therefore represents a finding of note and which may indicate some influence of selection on the genetic variation in these genes.

The findings from the Extended Haplotype homozygosity test did not find any evidence to suggest recent positive selection, especially when the two borderline results noted were not observed on visualisation of the same genes for the HapMap populations using the Haplotter program which implements the similar iH algorithm (Voight et al. 2006). No trace of selection was found in any of the genes of the current study in the worldwide populations for the genes reported as displaying signatures of selection in the Voight *et al* study (Voight et al. 2006).

In conclusion, we find some indicators that the shaping of the extant genetic variation in at least some of the genes from the set of dopamine and serotonin neurotransmitter pathway genes presented in this study is due to historical population based forces which may have included selective processes. The failure to find any evidence for selection using the EHH based tests may be due to the different timescale that the tests used in the present study have power to detect selection at (Sabeti et al. 2006). The results of the allele frequency based tests may indicate a selective event at a deeper timescale than that detected by the long range haplotype methods such as the EHH and the iH.

Acknowledgements

We thank Anna González-Neira in the Centro Nacional de Genotipación, Madrid for help with the SNP genotyping; and Howard Cann for his effort in providing the CEPH samples. We are grateful to Tomàs Marquès, Martin Sikora, Andrés Moreno and Anna Ferrer for suggestions and technical help. The present study was supported by the Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain (BFU2006-01235/BMC), Direcció General de Recerca, Generalitat de Catalunya (2005SGR/00608) and "Fundación Genoma España" (proyectos piloto CEGEN 2004-2005). M.G. received a fellowship (2004FI-IQUC2/0047) from the International Graduate School of Catalunya (Generalitat de Catalunya).

References

Aghajanian GK, Marek GJ (2000) Serotonin model of schizophrenia: emerging role of glutamate mechanisms. Brain Res Brain Res Rev 31: 302-12

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12: 1805-14

Andres AM, Soldevila M, Navarro A, Kidd KK, Oliva B, Bertranpetit J (2004) Positive selection in MAOA gene is human exclusive: determination of the putative amino acid change selected in the human lineage. Hum Genet 115: 377-86

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human genome diversity cell line panel. Science 296: 261-2

Cheung KH, Miller PL, Kidd JR, Kidd KK, Osier MV, Pakstis AJ (2000) ALFRED: a Web-accessible allele frequency database. Pac Symp Biocomput: 639-50

Cravchik A, Goldman D (2000) Neurochemical individuality: genetic diversity among human dopamine and serotonin receptors and transporters. Arch Gen Psychiatry 57: 1105-14

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, Moyzis RK (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc Natl Acad Sci U S A 99: 309-14

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155: 1405-13

Gonzalez-Neira A, Calafell F, Navarro A, Lao O, Cann H, Comas D, Bertranpetit J (2004) Geographic stratification of linkage disequilibrium: a worldwide population study in a region of chromosome 22. Hum Genomics 1: 399-409

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4: 45-61

Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques Suppl: 56-8, 60-1

Rosenberg NA (2006) Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70: 841-7

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298: 2381-5

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832-7

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive natural selection in the human lineage. Science 312: 1614-20

Schneider S, Roessli, D., Excoffier, L. (2000) Arlequin ver. 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978-89

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4: e72

Walsh EC, Sabeti P, Hutcheson HB, Fry B, Schaffner SF, de Bakker PI, Varilly P, Palma AA, Roy J, Cooper R, Winkler C, Zeng Y, de The G, Lander ES, O'Brien S, Altshuler D (2005) Searching for signals of evolutionary selection in 168 genes related to immune function. Hum Genet: 1-11

Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK (2004) The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. Am J Hum Genet 74: 931-44

Winterer G, Weinberger DR (2004) Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. Trends Neurosci 27: 683-90

Wise RA (2002) Brain reward circuitry: insights from unsensed incentives. Neuron 36: 229-40

Legends

Figure 1: The MAF test for i) Yri (Immune function gene study data data) v Yoruban and ii) Yri v SSAFR, Our data is represented by triangles, the experimental data from the immune function gene study data is shown by black circles and the simulated data is shown by grey dots.

Figure 2: The MAF test for i) Ceu (Immune function gene study data) v French, ii) Ceu v EUR and iii) Ceu v EUR, MENA and CSASIA.

Figure 3: The MAF test for i) Han (Immune function gene study data) v Han, ii) Han v EASIA and iii) Han v EASIA, AME and OCE.

Figure 4 The DAF test for i) Yri (Immune function gene study data data) v Yoruban and ii) Yri v SSAFR

Figure 5 The DAF test for i) Ceu (Immune function gene study data) v French, ii) Ceu v EUR and iii) Ceu v EUR, MENA and CSASIA.

Figure 6: The DAF test for i) Han (Immune function gene study data data) v Han, ii) Han v EASIA and iii) Han v EASIA, AME and OCE

Figure 7: Comparison of the 6 F_{ST} distributions, The F_{ST} distributions are '19 genes': a set of 19 High Divergence genes, 'ALFRED': the ALFRED database, 'GLYCO': 26 genes involved in Glycosylation, 'MARSH':the Marshfield Indel set, 'Chr 22': SNPs from a 'neutral' gene free region of chromosome 22.

Figure 1 (i) & (ii)

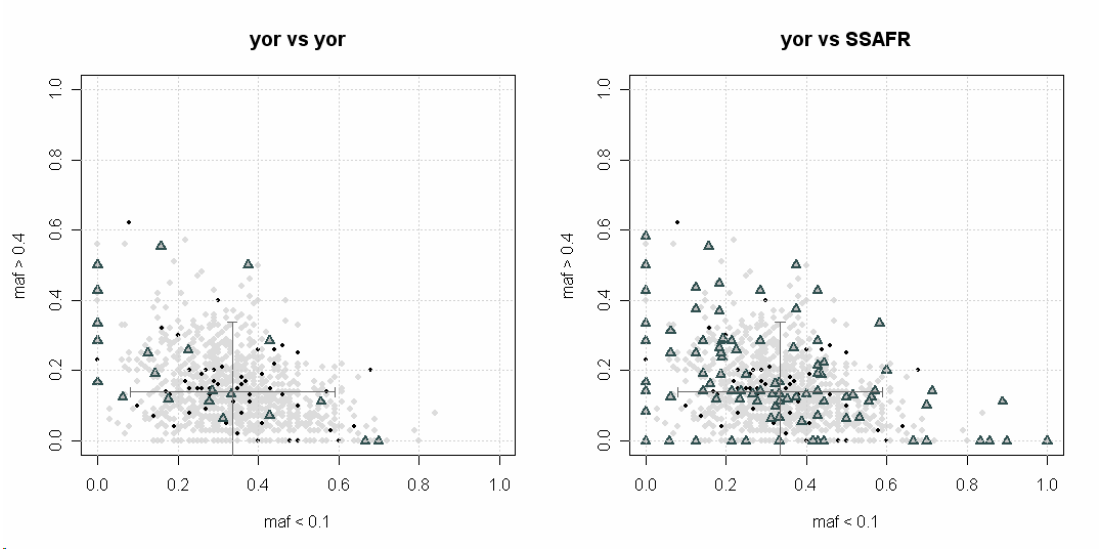


Figure 2 (i), (ii) & (iii)

