UNRAVELLING THE GENETICS OF SCHIZOPHRENIA AND ADHD

Steven C. Bakker

Unravelling the genetics of schizophrenia and ADHD

Unravelling the genetics of schizophrenia and ADHD

Op zoek naar de genetische basis van schizofrenie en ADHD

Met een samenvatting in het Nederlands

Proefschrift ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof.dr. W.H. Gispen,
ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op
31 mei 2005 des middags te 14:30 uur

door
Steven Cornelis Bakker
Geboren op 3 december 1970 te Woerden.

Promotoren: Prof.dr. P.L. Pearson, Faculteit der Geneeskunde, Universiteit Utrecht
Prof.dr. R.S. Kahn, Faculteit der Geneeskunde, Universiteit Utrecht
Copromotor: Dr. R.J. Sinke, Faculteit der Geneeskunde, Universiteit Utrecht

ISBN 90.393.3952.X

The research described in this thesis was performed at the Department of Biomedical Genetics and the Rudolf Magnus Institute of Neuroscience of the UMC Utrecht, as part of a UMC Genvlag project. ABC Neurogenomics, the Makaria Foundation and De Hersenstichting financially supported the work.

Cover: Michiel (aged 8 months) explores his mirror image. Alienation from one's own mirror image and impaired self-awareness are frequently encountered in schizophrenic patients (G.G. Gallup Jr. *et al*., in Kircher and David, *The Self in Neuroscience and Psychiatry*, Cambridge University press, 2003, pp 156-158). The back shows part of the genomic sequence of the neuregulin 1 gene (NRG1), with variations between human individuals. The polymorphic microsatellite D8S1810 is interrupted by single nucleotide polymorphism (SNP) rs2047850, and is flanked by SNP rs13260071.

Print: Febodruk BV, Enschede.

Publication of this thesis was financially supported by Eli Lilly BV, AstraZeneca BV,
Ingenieursbureau Floxair and Apotheek Varsseveld.

Voor Liesbeth en Michiel

Ter herinnering aan mijn moeder

Zelfs mentaal zijn wij veel onderhoriger aan natuurwetten dan wij denken, en de geest bezit al bij voorbaat, evengoed als die of die bedektbloeiende of een bepaalde grasachtige plant, de eigenaardigheden waar wij voor denken te hebben gekozen…
En het kan zijn dat terwijl je je ideeën als het resultaat ziet van een beredenering, en je lichaamskwalen als het gevolg van een nalatigheid in je hygiëne, je van je familie, net als vlinderbloemigen de vorm van hun zaad,
zowel de denkbeelden hebt waarop je teert als de ziekte waaraan je sterft.

Uit: M. Proust, A la recerche du temps perdu II (1918),
Nederlandse vertaling, De Bezige Bij, Amsterdam, vierde druk (2002), p.481

CONTENTS

SECTION I- GENETIC STUDIES IN SCHIZOPHRENIA

SECTION II GENETIC STUDIES IN ADHD

SECTION III- NEW GENOTYPING TECHNIQUES

SUMMARY

The susceptibility to common psychiatric disorders is largely determined by genetic factors. Knowledge of these factors may give insight into the causes of these disorders, and, more importantly, open new ways for prevention and treatment. Identifying the causal DNA variants, however, has proved to be difficult, despite considerable research efforts. One explanation is that many genes and environmental factors are involved, which would require large study populations to detect variants with modest effects. Some variants may be involved in specific disease symptoms only, or their relative importance could differ between population subgroups. In both cases, detecting such variants would be difficult in unselected, ethnically diverse patient samples.

This thesis describes genetic studies performed on clinically well-defined, homogeneous groups of Dutch patients suffering from schizophrenia and attention-deficit hyperactivity disorder (ADHD), and presents novel techniques for efficiently performing such studies.

In a large sample of unrelated schizophrenia patients, we investigated 12 functional candidate genes involved in dopamine neurotransmission. Dopamine dysregulation has long been implicated as a causal factor in schizophrenia, since most of the effective drug treatments block dopamine receptors in the brain. In this study, which was performed using microsatellite markers and pooled DNA from patients and controls, we found no evidence however for a significant contribution to schizophrenia from any of these genes. These results suggest that if dopamine genes are involved in schizophrenia, their role must be small.

Schizophrenia has a very diverse clinical presentation and disease course in individual patients, which suggests the existence of disease subtypes with different biological causes. To investigate this possibility, we collected patients with and without the deficit form of schizophrenia. The deficit syndrome is characterized by prominent, enduring negative symptoms, such as decreased interest, social withdrawal and flattened affect. In these groups, we investigated variants in several genes that have recently been implicated in schizophrenia. We found that the neuregulin 1 and RGS4 genes were almost entirely associated with the non-deficit group, which provided genetic support for a distinct aetiology of clinically recognizable forms of schizophrenia. The PIP5K2A gene, which has previously been implicated in bipolar disorder, was strongly associated with both deficit and non-deficit schizophrenia. For several other recently reported genes, such as dysbindin (DTNBP1) and G72/G30, we were able to rule out a substantial role in Dutch schizophrenic patients.

In ADHD, genes from the dopamine system have also been regarded as candidate genes, since the most effective medication blocks the dopamine transporter. We have investigated the dopamine transporter gene (DAT1), as well as the dopamine receptor genes 4 (DRD4) and 5 (DRD5), in a large family sample of Dutch children with ADHD. We found no evi-

dence for association of these genes with the disorder. In the same families with affected pairs of siblings, we have performed one of the first whole-genome linkage studies in ADHD. The results indicated that DNA regions on chromosomes 15q and 7p might contain genes that contribute to the disorder. On chromosome 7, we found no support for involvement of dopa decarboxylase (DDC), a functional candidate gene in the linkage region.

Besides specific genetic studies in these two disorders, two techniques are presented for the efficient screening of genetic markers. A new mathematical procedure was developed to correct the so-called stutter artefact in the analysis of microsatellite markers. This method enables the accurate and efficient estimation of microsatellite marker allele frequencies in DNA pools (combined DNA from large groups of individuals, which can be screened for differences between patients and controls). Finally, an inexpensive and efficient method was developed for analyzing single nucleotide polymorphisms (SNPs) on standard DNA sequencing equipment.

1 INTRODUCTION

Psychiatric diseases are a major cause of individual dysfunction and suffering. Many disorders have an early onset and persist throughout life, and since they occur with a relatively high frequency in the population, the burden on families and society is also large. Unfortunately, the options for prevention and treatment are still limited. Little is known about specific risk factors, and the current medication is usually not effective for all patients, while many drugs have substantial side effects.