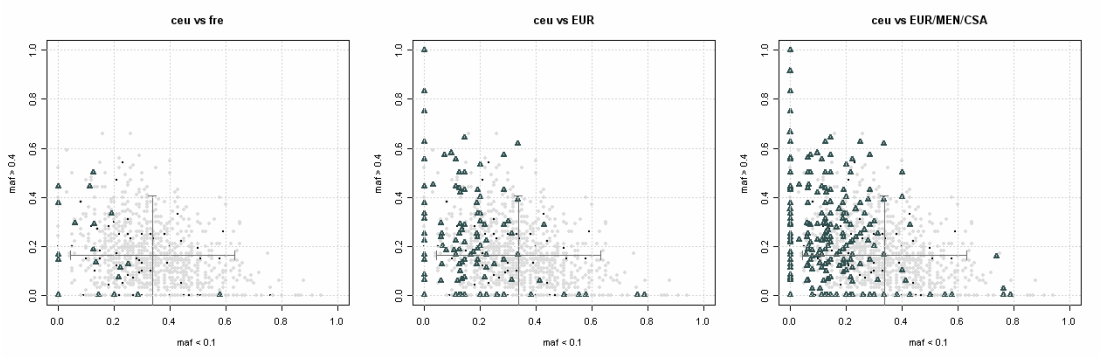


Figure 3 (i), (ii) & (iii)

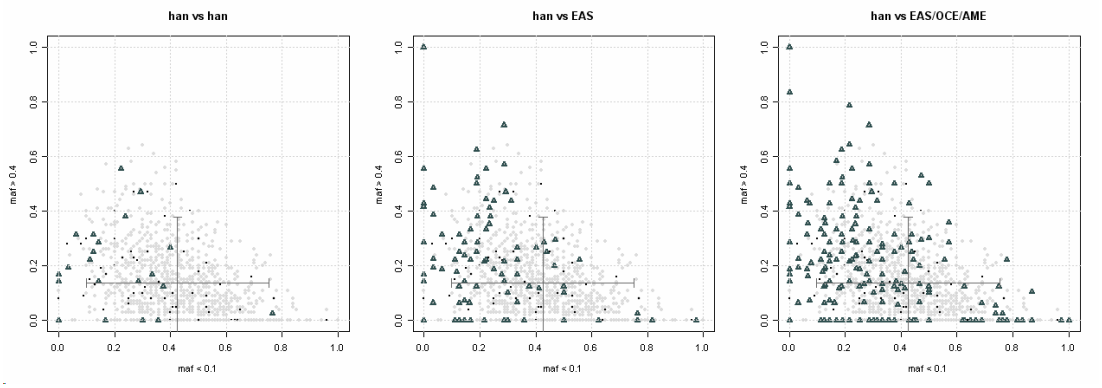


Figure 4 (i) & (ii)

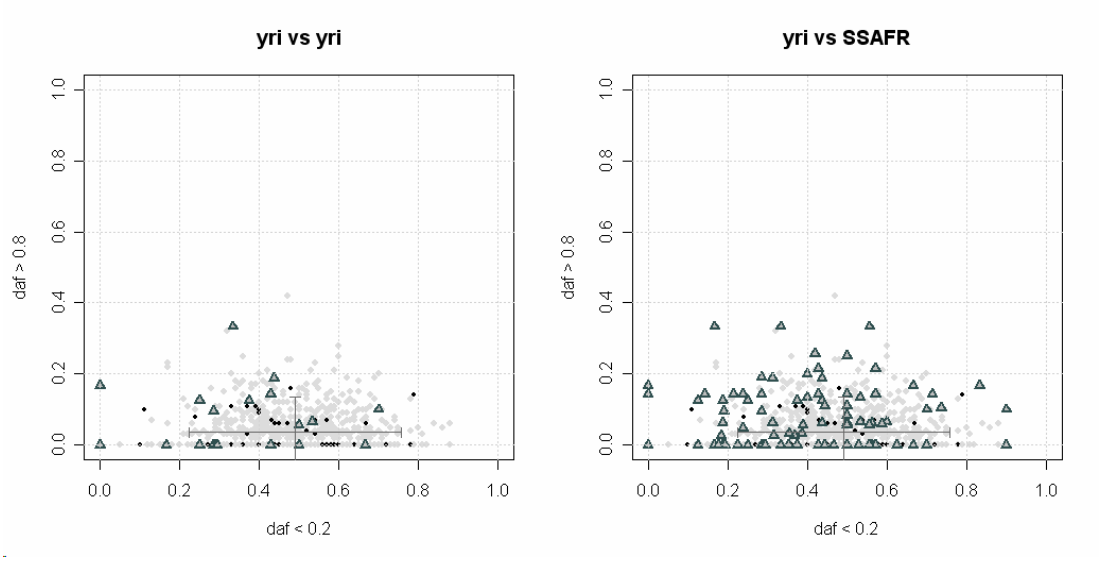


Figure 5 (i), (ii) & (iii)

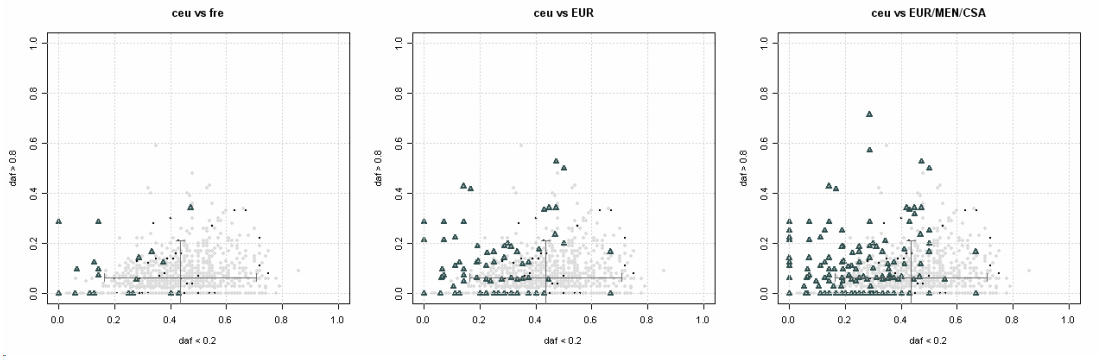


Figure 6 (i), (ii) & (iii)

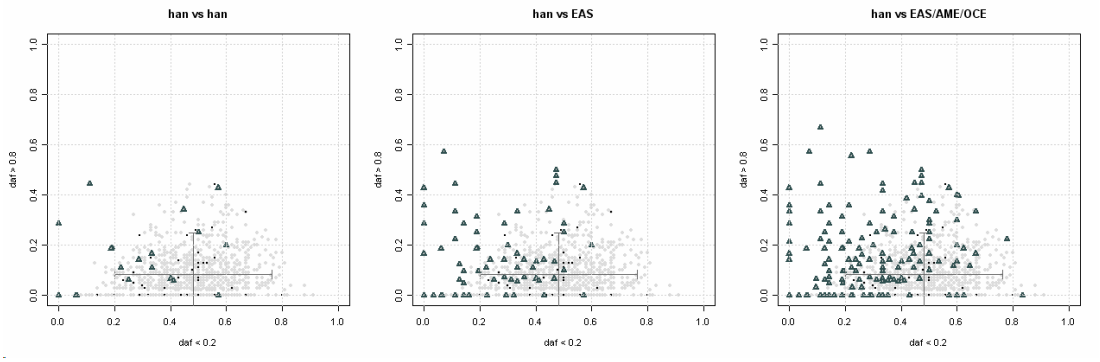


Figure 7

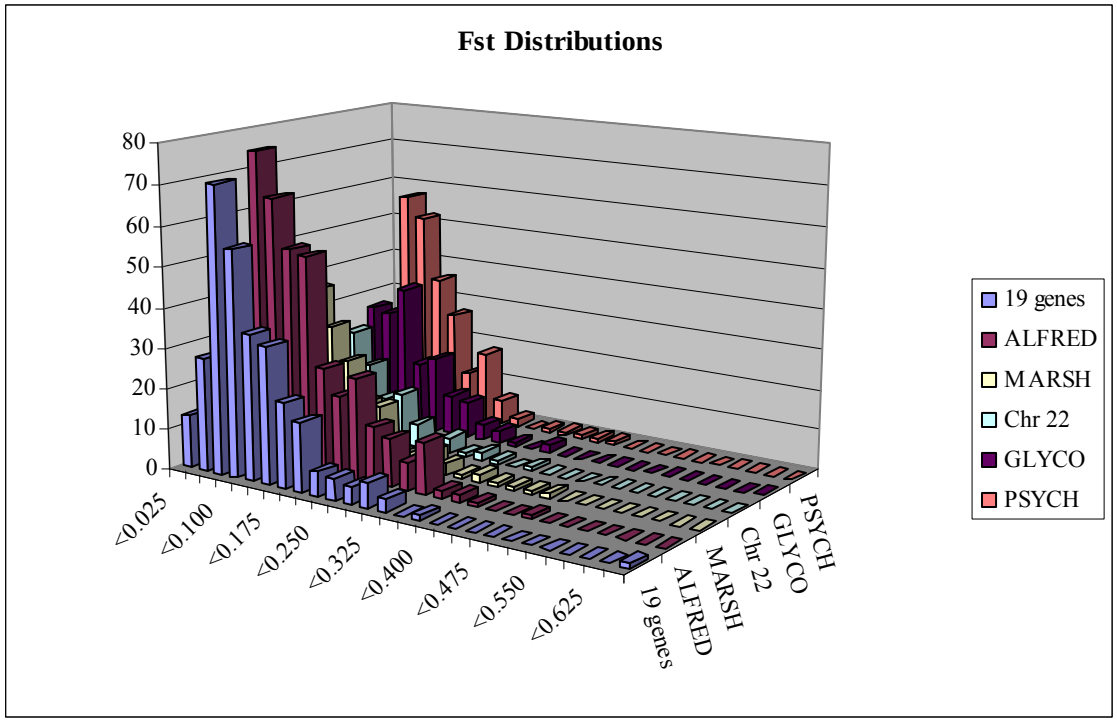

Fst Distributions
80
70
60
50
40
30
20
10
0
<0.025
<0.100
<0.175
<0.250
<0.325
<0.400
<0.475
<0.550
<0.625
19 genes
ALFRED
MARSH
Chr 22
GLYCO
PSYCH
19 genes
ALFRED
MARSH
Chr 22
GLYCO
PSYCH

Table 1

Continental Region	Population	N
America (AME)	Colombian	13
	Karitiana	24
	Maya	25
	Pima	25
	Surui	21
Central/South Asia (CSASIA)	Balochi	25
	Brahui	25
	Burusho	25
	Hazara	24
	Kalash	25
	Makrani	25
	North West China	29
	Pathan	25
	Sindhi	25
East Asia (EASIA)	Cambodian	11
	Han	44
	Japanese	29
	North East China	39
	South China	70
	Yakut	25
Europe (EUR)	Adygei	17
	Basque	24
	French	28
	North Italy	21
	Orcadian	16
	Russian	25
	Sardinian	28
Middle East/North Africa (MENA)	Bedouin	48
	Druze	47
	Mozabite	30
	Palestinian	51
Oceania (OCE)	NAN Melanesian	19
	Papuan	17
Sub Saharan Africa (SSAFR)	Bantu	20
	Biaka Pygmies	32
	Mandenka	24
	Mbuti Pygmies	15
	San	7
	Yoruba	24

Table 2

Gene	Pathway	Gene length (kb)	No. SNPs	SNP spacing (kb)
COMT	Dopamine	27254	16	4.89
DBH	Dopamine	22969	16	4.81
DDC	Dopamine	102615	21	7.74
DRD1	Dopamine	3126	8	8.40
DRD2	Dopamine	65575	16	7.40
DRD3	Dopamine	50199	14	8.00
DRD4	Dopamine	3398	6	8.36
DRD5	Dopamine	2029	7	8.09
HTR1A	Serotonin	1268	6	10.49
HTR1B	Serotonin	1259	10	17.25
HTR2A	Serotonin	62665	18	9.91
HTR2C	Serotonin	326073	39	6.14
HTR4	Serotonin	203037	32	7.23
MAOA	Dopamine	90601	12	7.86
MAOB	Dopamine	115835	10	9.91
PPP1R1B	Dopamine	9696	7	6.36
SLC6A3	Dopamine	52636	16	6.93
SLC6A4	Serotonin	37799	14	9.72
TH	Dopamine	7875	9	8.43
TPH1	Serotonin	21113	9	10.01
TPH2	Serotonin	93595	17	7.66

DISCUSSION

Genetic variation in Brain genes and implications for Association studies:

Behavioural variation in human beings encompasses a wide range of differences in personality and susceptibility to psychiatric illness arising from both genotype and environmental factors. The estimation of the contribution of a genetic component to human psychological traits varies widely between both the study and the trait measured but it averages at about 50% (Bouchard and McGue 2003; McGue and Bouchard 1998). This genetic component is likely to be attributable to genetic variation leading to functional variations in genes programming brain development and function (Cravchik and Goldman 2000).