Ignorance about the causes of psychiatric disorders hinders the development of specific preventive measures and medication. It has long been known that psychiatric illness runs in families, which could be explained by shared environmental risk factors, or by a heritable susceptibility. Studies in twins and adopted children have repeatedly indicated that the major psychiatric disorders are largely, although not completely, genetically determined [1]. The pattern of inheritance in families is not like that of "classical" monogenic genetic disorders, in which a single change, or mutation, in the heritable material is sufficient to cause disease. In addition, the risk for relatives of a patient to develop the same disorder is generally lower than would be expected if only one genetic factor were involved. Interplay of several genetic variants and environmental risk factors is therefore the most likely explanation for the occurrence of most psychiatric disorders. Like asthma, diabetes, and many other diseases, psychiatric disorders therefore belong to the so-called complex disorders.

Recently, the genetic causes of many monogenic disorders have been discovered. These successes have fuelled the hope that the same techniques could elucidate the causes of complex disorders.

1.1. General approaches for finding disease-related DNA variants in complex disorders

If heritable traits and disorders are transmitted through the DNA, the genetic susceptibility to disease must be the result of certain variants (specific forms of which are called alleles in genetic terms) somewhere in the four-letter DNA sequence. These differences between individuals may change the structure of proteins, which are produced using the information contained in the DNA, or the rate of production of a protein (gene expression). Proteins are essential for nearly all of the biological processes in the body, and a change in protein function may therefore lead to a disturbance in these processes, and ultimately to disease. Those parts of the DNA that contain the information for the production of proteins are the genes.

In a worldwide effort that took many years, the Human Genome Project and a private sequencing effort have recently revealed the nearly complete DNA sequence of a limited number of individuals [2, 3]. Using this information, thousands of genes, or predicted genes, have now been localized. Only a fraction of the entire DNA appears to consist of genes, while the function of the remaining DNA is largely unknown, but may involve regulation of gene expression (details are given in the text box).

All in our genes? By definition, disease-related variants in monogenic disorders have deleterious consequences, which usually result from changes that severely affect the structure of the protein. In complex disorders, on the other hand, the disease-related variants are not sufficient to cause disease by themselves. This could mean that other genes can normally compensate for the effects of a single deleterious variant, or the changes may have less serious effects. In addition to variants that change protein structure, changes may also involve regulatory sequences around otherwise intact genes, resulting in different amounts of protein being produced. Regulatory regions may be close to the gene, such as promoter regions or enhancer and silencer regions, but may also involve distant sequences, which regulate the expression of multiple genes. Of the few identified disease-related genes in complex disorders, a substantial number seems to involve changes in regulatory sequences [4]. Finally, the heritable differences between healthy individuals and patients may not always lie in the DNA sequence itself, but in so-called epigenetic factors such as DNA methylation [5]. This process renders DNA regions less available for transcription, thus reducing the function of genes contained in them.

It is technically not yet feasible to compare the entire DNA sequence in large numbers of patients and controls. Genetic investigations therefore have to reduce the work by focusing on DNA regions of special interest, or by reducing the number of variants to be studied [6].
Current genetic methods rely on the assumption that a certain disease-related variant arose at some point in history in a founder individual. This variant has been transmitted to later generations, and might still be present in a higher frequency in currently living individuals with a specific disease, than in control individuals. In order to be detectable, a variant should be present in a substantial number of patients. This could be the case, if specific variants in complex disorders play modest roles in causing disease, without being subject to strong negative selection (see below).

Common disease, common variant? The rapidly declining risk for more distant relatives of affected individuals indicates that at least several genetic variants are involved in most common psychiatric disorders, but the exact number of variants, or the number of different variants within single genes is unknown. The 'common disease-common variant' (CDCV) hypothesis predicts that only a limited number of predominating variants contribute substantially to disease. As was recently pointed out, this is a likely possibility if the current world population is the result of a rapid expansion from a small founder population, in which one or a few disease-related variants were already present at relatively high frequencies [4]. Such variants would be under limited or no selection pressure, and they would also be present at high frequencies in the unaffected population. A related model, which does not take into account population expansion, predicts that, in the absence of strong selection, some genes will by chance alone eventually contain disease-related variants with relatively high population frequencies [7]. Due to their high frequencies, however, such variants would contribute disproportionately to all disease cases at the population level. Genetic studies would detect these frequent variants, and the genes that contain them, in the first place [8, 9]. It is possible that, in addition to a few very frequent dominating alleles, many genes may also contain rare disease-related variants [7]. Moreover, many genes might not contain disease-related variants with high frequencies at all. The sparse empirical data in complex disorders suggest that common disease-related alleles are present in at least a number of common disorders studied to date [10]. Possibly, however, these findings represent the 'low-hanging fruit', and completely different approaches, such as extensive direct sequencing, may be required to detect additional low-frequency variants [11, 12].

Using genetic markers to detect disease-related variants. Not only the disease-related variant itself will be transmitted to affected children, but the surrounding DNA as well. Variants near a disease-related variant can therefore function as genetic markers, which indicate the presence of a true disease-related variant. Such markers are associated, or in linkage disequilibrium (LD) with the truly disease-related variant [13]. Any type of DNA variation that can be measured could be used as a genetic marker (see below). Currently, the most frequently used types of DNA variations are microsatellites and single nucleotide polymorphisms (SNPs).

Types of genetic markers, and methods for their analysis. Microsatellites, or short tandem repeats (STRs) are DNA sequences with short repeated motifs, mostly with a length of 2 to 5 nucleotides per repeated unit (e.g. CACACACA etc.), and with varying numbers of repeat units between individuals [14]. Although these repeats are abundant in the genome, their function is unknown. In rare cases, expanded repeats are known to influence gene expression, or to be disease-causing, most notably certain triplet repeat motifs [15], but in most instances the repeat number seems to be of little functional importance. Repeat numbers can be measured after making multiple copies of the repeat sequences with the polymerase chain reaction (PCR). Fragments with different repeat numbers will have a different length and electrical charge, and will migrate at different speeds through an electrically charged polymer gel. Recently, this electrophoresis process has been largely automated. Fragments are labelled with a fluorescent dye, which can be detected when it passes a laser beam. Fragments of many different individuals can be run in parallel through a gel, or arrays of gel-filled capillaries, and different markers can be run simultaneously by choosing different fragment sizes and dye colours, which results in high-throughput genotyping systems.

Single nucleotide polymorphisms (SNPs) are DNA variations at a single base pair position; for instance, some persons may have an A, where others have a G at the same position. SNPs are even more abundant in the genome than microsatellites. Many techniques for determining SNPs have been described [16, 17], and the choice will mainly depend on the genotyping needs and the budget of the investigator. Since SNPs generally consist of only two possible variants, they are generally less informative than microsatellites, which usually have many different repeat sizes, and are therefore said to be highly polymorphic. However, by combining information from several nearby SNPs, more informative marker haplotypes can be formed.