Gaining an understanding of the genetic variation in genes involved in processes which exhibit human specific characteristics such as cognition, emotion and behaviour is obviously of great interest as it has the potential to give us a greater comprehension of the underlying biology of these processes. It may also go some way towards explaining the reasons behind the observed phenotype variation in individuals in traits such as behaviour, personality and distinct susceptibility to psychiatric disorders.

In addition as genes involved in pathways of brain function and development are obvious candidate genes for various psychiatric diseases and behavioural disorders describing the genetic variation of these genes has applications to disease gene mapping studies aimed at pinpointing the genes involved in these disorders.

Recent advances such as the availability of the sequence of the human genome sequence (Venter et al. 2001), a wealth of new SNP polymorphism data

(Kruglyak and Nickerson 2001), whole 'haplotype maps' of the human genome (Altshuler 2005) and huge improvements in SNP genotyping technology have led to a massive surge in studies aimed at determining the genetic basis of psychiatric disorders.

Most psychiatric disorders are complex diseases displaying a complex mode of transmission. It has been suggested that common genetic variants lie behind susceptibility to the common complex diseases such as psychiatric disease, the so-called 'common variant- common disease' hypothesis (Lohmueller et al. 2003). These common variants are found at relatively high frequencies (>1%) in global populations and individually only explain a small portion of genetic susceptibility but together with other susceptibility genes and environmental factors may lead to expression of the disease.

The method of choice in recent years in the search for disease genes implicated in complex disease has been association studies (Risch and Merikangas 1996). Most association studies are 'indirect', that is anonymous markers are tested for association with disease, relying on the extensive Linkage Disequilibrium (LD) found throughout the genome, so that the marker may be in LD with the causal variant.

Gene mapping studies have been marked by an inability to replicate results leading to increasing scepticism about the ability of current methods to identify the genes involved in complex diseases such as psychiatric disorders (Lohmueller et al. 2003; Owen et al. 2000; Risch 2000). The reasons underlying this failure is unknown but may include population based factors such as ethnic differences in allele frequencies, haplotype composition and LD (Colhoun et al. 2003; Gardner et al.

2006; Zondervan and Cardon 2004). As LD varies worldwide (Gonzalez-Neira et al. 2004) so to may the LD between the marker allele and the disease locus. Also as mismatching between disease and marker allele frequencies may result in a lack of statistical power to detect associations and as allele frequencies may vary worldwide this may be a factor (Zondervan and Cardon 2004)

We present two studies which address this issue, the first focussing just on the Neuregulin 1 gene (*NRG1*) and the other on the twenty one dopamine and serotonin pathway genes. *NRG1* is a gene which had been previously associated with Schizophrenia, initially in the homogeneous Icelandic population (Stefansson et al. 2002). This represented an exciting candidate gene for schizophrenia as it agreed with two of the common hypotheses for aetiology of the disease, the neurodevelopmetal theory (Rapoport et al. 2005) and the glutamate theory (Collier and Li 2003). Thus a flurry of publications ensued, some replicating the finding, (Corvin et al. 2004; Stefansson et al. 2003; Williams et al. 2003; Yang et al. 2003) while others refuted it (Ingason et al. 2006; Iwata et al. 2004; Tang et al. 2004; Thiselton et al. 2004; Venter et al. 2001).

We investigated whether some of the discrepancy might have had its basis in population differences by examining allele frequencies and distribution of the 'core haplotype' that was found associated with schizophrenia in the Icelandic population in a set of populations representing total worldwide diversity, the human genome diversity panel (HGDP), (Cann et al. 2002). We reported differences between continental groups in allele frequency and also frequency of the 'core haplotype' (Chapter 1). This finding in global 'unaffected' populations has implications for future association studies with this gene, but also in general reiterates the need to

strongly keep in mind the evolutionary history of the population where the association study is being carried out.

Due to the mixed results so far with regard to the association of *NRG1* with schizophrenia (as mentioned before a plethora of studies for and against association with this gene) any further studies in new populations are extremely valuable. A family based association study of *NRG1* with psychosis in a Spanish population was carried out (Chapter 2). This study included SNPs on the 'core haploype' found associated in Icelandic schizophrenia families as well as a number of SNPs spaced further along the gene. There was no statistically significant finding of association for any of the genes or haplotypes examined, however there was one borderline significant result for a missense mutation at the other end of the gene than that of the Icelandic finding of an associated haplotype. While the results of this study do not neccesarily point to population based motive based on allele or LD frequencies behind the failure to replicate the finding this cannot be ruled out either, as an association study an evolutionary similar population found similar results (Petryshen et al. 2004).

We also investigated the baseline genetic variation in the same worldwide population set (of in principal unaffected individuals) in the set of twenty one dopamine and serotonin genes (Chapter 4). Most of these genes have been investigated in relation to their role in psychiatric diseases, as both the serotonin and dopamine neurotransmitter pathways have been implicated in psychiatric disease pathophysiology from numerous functional studies involving the dopamine and serotonin pathways (Kapur and Remington 1996; Seeman et al. 1975; Seeman et al. 1976). There have been literally thousands of association studies and attempts at

replication, with genes from both pathways having been associated with a wide range of disorders from autism (Brookes et al. 2006) to schizophrenia (Palmatier et al. 2004), with just a few confirmed associations (Hirschhorn et al. 2002).