Linkage disequilibrium declines in time. With each new generation, the region of DNA around a disease-related variant that is shared with the founder will become smaller, due to a process called recombination. The entire DNA is present in two copies in each individual. One copy was obtained from the mother, the other from the father. During the formation of reproductive cells, the two linear DNA strands line up, random breaks are made at the same positions in both strands, and DNA fragments of equal length are exchanged between the two DNA copies. As a result of this recombination, the single DNA copy that is transmitted to a child is a unique blend of the two DNA copies of the parent. Only a limited number of recombinations take place with each generation, and the DNA regions shared between parents and children, or between siblings, are therefore large on average. With each following generation, new recombinations can take place at random positions around the disease-related variant. The DNA region around this variant that is still shared with the founder will thus become smaller in time.

Genetic association studies

Instead of investigating all possible DNA variation, genetic association studies reduce the workload by studying a limited number of genetic markers for association, or LD, with a disease-related variant [18]. Markers close to a disease-related variant will generally be in strong LD with it, while LD will decay with increasing distance between the marker and the disease-related variant, among other factors (see below).

Patterns of linkage disequilibrium in the genome. Besides the number of generations since two variants first occurred together, the distance between them, and the recombination rate in their genomic region will also determine the chance of a recombination [13]. Recent studies showed that fine-scale patterns of linkage disequilibrium (LD) are irregular and only partially explained by the distance between markers [11, 19]. LD reportedly occurs in a block-like pattern, in which regions of low recombination and high LD seem to be separated by recombination 'hot spots' [20, 21]. This concept of so-called haplotype blocks suggests that many SNPs within a block occur in a limited number of combinations (haplotypes). Most SNPs will then be redundant, and characterized by a limited number of 'haplotype-tagging' SNPs (htSNPs) within each block [22]. Data on LD patterns and haplotype blocks are now being collected in large collaborative efforts, such as the HapMap project, in the hope of greatly reducing the number of SNPs to be typed in a region of interest [23].

The amount of LD can be calculated in different ways. Frequently used measures are D' and r^2, which vary between zero (both variants are inherited completely independently) and one (complete LD, i.e. specific variants are always found together) [24].

The average distance at which LD between variants can still be detected will determine the density of genetic markers that is necessary to find a disease-related variant through association studies. The number of markers needed for a comprehensive genome screen has been suggested to be several hundreds of thousands in each individual. Although genotyping techniques are advancing rapidly, such studies are still barely feasible in large numbers of patients and controls. As a result, genetic association studies have so far been restricted to candidate genes or candidate DNA regions.

Candidate gene studies. Assuming that susceptibility to a disease is the result of variation in protein function, one could start with studying those parts of the DNA that code for proteins (genes), ignoring possible distant regulatory sequences [25]. A further reduction of research work is possible by focusing on genes with a function that is likely to be involved in the development of a disorder (functional candidate genes). For instance, it may be known that drugs that are used to effectively treat a disorder modulate the products of such genes. Alternatively, a gene or its product may have been found to be defective in animals with behavioural disturbances that resemble a psychiatric disorder [26, 27].

Functional candidate gene studies require a priori assumptions about the biological causes of a disorder. If unknown biological systems are involved, they are not likely to be identified.

Choosing controls. Case-control designs carry a risk of spurious findings due to a phenomenon called population stratification [28]. In the presence of population subgroups with different frequencies of marker alleles as well as different disease frequencies, an association of marker alleles with the disease may be noticed that is not caused by the presence of a disease-related variant. Several methods have recently been proposed to detect population subgroups, or to correct for the presence of base-line allele frequency differences between cases and controls [29, 30].

Another drawback of the use of unrelated patients and controls is that haplotypes must be estimated in the absence of phase-known haplotypes from parents. The transmission disequilibrium test (TDT) is a particular form of association study, in which the DNA from the parents instead of unrelated individuals is used as a control. This design protects against population stratification, since the patient and the controls will always share the same population subgroup. However, collecting DNA from parents can be difficult, especially in late-onset disorders, or in psychiatric disorders which result in a high frequency of disrupted families.

Linkage studies in affected sibling pairs

A limiting factor in whole-genome association studies are the enormous number of markers that would need to be genotyped, due to the limited LD between distantly related individuals. Marker numbers can be reduced, however, by studying families instead of individuals [31]. Like in association studies, the underlying assumption of linkage studies is that a disease-related variant was transmitted from a founder person to currently living patients. In a similar fashion, we expect that within one family, siblings with a disorder received the same disease-causing variant from a parent. However, within this family, the number of recombinations from parent to children will be low, causing affected siblings to share much larger DNA regions around the disease-related variant, than distantly related individuals. Consequently, only a small number of genetic markers (typically 300-800) will be sufficient to determine which DNA regions are shared by affected family members [32]. Siblings have a 50% chance of sharing any random part of the DNA. Large numbers of affected sib(-ling) pairs must therefore be studied to determine if certain DNA regions are shared by significantly more than 50% of sibling pairs. If so, this DNA region could contain a disease-related variant.

The limited number of markers required for affected sib pair studies makes linkage studies very efficient for screening the entire DNA for involvement of specific regions in a disorder. However, identified linkage regions will typically be very large, and contain tens or hundreds of candidate genes. Therefore, in a second phase, linkage regions are usually fine-mapped by association studies in unrelated individuals, using high-density marker sets, in order to pinpoint the disease-related gene.

1.2. Genetics of schizophrenia

Schizophrenia is a psychiatric disorder that generally presents itself in late adolescence or early adulthood, and affects approximately one percent of people worldwide [33]. Broadly, two classes of symptoms can be distinguished. Psychotic, or positive, symptoms include disturbances of perception (hallucinations) and delusions, while diminished interest and drive, disorganization of thoughts and speech, flattening of affect, and social withdrawal are regarded as negative symptoms. Positive or negative symptoms may predominate in individual patients, and the course of the disease is highly variable. Some patients have a relatively benign course with exacerbations and remissions, but in other patients, negative symptoms are persistent. Different criteria have been proposed to distinguish clinical schizophrenia subtypes, with the rationale that the differences in clinical presentation and disease course may reflect different underlying disease causes [34]. Recognition of schizophrenia subtypes might then be very valuable for predicting the disease course in an individual patient, and for choosing the appropriate therapy. Following the disease descriptions by Kraepelin, almost one century ago, a Kraepelinian form of schizophrenia was defined by the criterion that the patient had not been able to take care of himself for the last five years [35]. A more elaborate list of criteria has been proposed to delineate the deficit syndrome, a form of schizophrenia in which negative symptoms are prominent, primary, and enduring [36].

The deficit syndrome. Clinically, the 'deficit syndrome' was defined as a form of schizophrenia with predominantly negative symptoms and a chronic course with a poor prognosis. The syndrome is diagnosed using the Schedule for the Deficit syndrome (see page 158) [36, 37]. It is estimated that 15% of first-episode patients, and 25-30% of chronic patients will meet deficit criteria [38]. Several observations suggest that deficit schizophrenia may represent a distinct biological entity [38, 39]. For example, deficit patients were found to have higher blood levels of homovanilic acid (HVA), a breakdown product of dopamine, than other schizophrenia patients and eye-tracking abnormalities are especially prominent in deficit schizophrenia [40, 41]. Family studies indicate that the deficit syndrome aggregates in families and that it is significantly associated with a positive family history of schizophrenia [42]. The risk to develop schizophrenia is estimated to be 1.75 times greater for first-degree relatives of deficit patients than for relatives of non-deficit patients. In addition, there is a lower association of the deficit syndrome with a family history for other psychiatric disorders [43]. Together, these observations suggest a high genetic contribution to the deficit syndrome.

The underlying pathophysiology of schizophrenia has been proposed to be a deficit in sensory gating, meaning that persons with schizophrenia respond to extraneous stimuli as well as to internal thoughts that other people can ignore. The causes for the inability to discern relevant from irrelevant information, however, are unknown. No gross abnormalities of brain structure have been reported in schizophrenic patients.

Imaging studies have consistently shown a small reduction of brain volume, as well as an increased volume of the lateral ventricle in schizophrenic patients [44]. The loss of volume appears to be progressive during the course of illness [45]. Functional MRI studies, which measure local brain activity, suggest disruption of functional circuits, rather than local dysfunction of specific brain regions. Reported environmental risk factors include complications during pregnancy and birth, lower socio-economic status, living in urban areas, and migration [46, 47]. Together, however, environmental factors seem to play a minor role.

Twin and adoption studies have consistently shown that genetic factors account for approximately 80 percent of the susceptibility for developing schizophrenia, and that most likely several or multiple genes are involved [48]. Different types of studies have been used in attempts to identify these genes.

Functional candidate genes

Genes involved in dopaminergic and serotonergic neurotransmission have been investigated repeatedly, because antipsychotic medication is known to influence these systems. Results have been both positive, as well as negative, and the findings were not very consistent. Recent meta-analyses of all the available data have suggested small, but significant roles for several of these genes, e.g. the serotonin 2a (5HT2a) receptor [49, 50], the dopamine D2 [51] and D3 receptors [52-54], and the COMT gene [55].

Chromosomal abnormalities

In a large Scottish family, psychiatric disorders, including schizophrenia, are inherited together with a translocation (exchange of genetic material) between chromosomes 1 and 11 [56]. The breakpoint of this translocation disrupts two genes, which have been named DISC1 and DISC2, for Disrupted In SChizophrenia. This observation was regarded as a curiosity for some time, but recently the DISC1 gene was found to be associated with schizophrenia, schizoaffective disorder, and bipolar disorder in a sample of unrelated cases and controls [57]. Small deletions in the chromosome 22q11 region result in the Velo-Cardio-Facial Syndrome (VCFS), which is associated with significantly increased rates of psychosis and schizophrenia [58, 59]. This syndrome is quite rare, and in itself will not explain the majority of schizophrenia cases, but some of the genes contained in the region of deletion may play a role in other schizophrenia patients, or provide clues to relevant genes and pathways. Candidate genes in this region, for which association with schizophrenia has been reported, are COMT, PRODH2 and ZDHHC8 [60-62].

Linkage studies

Taken together, the known functional candidate genes explain only a fraction of the total genetic contribution to schizophrenia, indicating that other genes must be involved. In search of such genes with unknown functions, numerous schizophrenia linkage studies have been performed. Unfortunately, the evidence for involvement of certain genomic regions was not highly significant in most studies, and positive findings were frequently hard to replicate. Still, several genomic regions are supported by different linkage studies, some of which have significant evidence at the whole-genome level (on chromosome regions 1q, 6p. and 13q). Recent meta-analyses found the strongest combined evidence for linkage on chromosome regions 1p-q, 2p-q, 3p, 5q, 6p, 8p, 11q, 13q, 14pter, 20p and 22q [63, 64].

Genetic association studies in linkage regions

These findings have prompted association studies in regions of linkage, and several putative schizophrenia susceptibility genes have recently been identified [65]. Among these genes are neuregulin 1 (NRG1) on chromosome region 8p [66], dysbindin (DTNBP1) on chromosome region 6p [67], and G72/G30 on chromosome region 13q [68]. The regulator of G-protein signalling 4 (RGS4) gene is located in an area of linkage on chromosome region 1q21-22. Moreover, the expression of this gene was found to be downregulated in the brains of schizophrenic patients [69]. For each of these genes, replications have been reported, and although no causal variants have been detected yet, they are regarded as serious candidates. A recent report suggested that these genes, as well as a number of other putative schizophrenia susceptibility genes, might all be involved in glutamate neurotransmission [70]. Like the functional candidate genes, however, these positional candidate genes only marginally increase the risk to develop schizophrenia, typically with a factor of less than two, and together they still leave a large proportion of the total genetic risk for schizophrenia unexplained. This raises the question, whether the traditional genetic approaches are suitable for identifying the majority of heritable factors involved in this disorder.

1.3. Genetics of attention-deficit hyperactivity disorder (ADHD)

ADHD is a childhood-onset psychiatric condition characterized by excessive motor activity, impaired attention and impulsive behaviour [71]. It affects up to 5% percent of children in western societies [72], with boys being affected three times more frequently than girls. Symptoms persist into adulthood in a substantial number of patients (30 to 60 percent). ADHD is frequently accompanied by other psychiatric disorders, such as oppositional defiant disorder (ODD), substance abuse, anxiety and mood disorders, and language disorders [73].

We know very little about the causes of ADHD. Although gross structural brain abnormalities are absent, multiple studies have demonstrated reductions in cerebral volume in patients with ADHD, as well as healthy siblings [74]. Measurements of the electrical activity of the brain using EEG have repeatedly shown deficits in event-related potentials (ERP) and other changes, which may differ between diagnostic subgroups [75, 76]. Poor inhibitory control and deficits in fronto-striatal circuitry are regarded as core problems.