In the first place allele frequency differences and LD were examined. The analysis of the genetic variation in a large set of genes which have been implicated in various psychiatric diseases and behavioural disorders worldwide has provided a valuable guide for researchers investigating these genes for association with disease. Allele frequency differences were shown, with important differences in certain individual marker allele frequencies demonstrated (as measured by the F_{ST} value). On the whole however we showed that variation in allele frequencies is less than that of other genome-wide sets of genes, with the median F_{ST} value of the dopamine and serotonin gene distribution less than that of the other F_{ST} distributions. A multidimensional scaling plot of the values per population showed that populations within a continent clustered closely and that Sub Saharan African followed by American populations showed the most separation from other worldwide populations and the most intra-continental differences based on genetic distances. Linkage disequilibrium was also seen to vary in quantity across continental regions when all the continental groups were examined as a whole, this was mostly explained by the lower LD in Sub Saharan African populations. We also a finding of relative homogeneity of LD patterns across populations within the continental groupings. The results for allele frequencies and LD taken together indicate that factors other than population based factors need to be examined in the case of these twenty one genes as the reasons behind failure to replicate association studies.

The utilisation of tagSNPs in association studies takes advantage of LD to reduce the number of SNPs typed, allowing a reduction of cost and time. We carried out a tagSNP transferibility study for tagSNPs described in the HapMap populations for the genes to worldwide populations. In general results for serotonin and dopamine genes find in favour of the transferability of HapMap data to other population samples. Some indications for future associations studies can be derived from this work such as populations or genes requiring more or less SNPs to be typed.

Investigation of possible traces of Natural Selection in Brain Genes:

That human beings are highly different from our most closely related species, the chimpanzee and other great apes, with regard to cognition, intelligence, behaviour etc. is abundantly clear and that the difference in brain size and functioning underlying these distinctions is genetic is without doubt. There have been a number of studies aiming to discover the genes underlying these differences with more or less convincing findings, *ASPM* in brain size (Evans et al. 2004; Mekel-Bobrov et al. 2005; Mekel-Bobrov et al. 2007) and *FOXP2* in speech and language (Enard et al. 2002). Although there is still a long way to go to discover the 'human genes'. What has been less clear is that the differences between human individuals with regard to cognition, intelligence, behaviour and social functioning also has a genetic basis.

Natural Selection has been one of the greatest forces in the shaping of our genome (Sabeti et al. 2006), and the traces the different types of selective forces

leaves on the genetic variation can be tested for providing us with vital clues as to the evolutionary forces suffered by the region of the genome in question, and even in the case of local adaptive selection providing some indicators as to the environmental factors adapted to.

Local selection causes a characteristic higher than expected differentiation between populations based on allele frequencies, which can be measured by the F_{ST} statistic (Sabeti et al. 2006). The results of the first *NRG1* study (Chapter 1), showing allele frequency differences between worldwide populations revealed a finding of significant interest: two markers with extremely high F_{ST} values in the middle of the first intron of the gene and on a large block of LD. Comparing these individual marker values with the genome wide values available at that time suggested that these values were in the top five percent of the genomewide distribution of F_{ST} values. This finding was of sufficient interest to warrant a follow up study. We typed an additonal 32 SNPs around the block of LD in the full human genome diversity panel (HGDP) and aimed to determine if there was any evidence for selection having acted on the gene. Subsequent analysis could not find any evidence for selection having acted on this region. This finding urges a note of caution to researchers, as many claims of selection have been based on a finding of high F_{ST}. A more in depth analysis such as the one we have undertaken may well reveal that such values have another explanation such as genetic drift.

The search for traces of selection in genes related to brain function is of great interest as mentioned before, and nowhere more so than in the genes underlying two of the most important neurotransmitter pathways, dopamine and serotonin. These neurotransmitter pathways control a number of important behavioural and emotional

traits (Cravchik and Goldman 2000). A number of the genes in the current data set have been previously reported as having undergone positive selection, *MAOA* (Andres et al. 2004), *SLC6A4* (Voight et al. 2006) and *DRD4* (Ding et al. 2002; Wang et al. 2004) both in the human lineage and population specific selection. We undertook to search for any evidence of selection in the set of twenty one dopamine and serotonin genes in the set of worldwide populations constituting the HGDP.

In general we found less genetic differentiation based on allele frequencies and measured by the F_{ST} value in this set of 'Brain genes' than in other large gene sets for distinct gene classes, indicating that this set of dopamine and serotonin genes is more conserved across worldwide populations than expected. This was also reflected in the lower median F_{ST} value for the F_{ST} distribution. However a number of individual marker high F_{ST} values were noted in a few of the genes which may indicate local selection (such as adaptation to different environments). Other allele frequency based tests for signatures of selection such as selective sweeps, balancing selection and hitchhiking found indicators of the genetic variation in the same genes having been shaped by population history. It remains to be examined whether those population history forces may include selection. The results are intriguing, at once a finding of less diversity in 'brain genes' yet in that diversity seeing some hints of the forces of selection having acted.

It is clear from the genetic studies of recent years aimed at gaining an understanding of the human genome and its variation, the forces that have shaped it and what effects this genetic variation has on our susceptibility to disease that we are only beginning to comprehend the complexity of the human genome. This is evident from the difficulty in discovering the genetic basis for common diseases and the

mammoth task of teasing apart the signatures of selection from other demographic forces. Nevertheless the payoff for achieving just these tasks will be immense. A greater understanding of our species, particularly in the case of ‘brain genes’ where we may learn the genetic and therefore the biological basis of our differences from one another and from our nearest relatives from the primate order with respect to those ‘human traits’ of cognition, intelligence, behaviour, social function among others. We may discover that it is common variation that in some long gone environment provided an advantage and therefore persists in low frequency still, that now burdens us with disease. While these may seem like grandiose aspirations, they are in fact some of the most fundamental questions a species blessed or burdened with conciousness could ask, and while the road leading to this aspiration may be long and tenuous, it is also one of the most truly interesting journeys: towards self knowledge.

BIBLIOGRAPHY

Abdolmaleky HM, Thiagalingam S, Wilcox M (2005) Genetics and epigenetics in major psychiatric disorders: dilemmas, achievements, applications, and future scope. Am J Pharmacogenomics 5: 149-60

Adam D (2001) Genetics group targets disease markers in the human sequence. Nature 412: 105

Aghajanian GK, Marek GJ (2000) Serotonin model of schizophrenia: emerging role of glutamate mechanisms. Brain Res Brain Res Rev 31: 302-12

Aidoo M, Terlouw DJ, Kolczak MS, McElroy PD, ter Kuile FO, Kariuki S, Nahlen BL, Lal AA, Udhayakumar V (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet 359: 1311-2

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2: e286

Altshuler D, Brooks, L.D., Chakravarti, A., Collins, A.S., Daly, M.J., and Donnelly, P. (2005) A haplotype map of the human genome. Nature 437: 1299-320

Andreasen NC, Flaum M, Arndt S (1992) The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. Arch Gen Psychiatry 49: 615-23

Andres AM, Soldevila M, Navarro A, Kidd KK, Oliva B, Bertranpetit J (2004) Positive selection in MAOA gene is human exclusive: determination of the putative amino acid change selected in the human lineage. Hum Genet 115: 377-86

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3: 299-309

Backman L, Farde L (2001) Dopamine and cognitive functioning: brain imaging findings in Huntington's disease and normal aging. Scand J Psychol 42: 287-96

Barrett JC, Cardon LR (2006) Evaluating coverage of genome-wide association studies. Nat Genet 38: 659-62

Bergen AW, Yeager M, Welch RA, Haque K, Ganjei JK, van den Bree MB, Mazzanti C, Nardi I, Fichter MM, Halmi KA, Kaplan AS, Strober M, Treasure J, Woodside DB, Bulik CM, Bacanu SA, Devlin B, Berrettini WH, Goldman D, Kaye WH (2005) Association of Multiple DRD2

Polymorphisms with Anorexia Nervosa. Neuropsychopharmacology 30: 1703-10

Berry N, Jobanputra V, Pal H (2003) Molecular genetics of schizophrenia: a critical review. J Psychiatry Neurosci 28: 415-29

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74: 1111-20

Bertranpetit J, Calafell F, Comas D, Gonzalez-Neira A, Navarro A (2003) Structure of linkage disequilibrium in humans: genome factors and population stratification. Cold Spring Harb Symp Quant Biol 68: 79-88

Bertranpetit JaC, F. (2004) Genome Views on human evolution. Evolution: From Molecules to Ecosystems (A.Moya & E. Font Eds). Oxford University Press.: 260-271

Blaveri E, Kalsi G, Lawrence J, Quested D, Moorey H, Lamb G, Kohen D, Shiwach R, Chowdhury U, Curtis D, McQuillin A, Gramoustianou ES, Gurling HM (2001) Genetic association studies of schizophrenia using the 8p21-22 genes: prepronociceptin (PNOC), neuronal nicotinic cholinergic receptor alpha polypeptide 2 (CHRNA2) and arylamine N-acetyltransferase 1 (NAT1). Eur J Hum Genet 9: 469-72

Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl: 228-37

Bouchard TJ, Jr., McGue M (2003) Genetic and environmental influences on human psychological differences. J Neurobiol 54: 4-45

Brookes K, Xu X, Chen W, Zhou K, Neale B, Lowe N, Aneey R, Franke B, Gill M, Ebstein R, Buitelaar J, Sham P, Campbell D, Knight J, Andreou P, Altink M, Arnold R, Boer F, Buschgens C, Butler L, Christiansen H, Feldman L, Fleischman K, Fliers E, Howe-Forbes R, Goldfarb A, Heise A, Gabriels I, Korn-Lubetzki I, Marco R, Medad S, Minderaa R, Mulas F, Muller U, Mulligan A, Rabin K, Rommelse N, Sethna V, Sorohan J, Uebel H, Psychogiou L, Weeks A, Barrett R, Craig I, Banaschewski T, Sonuga-Barke E, Eisenberg J, Kuntsi J, Manor I, McGuffin P, Miranda A, Oades RD, Plomin R, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Taylor

E, Thompson M, Faraone SV, Asherson P, Johansson L (2006) 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: 934-53

Brune M (2004) Schizophrenia-an evolutionary enigma? Neurosci Biobehav Rev 28: 41-53

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human genome diversity cell line panel. Science 296: 261-2

Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106: 479-99

Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, Gottesman, II, Farmer AE, McGuffin P, Reveley AM, Murray RM (1999) Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Arch Gen Psychiatry 56: 162-8

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2: 91-9

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74: 106-20

Colhoun HM, McKeigue PM, Davey Smith G (2003) Problems of reporting genetic associations with complex outcomes. Lancet 361: 865-72

Collier DA, Li T (2003) The genetics of schizophrenia: glutamate not dopamine? Eur J Pharmacol 480: 177-84

Corfas G, Rosen KM, Aratake H, Krauss R, Fischbach GD (1995) Differential expression of ARIA isoforms in the rat brain. Neuron 14: 103-15

Corfas G, Roy K, Buxbaum JD (2004a) Neuregulin 1-erbB signaling and the molecular/cellular basis of schizophrenia. Nat Neurosci 7: 575-80

Corfas G, Velardez MO, Ko CP, Ratner N, Peles E (2004b) Mechanisms and roles of axon-Schwann cell interactions. J Neurosci 24: 9250-60

Corvin AP, Morris DW, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Clair DS, Waddington JL, Gill M (2004) Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. Mol Psychiatry 9: 208-13

Cravchik A, Goldman D (2000) Neurochemical individuality: genetic diversity among human dopamine and serotonin receptors and transporters. Arch Gen Psychiatry 57: 1105-14

Crawford DC, Akey DT, Nickerson DA (2005) The patterns of natural variation in human genes. Annu Rev Genomics Hum Genet 6: 287-312

Crawford DC, Carlson CS, Rieder MJ, Carrington DP, Yi Q, Smith JD, Eberle MA, Kruglyak L, Nickerson DA (2004) Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am J Hum Genet 74: 610-22

Crow TJ (2000) Schizophrenia as the price that homo sapiens pays for language: a resolution of the central paradox in the origin of the species. Brain Res Brain Res Rev 31: 118-29

Darwin C (1859) Origin of the species.

Darwin CW, A. R. (1858) Proceedings of Linnean Society of London 3

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, Moyzis RK (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc Natl Acad Sci U S A 99: 309-14

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418: 869-72

Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT (2004) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Hum Mol Genet 13: 489-94

Falls DL (2003) Neuregulins: functions, forms, and signaling strategies. Exp Cell Res 284: 14-30

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155: 1405-13

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296: 2225-9

Gardner M, Gonzalez-Neira A, Lao O, Calafell F, Bertranpetit J, Comas D (2006) Extreme population differences across Neuregulin 1 gene, with implications for association studies. Mol Psychiatry 11: 66-75

Gardner M, Williamson S, Casals F, Bosch E, Navarro A, Calafell F, Bertranpetit J, Comas D (2007) Extreme individual marker F(ST)values do not imply population-specific selection in humans: the NRG1 example. Hum Genet

Goldner EM, Hsu L, Waraich P, Somers JM (2002) Prevalence and incidence studies of schizophrenic disorders: a systematic review of the literature. Can J Psychiatry 47: 833-43

Goldstein DB (2001) Islands of linkage disequilibrium. Nat Genet 29: 109-11

Goldstein DB, Cavalleri GL (2005) Genomics: understanding human diversity. Nature 437: 1241-2