Family, twin, and adoption studies have demonstrated a large genetic contribution to the development of the disorder, with an estimated heritability of 80 percent [77]. Although some authors have suggested that one gene could predominate in causing the disorder [78], the large difference in concordance rates between monozygotic and dizygotic twins, and the relatively modest recurrence risk to first-degree relatives are more compatible with the involvement of multiple genes and environmental factors. Most genetic studies in ADHD have focused on functional candidate genes [79]. Genes involved in dopaminergic and noradrenergic neurotransmission have been extensively studied, since effective medication used to treat ADHD changes the function of these systems [80]. Consequently, other genes that may influence the levels of these neurotransmitters in the brain are interesting ADHD candidate genes. Recent meta-analyses of genetic studies have indicated that variants in the dopamine transporter gene (DAT1) and the dopamine receptor genes 4 and 5 (DRD4 and DRD5) are associated with a small increase in the risk of developing ADHD [81-83]. Results for many other investigated genes were not always consistent, however. Together, the functional candidate genes investigated so far only explain a minor proportion of the total genetic susceptibility to ADHD. This suggests that other, as yet unknown, genes or gene systems are involved in the development of the disorder. Linkage studies in ADHD, which could help to identify such genes, had not been performed when the research described in this thesis started in 1999.

1.4. DNA pooling

Large samples of patients will have to be analyzed for the detection of genetic variants with small effects, as in a typical complex disorder. DNA pooling is a way to reduce the workload of such studies by combining equal amounts of DNA from many individuals before the analysis [84, 85]. Compared with individual genotyping, which results in a signal that represents the two alleles (specific forms of variations) in each individual, the signal from a DNA pool will somehow represent the total spectrum of alleles that is present in the DNA pool. DNA pools of patients and controls can then be screened for different frequencies of variants. Both SNPs and microsatellites have been analyzed in DNA pools.

The problem of different signal intensities of alleles. Most SNP genotyping techniques produce pool signals in which the signal intensity of an allele does not correspond directly with the number of alleles present in the pool. Factors like differential amplification of alleles during PCR, the use of different fluorescent labels for both alleles, or different annealing properties of allele-specific probes can all cause such discrepancies. A systematic over- or under-estimation of allele frequencies may bias statistical tests, and possibly result in false-positive and/or false-negative findings [86]. Pool signals can be corrected for different signal intensities, after first measuring a number of heterozygous individuals (both alleles are present in equal amounts) [86]. This approach was reported to result in highly accurate estimates of allele frequencies. Especially in the case of infrequent alleles, however, finding sufficient heterozygous individuals may require a considerable amount of additional genotyping.

A major drawback of the analysis of SNPs in DNA pools is that it is impossible to construct multi-marker haplotypes, since specific combinations of SNPs in individuals cannot be determined. Haplotypes can be reconstructed, if pools of only a few individuals are used [87], which obviously greatly reduces the efficiency of pooling studies. In many situations, SNP haplotypes may have greater power to detect association than single SNPs [88, 89]. Multi-allelic microsatellite markers may be particularly suitable for analysis in DNA pools, since they carry the information comparable with a SNP haplotype in a single marker.

Compared with SNPs, however, the occurrence of PCR-induced stutter artefacts hinders deriving accurate allele frequencies from microsatellite genotype patterns. After PCR amplification and electrophoresis of a microsatellite, peaks are visible with a length of one to several repeat units less than the original microsatellite. When repeats of different lengths are present in a DNA pool, these so-called stutter peaks will contribute to neighbouring peaks, which makes it impossible to directly derive allele frequencies from the electrophoresis pattern (for a graphical example and a detailed introduction to stutter see page 81 onwards).

Some authors preferred to neglect stutter, with the argument that patient and control pools will suffer from stutter to the same extent, and they quantified differences in the total pool patterns between cases and controls. This approach does not allow the comparison of differences for single variants, or the combination of results from different experiments [90]. Moreover, biased estimates of the frequencies of different variants may lead to an under- or overestimation of differences between patient and control pools. A fundamentally different way to compare pool patterns is to correct pool patterns for stutter, in an attempt to obtain accurate estimates of allele frequencies from the signal, and subsequently analyze these as if they were results obtained from individual genotyping. All described methods for stutter correction in DNA pools require additional genotyping of individual DNA samples to measure the intensity of the stutter artefact, which can vary between markers. Additional individual genotyping will obviously reduce pooling efficacy, and some authors therefore suggested that stutter correction is not worth the effort. Uncertainty about the optimal way to handle the stutter artefact may explain why DNA pooling using microsatellites is not widespread. DNA pooling has been used, however, for genome-wide association studies in several complex genetic traits, in studies involving up to 6000 microsatellites (see below).

The analysis of microsatellite patterns obtained from DNA pools.

Methods without stutter correction. Visual comparison of banding patterns from amplified DNA pools on electrophoresis gels resulted in the identification of obesity loci in mice [91]. In humans, several markers associated with Tourette syndrome were found using the same approach [92]. Daniels *et al*. obtained pool signals from fluorescently labelled PCR products [93]. Differences between patterns were detected by overlaying the pool patterns of cases and controls and comparing the shared area under the curves (ΔAIP method). The authors were able to replicate a previously found association of two markers with hemochromatosis. This method was also used to search for loci associated with high cognitive ability in children [94, 95]. The shared area under the curve could be influenced by differences in running conditions between the pools and by the scaling of pool patterns to their highest peaks. This might explain the frequent false-positive findings that were reported. Collins *et al*. found peak height to be a more reliable measure than area under the curve [96]. They proposed a method that compares the sum of the absolute differences in peak height of all peaks in the pool patterns (ΔTAC method). Although the true frequencies could not be derived from the pools, the correlation between true allele frequencies in pools and ΔTAC was high, and considerably better than for the ΔAIP statistic. Determining the significance levels of differences between pool patterns is difficult. Both the ΔAIP and the ΔTAC method require simulations of large series of pool patterns, which are based on a very simplified general model that assumes equal stutter for all markers. Instead of comparing a measure of the total difference for all alleles, peak heights from pool patterns could also be analyzed as if they represented true allele frequencies. This method has been applied to a whole-genome screen with nearly 6000 polymorphic markers in multiple sclerosis [97]. Several loci were identified that had previously been suggested by linkage studies. As was pointed out, PCR artefacts like stutter are likely to distort the estimates and have a negative effect on the power to detect differences [98].