Goldstein DB, Weale ME (2001) Population genomics: linkage disequilibrium holds the key. Curr Biol 11: R576-9

Gonzalez-Neira A, Calafell F, Navarro A, Lao O, Cann H, Comas D, Bertranpetit J (2004) Geographic stratification of linkage disequilibrium: a worldwide population study in a region of chromosome 22. Hum Genomics 1: 399-409

Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J (2006) The portability of tagSNPs across populations: a worldwide survey. Genome Res 16: 323-30

Gould TD, Gottesman, II (2006) Psychiatric endophenotypes and the development of valid animal models. Genes Brain Behav 5: 113-9

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK (2007) Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. Eur J Hum Genet

Gutierrez B, Bertranpetit J, Guillamat R, Valles V, Arranz MJ, Kerwin R, Fananas L (1997) Association analysis of the catechol O-methyltransferase gene and bipolar affective disorder. Am J Psychiatry 154: 113-5

Halldorsson BV, Istrail S, De La Vega FM (2004) Optimal selection of SNP markers for disease association studies. Hum Hered 58: 190-202

Hamblin MT, Thompson EE, Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70: 369-83

Harrison PJ, Owen MJ (2003) Genes for schizophrenia? Recent findings and their pathophysiological implications. Lancet 361: 417-9

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLoS Biol 3: e193

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307: 1072-9

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4: 45-61

Hoenicka J, Aragues M, Ponce G, Rodriguez-Jimenez R, Jimenez-Arriero MA, Palomo T (2007) From dopaminergic genes to psychiatric disorders. Neurotox Res 11: 61-72

Howie BN, Carlson CS, Rieder MJ, Nickerson DA (2006) Efficient selection of tagging single-nucleotide polymorphisms in multiple populations. Hum Genet 120: 58-68

Ingason A, Soeby K, Timm S, Wang AG, Jakobsen KD, Fink-Jensen A, Hemmingsen R, Berg Rasmussen H, Werge T (2006) No significant association of the 5' end of neuregulin 1 and schizophrenia in a large Danish sample. Schizophr Res 83: 1-5

Ioannidis JP (2003) Genetic associations: false or true? Trends Mol Med 9: 135-8

Iwata N, Suzuki T, Ikeda M, Kitajima T, Yamanouchi Y, Inada T, Ozaki N (2004) No association with the neuregulin 1 haplotype to Japanese schizophrenia. Mol Psychiatry 9: 126-7

Jablensky A (1997) The 100-year epidemiology of schizophrenia. Schizophr Res 28: 111-25

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29: 233-7

Kang SJ, Scott WK, Li YJ, Hauser MA, van der Walt JM, Fujiwara K, Mayhew GM, West SG, Vance JM, Martin ER (2006) Family-based case-control study of MAOA and MAOB polymorphisms in Parkinson disease. Mov Disord 21: 2175-80

Kapur S, Remington G (1996) Serotonin-dopamine interaction and its relevance to schizophrenia. Am J Psychiatry 153: 466-76

Kauppi L, Jeffreys AJ, Keeney S (2004) Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 5: 413-24

Kendler KS, Gardner CO (1997) The risk for psychiatric disorders in relatives of schizophrenic and control probands: a comparison of three independent studies. Psychol Med 27: 411-9

Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Easter SM, Webb BT, Zhang J, Walsh D, Straub RE (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 153: 1534-40

Kendler KS, Myers JM, O'Neill FA, Martin R, Murphy B, MacLean CJ, Walsh D, Straub RE (2000) Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 157: 402-8

Kendler KS, Tsuang MT, Hays P (1987) Age at onset in schizophrenia. A familial perspective. Arch Gen Psychiatry 44: 881-90

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 624-6

Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-6

Law A (2003) Schizophrenia, IV, Neuregulin-1 in the Human Brain. Am J Psychiatry 160: 1392

Lawrence RW, Evans DM, Cardon LR (2005) Prospects and pitfalls in whole genome association studies. Philos Trans R Soc Lond B Biol Sci 360: 1589-95

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lonnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfson J, Sigmundsson T, Petursson H, Jazin E, Zoega T, Helgason T (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am J Hum Genet 73: 34-48

Li T, Stefansson H, Gudfinnsson E, Cai G, Liu X, Murray RM, Steinthorsdottir V, Januel D, Gudnadottir VG, Petursson H, Ingason A, Gulcher JR, Stefansson K, Collier DA (2004) Identification of a novel neuregulin 1 at-risk haplotype in Han schizophrenia Chinese patients, but no association with the Icelandic/Scottish risk haplotype. Mol Psychiatry 9: 698-704

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33: 177-82

Lohmueller KE, Wong LJ, Mauney MM, Jiang L, Felder RA, Jose PA, Williams SM (2006) Patterns of genetic variation in the hypertension candidate gene GRK4: ethnic variation and haplotype structure. Ann Hum Genet 70: 27-41

Mateu E, Calafell F, Lao O, Bonne-Tamir B, Kidd JR, Pakstis A, Kidd KK, Bertranpetit J (2001) Worldwide genetic analysis of the CFTR region. Am J Hum Genet 68: 103-17

Mayr E (1963) Animal Species and Evolution. Harvard University Press, Cambridge, Massachusetts

McGue M, Bouchard TJ, Jr. (1998) Genetic and environmental influences on human behavioral differences. Annu Rev Neurosci 21: 1-24

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT (2005) Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science 309: 1720-2

Mekel-Bobrov N, Posthuma D, Gilbert SL, Lind P, Gosso MF, Luciano M, Harris SE, Bates TC, Polderman TJ, Whalley LJ, Fox H, Starr JM, Evans PD, Montgomery GW, Fernandes C, Heutink P, Martin NG, Boomsma DI, Deary IJ, Wright MJ, de Geus EJ, Lahn BT (2007) The ongoing adaptive evolution of ASPM and Microcephalin is not explained by increased intelligence. Hum Mol Genet 16: 600-8

Merikangas KR, Risch N (2003) Will the genomics revolution revolutionize psychiatry? Am J Psychiatry 160: 625-35

Michailov GV, Sereda MW, Brinkmann BG, Fischer TM, Haug B, Birchmeier C, Role L, Lai C, Schwab MH, Nave KA (2004) Axonal neuregulin-1 regulates myelin sheath thickness. Science 304: 700-3

Nurminsky DI (2001) Genes in sweeping competition. Cell Mol Life Sci 58: 125-34

Owen MJ, Cardno AG, O'Donovan MC (2000) Psychiatric genetics: back to the future. Mol Psychiatry 5: 22-31