Methods with stutter correction. Daniels *et al*. obtained the relative area under the first and second stutter peaks from several homozygous individual genotypes and subtracted these stepwise from the peaks in the pool patterns, starting with the longest allele. This method, which did not take into account allele-specific stutter, did not produce sufficiently accurate allele estimates [93]. Perlin *et al*. constructed a more sophisticated allele-specific correction matrix for stutter and used it in a deconvolution algorithm, which resulted in highly accurate correction of individual genotypes [99]. This method was applied to pool patterns by analyzing one dinucleotide marker associated with hemochromatosis [100]. Twenty individuals were genotyped to obtain correction values for stutter and differential amplification (DA). After correction, estimated frequencies closely resembled true frequencies and the previously reported association with the marker could be replicated (although it was pointed out that this was a relatively strong association, which was also apparent from uncorrected comparison of the pools) [100]. Kirov *et al*. combined elements from different methods [90]. Ten individual genotypes were used to construct a correction model for stutter and to determine differential amplification, assuming a linear increase with allele length for both artefacts. Peak heights were corrected for stutter by stepwise subtraction, as first applied by Daniels *et al*., followed by correction for differential amplification. In several dinucleotide repeat markers correction resulted in allele frequencies that approached true frequencies. Moreover, by adding alleles to pools it was shown that correction resulted in the detection of differences that were significant in individual genotyping, whereas both the ΔAIP method and uncorrected comparison underestimated the differences [90]. Several groups reported tetranucleotide markers to more or less reflect true allele frequencies without any stutter correction [90, 100, 101].

1.5. Sample heterogeneity in genetic studies

Results of genetic studies in complex disorders have generally been difficult to replicate. The effect of single variants appears to be small, which can be explained in several ways. First, specific disease-related variants may be present in all patients, with only a small contribution to the development of the disease. In this situation, very large patient samples would be needed to detect these variants. It is also possible that the relative contribution of specific variants differs between populations, or between groups of patients with specific characteristics. In unselected, heterogeneous patient samples, single variants will then contribute mainly to small subgroups of patients, and stratification of patient samples according to ethnic background or specific characteristics could be expected to facilitate their detection.

Reducing ethnic heterogeneity

Some patients may belong to a particular ethnic subgroup, in which a certain disease-related variant is present in higher frequencies than in other populations. The occurrence of a population bottleneck in history, with a small number of founder individuals, would increase the genetic homogeneity of this subpopulation. The study of geographically or socially isolated populations has therefore been suggested as a way of increasing the power of genetic studies [102-106]. Moreover, LD in young isolates may be more extensive than in the general population, which would allow the use of marker sets with a lower density in association studies [107-109].

Studying isolates may be an efficient way to identify genetic variants involved in complex disorders, but results from such studies may be specific for the isolated population under study, and not necessarily reflect the most important genetic risk factors in the general population. The majority of genetic studies have been performed in ethnically unselected samples, and there is little actual data on how genetic heterogeneity in the general population could influence such studies. Recent studies suggest that extensive heterogeneity is present across ethnic subgroups [110], and that even in populations considered to be relatively homogeneous, like the Icelandic population, genetic subgroups can be identified [111]. Genetic heterogeneity may therefore be a significant problem in the study of complex disorders.

Reducing clinical heterogeneity

In the absence of reliable biological disease markers, disease classification in psychiatry is entirely based on clinical criteria, such as those laid down in the Diagnostical and Statistical Manual (DSM-IV) [71]. However, in patients assumed to suffer from the same disorder, there may in fact be partially different underlying causes. Alternatively, psychiatric symptoms that are observed in distinct disorders may share common disease mechanisms. This is, for example, suggested by the familial co-occurrence of ADHD and autism [112].

The study of endophenotypes has therefore been proposed as another way to increase the genetic homogeneity of study samples [113, 114]. An endophenotype can be any measurable specific characteristic (trait) that is associated with the disorder, and which can be present in patients, as well as in relatives that do not meet all the criteria for psychiatric disease. This could be a specific behavioural or cognitive deficit, such as short-term memory, or more basal physiological disturbances, such as eye-tracking patterns, or electrical activity in the brain in reaction to stimuli. An endophenotype that is part of the entire, clinically recognized disease syndrome might involve a smaller number of genetic variants, which could therefore be more easily detectable.

1.6. Outline of this thesis

The work presented in this thesis had three aims:

- The identification of genes or chromosomal regions involved in the development of schizophrenia and ADHD;
- Investigating genetic differences between schizophrenia patients with and without prominent negative symptoms;
- The development of techniques for efficient marker analysis in genetic association studies.

Section I: Genetic studies in schizophrenia

When the work described in this thesis started in 1999, many different schizophrenia linkage studies had been published with apparently inconsistent results. Likewise, results of studies of functional candidate genes were difficult to interpret, and hard to replicate.

Hypothesizing that part of the problems in the genetic study of complex disorders could be the result of sample heterogeneity, we collected a large, ethnically homogeneous sample of unrelated schizophrenia patients for association studies. Special efforts were made to include a large number of patients with deficit schizophrenia, to allow the study of genetic variants in patient subgroups with different clinical characteristics.

Chapter two describes a systematic screening using DNA pooling of schizophrenia functional candidate genes involved in dopamine neurotransmission. Chapter three presents the analysis of the neuregulin 1 gene in patients with different clinical characteristics (deficit and non-deficit schizophrenia), while chapter four describes a similar study of the dysbindin, G72/G30, RGS4 and PIP5K2A genes.

Section II: Genetic studies in ADHD

The starting point for the genetic studies in ADHD was a multitude of published studies in functional candidate genes with inconsistent results, while no whole-genome linkage studies had been performed.

We collected an ethnically homogeneous sample of families with multiple children affected with ADHD for linkage and association studies. In chapter five, markers in the dopamine transporter gene and the dopamine receptor genes 4 and 5 were investigated for association with ADHD. A whole-genome scan in 164 affected sibling pairs with ADHD is presented in chapter six. These results are followed up in chapter seven, which describes an association analysis of the dopamine decarboxylase gene (DDC), located in a linkage peak on chromosome 7.

Section III: New genotyping techniques

Genetic studies in complex disorders are limited by the capacity of genotyping techniques.

Large-scale application of DNA pooling using microsatellites appeared to be hindered by PCR-induced artefacts, and uncertainty about ways to handle these. In chapter eight, the characteristics of PCR-induced stutter artefacts in microsatellite genotyping are investigated in detail. These results are worked out in chapter nine, in which a novel method was developed for correcting the stutter artefact in DNA pools.

Single nucleotide polymorphisms (SNPs) are becoming increasingly important as genetic markers, but methods for their analysis are usually either time-consuming, or they require considerable investments in genotyping equipment. In chapter ten, we describe the development of an efficient method for SNP genotyping on standard DNA sequencing equipment, which is present in most genetic laboratories.

Finally, in chapter eleven the results are discussed and compared with other studies, followed by some speculations about future developments in genetic research into psychiatric disorders.

SECTION I

GENETIC STUDIES IN SCHIZOPHRENIA

2 NO ASSOCIATION BETWEEN 12 DOPAMINERGIC GENES AND SCHIZOPHRENIA IN A LARGE DUTCH SAMPLE

M.L.C. Hoogendoorn, S.C. Bakker, H.G. Schnack, J-P.C. Selten, H.G. Otten, W. Verduijn, F.M.M.A. van der Heijden, P.L. Pearson, R.S. Kahn and R.J. Sinke

Abstract

It has been suggested that genes involved in dopamine neurotransmission contribute to the pathogenesis of schizophrenia. However, reported associations of the disorder with genetic markers in dopaminergic genes have yielded inconsistent results. Possible explanations are differences in phenotyping, genetic heterogeneity, low marker informativity, and the use of small sample sizes. Here, we present a two-stage analysis of twelve dopaminergic genes in a large sample of Dutch schizophrenic patients. To reduce genetic heterogeneity, only patients with at least three Caucasian grandparents of Dutch ancestry were ascertained. An efficient genotyping strategy was used, in which polymorphic microsatellite markers were first screened for association in DNA pools. Promising results were followed up by individual genotyping in an extended sample. The pooled samples consisted of 208 schizophrenic patients and 288 unmatched control individuals. For each of the genes, more than one microsatellite marker was selected where possible, either intragenic or close to the gene. After correcting for multiple testing, significantly different allele frequencies were detected for DRD5 marker D4S615. Subsequently, we individually genotyped this particular marker and another DRD5 marker, as well as a DRD3 marker that could not be analyzed using the pooling strategy. This was done in an extended sample of 282 schizophrenic patients and a control sample of 585 individuals. In this second stage of the study, we found no association between these three markers and schizophrenia. The results of our comprehensive analysis provide no evidence for association between schizophrenia and 12 dopaminergic genes in a large Dutch sample.

Introduction

Multiple data from pharmacological and neuroimaging studies suggest that dopaminergic neurotransmission is dysregulated in schizophrenia. Since genetic factors contribute about 80% to the aetiology of this complex disorder [48], dopaminergic genes can be considered as schizophrenia candidate genes. However, previous studies on genes implicated in dopaminergic neurotransmission were inconclusive [65], although meta-analyses of dopamine D2 and D3 receptor polymorphisms have suggested small but significant effects on the susceptibility to schizophrenia [53, 115]. There may be several explanations for the lack of conclusive results, e.g. clinical heterogeneity caused by the use of different phenotypes or phenotyping instruments, genetic heterogeneity, lack of power due to small sample sizes, and low marker informativity. Most previous association studies of schizophrenia and the dopamine system examined only one or a few dopaminergic genes, and they most often tested for association with only one or few specific polymorphisms. The purpose of the current study was to perform a comprehensive candidate gene analysis between 12 genes involved in dopaminergic transmission and schizophrenia in a sample of Dutch patients. Our aim was to identify any possible association for any of these genes, rather than to restrict our analyses to previously reported singular, associated functional variants. We therefore set out to screen for association using microsatellite polymorphisms, as these markers are multi-allelic and thus highly informative.

The study was conducted in two stages. Firstly, a DNA pooling technique was used for a fast screening of microsatellite markers within and around the 12 genes. A major advantage of pooling is that it greatly reduces the amount of genotyping by combining patient and control samples separately before the analysis [85, 90, 93, 116]. Secondly, if any of the markers showed significant differences in allele frequencies between the groups of pooled cases and pooled controls, individual genotyping was performed in an extended case sample and an independent control sample.

Materials and methods

Patients. The sample was ascertained through mental health services and relatives' support groups in the Netherlands. We diagnosed 208 unrelated patients with a primary diagnosis of schizophrenia (DSM-IV) using the Comprehensive Assessment of Symptoms and History (CASH, [117]) and additional information from medical records and clinicians. All patients were diagnosed by well-trained raters. To reduce genetic heterogeneity, only patients with at least three Caucasian Dutch-born grandparents were included. All participating patients provided written informed consent and the Medical Research Ethics Board of each participating institution approved the project. For replication of pooling results using individual genotypes, the sample was extended with 74 patients to a total of 282 patients. The inclusion criteria for these 74 patients were the same as for those who were ascertained for the original sample.

Controls. For the DNA pooling study, a total of 288 samples were obtained from 179 anonymous blood bank donors, and 109 control individuals from our department. For replication of pooling results by individual genotyping, a largely independent sample of 585 unmatched controls was assembled from 472 different anonymous blood bank donors, and 113 controls from our department. All subjects gave permission for their blood to be used for research purposes. Information about descent was not available from the control subjects.

Genes of interest. Twelve genes were studied (Table 2.1), three of which are involved in the synthesis of dopamine: Tyrosine Hydroxylase (TH), Dopa Decarboxylase (DDC), and Dopa-ß-hydroxylase (DBH). Five of the genes are coding for the dopamine receptors: DRD1, DRD2, DRD3, DRD4, and DRD5. One gene codes for the dopamine transporter: DAT1, and lastly, three genes are involved in the degradation of dopamine: Monoamine oxidase A (MAOA), Monoamine oxidase B (MAOB), and Catechol-O-Methyltransferase (COMT).

Table 2.1. Microsatellite marker information

Gene	Marker	Repeat[a]	Distance[b]	Primers
Tyrosine Hydroxylase (TH)	TH01	Tetra	Intragenic	GTGGGCTGAAAAGCTCCCGATTAT ATTCAAGGGTATCTGGGCTCTGG
Dopa Decarboxylase (DDC)	DDC_Intra	Di	Intragenic	TAGCTTATTGCTAGGATATTAGG CTTTCCCAGCTATCTCTCTC
Dopa-beta-Hydroxylase (DBH)	DH59AC	Di	6 kb down	GCAGTCACGCATCCTTATGG CAGCTCTGGGCTCATGCTC
	DBH_Up	Di	31 kb up	AGACAGACACCCTACCTCAC GTGTCTTGTTTTCAGGGAAGTT
Dopamine Transporter 1 (DAT1)	D5S2005	Di	52 kb up	CCTCAGGTGGGTTATTGAC CCCAGGGCTTTACGAGT
Dopamine Receptor 1 (DRD1)	DRD1_Dwn55kb	Di	55 kb down	GTGTCTTACAACAATTTGGGAAGAGAT CCATTTTCAGTAGCAACAAA
	DRD1_Dwn57kb	Tetra	57 kb down	GTGTCTTCTAGCACCAAGGTCAAAT GGGCTTACATGTCTAGGTGA
Dopamine Receptor 2 (DRD2)	D11S3179	Di	166 kb down	CTTCCTACCAAAGGGGC ATCAATCCATCAGTGGGG
	RH27315	Di	Intragenic	GGAGGGCGGTGCGGTCAT CAGGAGCACGTTTCTCATAC
Dopamine Receptor 3 (DRD3)	DRD3_Dwn	Di	48 kb down	AGGTTGCCATTTAATTCTGT GTGTCTTGCCTATAATCCCAAAACTT
Dopamine Receptor 4 (DRD4)	DRD4_Mono	Mono	Intragenic	ACAGGCCCTGAGGTTTCC GTGGGGAAGGGGTGTTTC
Dopamine Receptor 5 (DRD5)	DRD5_Dwn	Di	3 kb down	AGGGAGTTTCACCATGTTAG ATGTTGTTTTACCCACTGGT
	DRD5	Di	17 kb down	CGTGTATGATCCCTGCAG GCTCATGAGAAGAATGGAGTG
	D4S615	Di	128 kb up	CTATACATCACCATTTGTCTGTGGC GCTAAGCTATTGCAGTAATTTGCTAC
Monoamine oxidase A (MAOA)	MAOA_Dwn	Di	65 kb down	GTGTCTTAGGAGCACCTACCTTTATGA GAGCAAAAGAATGAAACTCC
Monoamine oxidase B (MAOB)	MAOB_Up	Di	70 kb up	CTGTCTCCCAAATATGTCCT GTGTCTTCAGATGGTTATTCCTTCCAT
Catechol-O-methyltransferase (COMT)	COMT_Intra	Mono	Intragenic	GTGTCTTGAGACAGCAGAATTGCTTA GCAACTGTGAATGGATACAG