Palmatier MA, Pakstis AJ, Speed W, Paschou P, Goldman D, Odunsi A, Okonofua F, Kajuna S, Karoma N, Kungulilo S, Grigorenko E, Zhukova OV, Bonne-Tamir B, Lu RB, Parnas J, Kidd JR, DeMille MM, Kidd KK (2004) COMT haplotypes suggest P2 promoter region relevance for schizophrenia. Mol Psychiatry 9: 859-70

Payton A, Holmes J, Barrett JH, Hever T, Fitzpatrick H, Trumper AL, Harrington R, McGuffin P, O'Donovan M, Owen M, Ollier W, Worthington J, Thapar A (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: 464-70

Petryshen TL, Middleton FA, Kirby A, Aldinger KA, Purcell S, Tahl AR, Morley CP, McGann L, Gentile KL, Rockwell GN, Medeiros HM, Carvalho C, Macedo A, Dourado A, Valente J, Ferreira CP, Patterson NJ, Azevedo MH, Daly MJ, Pato CN, Pato MT, Sklar P (2004) Support for involvement of neuregulin 1 in schizophrenia pathophysiology. Mol Psychiatry

Polimeni J, Reiss JP (2003) Evolutionary perspectives on schizophrenia. Can J Psychiatry 48: 34-9

Pritchard JK, Cox NJ (2002) The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 11: 2417-23

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69: 1-14

Ramirez-Soriano A, Lao O, Soldevila M, Calafell F, Bertranpetit J, Comas D (2005) Haplotype tagging efficiency in worldwide populations in CTLA4 gene. Genes Immun 6: 646-57

Rapoport JL, Addington AM, Frangou S, Psych MR (2005) The neurodevelopmental model of schizophrenia: update 2005. Mol Psychiatry 10: 434-49

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411: 199-204

Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, Carracedo B, Gonzalez E, Barroso E, Fernandez LP, Yankilevich P, Robledo M, Carracedo A, Benitez J (2006) Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. Hum Genet 118: 669-79

Riley B, Kendler KS (2006) Molecular genetic studies of schizophrenia. Eur J Hum Genet 14: 669-80

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516-7

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405: 847-56

Rosenberg NA (2006) Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70: 841-7

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298: 2381-5

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ,

Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832-7

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive natural selection in the human lineage. Science 312: 1614-20

Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N, Cooper R, Reich D, Altshuler D, O'Brien S, Lander ES (2005) The case for selection at CCR5-Delta32. PLoS Biol 3: e378

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-33

Seeman P, Chau-Wong M, Tedesco J, Wong K (1975) Brain receptors for antipsychotic drugs and dopamine: direct binding assays. Proc Natl Acad Sci U S A 72: 4376-80

Seeman P, Lee T, Chau-Wong M, Wong K (1976) Antipsychotic drug doses and neuroleptic/dopamine receptors. Nature 261: 717-9

Seeman P, Tedesco JL, Lee T, Chau-Wong M, Muller P, Bowles J, Whitaker PM, McManus C, Tittler M, Weinreich P, Friend WC, Brown GM (1978) Dopamine receptors in the central nervous system. Fed Proc 37: 131-6

Siever LJ, Davis KL (2004) The pathophysiology of schizophrenia disorders: perspectives from the spectrum. Am J Psychiatry 161: 398-413

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506-16

Stead JD, Hurles ME, Jeffreys AJ (2003) Global haplotype diversity in the human insulin gene region. Genome Res 13: 2101-11

Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnarsdottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR, Stefansson K, St Clair D (2003) Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. Am J Hum Genet 72: 83-7

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, Hjaltason O, Birgisdottir B, Jonsson H, Gudnadottir VG, Gudmundsdottir E, Bjornsson A, Ingvarsson B, Ingason A, Sigfusson S, Hardardottir H, Harvey RP, Lai D, Zhou M, Brunner D, Mutel V, Gonzalo A, Lemke G, Sainz J, Johannesson G, Andresson T, Gudbjartsson D, Manolescu A, Frigge ML, Gurney ME, Kong A, Gulcher JR, Petursson H, Stefansson K (2002) Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71: 877-92

Stefansson H, Steinthorsdottir V, Thorgeirsson TE, Gulcher JR, Stefansson K (2004) Neuregulin 1 and schizophrenia. Ann Med 36: 62-71

Steinthorsdottir V, Stefansson H, Ghosh S, Birgisdottir B, Bjornsdottir S, Fasquel AC, Olafsson O, Stefansson K, Gulcher JR (2004) Multiple novel transcription initiation sites for NRG1. Gene 342: 97-105

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239: 1263-8

Tang JX, Chen WY, He G, Zhou J, Gu NF, Feng GY, He L (2004) Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population. Mol Psychiatry 9: 11-2

Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 9: 578-94

Thiselton DL, Webb BT, Neale BM, Ribble RC, O'Neill FA, Walsh D, Riley BP, Kendler KS (2004) No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). Mol Psychiatry 9: 777-83; image 729

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski

M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. Science 291: 1304-51

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4: e72

Wall JD, Pritchard JK (2003) Assessing the performance of the haplotype block model of linkage disequilibrium. Am J Hum Genet 73: 502-15

Walsh EC, Sabeti P, Hutcheson HB, Fry B, Schaffner SF, de Bakker PI, Varilly P, Palma AA, Roy J, Cooper R, Winkler C, Zeng Y, de The G, Lander ES, O'Brien S, Altshuler D (2005) Searching for signals of evolutionary selection in 168 genes related to immune function. Hum Genet: 1-11

Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK (2004) The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. Am J Hum Genet 74: 931-44

Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammit S, O'Donovan MC, Owen MJ (2003) Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. Mol Psychiatry 8: 485-7

Winterer G, Weinberger DR (2004) Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. Trends Neurosci 27: 683-90

Wright S (1931) Evolution in Mendelian Populations. Genetics 16: 97-159

Wright S (1938) The Distribution of Gene Frequencies Under Irreversible Mutation. Proc Natl Acad Sci U S A 24: 253-9

Wright S (1969) Evolution and the genetics of populations. Volume 2: The theory of gene frequencies. Chicago: University of Chicago Press; 511

Yang JZ, Si TM, Ruan Y, Ling YS, Han YH, Wang XL, Zhou M, Zhang HY, Kong QM, Liu C, Zhang DR, Yu YQ, Liu SZ, Ju GZ, Shu L, Ma DL, Zhang D (2003) Association study of neuregulin 1 gene with schizophrenia. Mol Psychiatry 8: 706-9

Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. Nat Rev Genet 5: 89-100