[a] Mono,mononucleotide repeat; di,dinucleotide repeat; tetra, tetranucleotide repeat.

[b] kb, kilobase; up, upstream from gene; down, downstream from gene.

Markers. Marker and primer selection was done as described by Schnack *et al.* [118]. All polymorphic microsatellite markers were within or near the selected genes of interest. For marker information see Table 2.1.

Genotyping and analysis. Preparation of pools, PCR and genotyping were performed as described by Schnack *et al.* [118]. Volumes containing equal amounts of DNA of individual samples were combined into two pools of patients (n=104 and n=104) and three pools of controls (n=109, n=90, and n=89). Genotypes of individual DNA samples were used to correct patterns of dinucleotide markers for PCR induced stutter artefacts (for details of the correction method see Schnack *et al.*[118]). After genotyping of the 10 individual samples, stutter correction models were derived for each dinucleotide marker. Pool patterns were then corrected for the artefact, in order to accurately determine the allele frequencies in the pools of cases and controls. No correction method was used for tetranucleotide markers, as these markers show far less stutter artefacts, and for mononucleotide markers, for which the contribution of stutter could not always be determined reliably.

Statistical analysis. The allele frequencies obtained from the stutter-corrected pool patterns were converted to counts and summed. Alleles with frequencies of less than 5% were clustered in one rest group. The allele counts of the combined pools of patients and controls were tested for significant differences with the CLUMP program [116]. For each comparison 10,000 Monte Carlo simulations were performed. We expect markers in close proximity of each other not to be completely independent because of linkage disequilibrium. Bonferroni correction would therefore have been too conservative, and correction was thus applied for the total number of tested genes rather than the total number of markers.

Individual genotyping was performed for each marker for which the allele frequencies differed significantly between patient pools and control pools, unless the difference was caused by alleles with a frequency of less than 5%. DNA extraction, PCR parameters, amplification and statistical analysis were the same as described for the pools. Two independent raters analyzed the genotypes, which included 99 random control duplicates. Hardy-Weinberg equilibrium was calculated using the GENEPOP program [119].

Results

Table 2.2 shows the results of the comparisons of the allele counts derived from the combined patient pools and combined control pools. Comparisons were made after correction for PCR artefacts. All markers were successfully genotyped except the DRD3 marker, which showed alleles that were separated by less than 2 base pairs and could therefore not be corrected for stutter artefacts. For marker DRD1_Down57kb, the genotyping of one of the control pools failed (n=109). Three markers reached P values < 0.05. However, when we combined the alleles with a frequency of less than 5% for these markers, and corrected for testing

11 genes, the allele frequencies between the pools of patients and the pools of controls were significantly different only for marker D4S615 (DRD5).

Table 2.2. Results of comparisons of allele counts derived from patient and control pools, after stutter correction for dinucleotide markers

Gene	Marker[a]	P value[b]
Tyrosine Hydroxylase (TH)	TH01	0.157
Dopa Decarboxylase (DDC)	DDC_Intra	0.672
	DBH_Up	**0.009**
Dopamine Receptor 1 (DRD1)	DRD1_Dwn55kb	0.101
	DRD1_Dwn57kb	0.734
Dopamine Receptor 2 (DRD2)	D11S3179	0.082
	RH27315	0.599
Dopamine Receptor 3 (DRD3)	DRD3_Dwn	n.a.
Dopamine Receptor 4 (DRD4)	DRD4_Mono	0.160
Dopamine Receptor 5 (DRD5)	DRD5_Dwn	**0.025**
	DRD5	0.233
	D4S615	***0.002***
Dopamine Transporter (DAT1)	D5S2005	0.082
Monoamine oxidase A (MAOA)	MAOA_Dwn	0.747
Monoamine oxidase B (MAOB)	MAOB_Up	0.198
	MAOB_Intra	0.510
Catechol-O-methyltransferase (COMT)	COMT_Intra	1.000

[a] Mono, mononucleotide repeat; kb, kilobase; Up, upstream from gene; Dwn, downstream.

[b] *P* value < 0.05 is printed in bold; *P* value < 0.00455 is printed in bold and italics (*P* value considered to be significant after Bonferroni correction for testing 11 genes).

Subsequently, three markers were genotyped individually in the extended patient sample and in the independent control sample: DRD3_Dwn, D4S615 and DRD5_Dwn. Because of failure of the stutter correction of marker DRD3_Dwn, we decided to genotype this marker individually, along with the DRD5 markers, DRD5_Dwn and D4S615. Marker DRD5_Dwn, although not reaching a *P* value < 0.00455, was selected because this marker is in much closer proximity to the DRD5 gene than marker D4S615. The difference for marker DBH_up was entirely due to the rare alleles that were combined in the rest group, and this marker was therefore not genotyped individually. Duplicate genotypes were in agreement, and the three markers were in Hardy-Weinberg equilibrium. No significant difference in allele frequencies was found for the DRD3_Dwn marker (p=0.43). Nor was there evidence for significantly different allele frequencies for the two DRD5 markers (DRD5_Dwn: p=0.35; D4S615: p=0.63) that were used.

Discussion

Using a comprehensive candidate gene analysis, we examined the association between schizophrenia and 12 genes involved in dopaminergic neurotransmission. In this large sample of 282 narrowly defined Dutch patients, we found no evidence for association with any of the markers close to or within these dopaminergic genes and schizophrenia. One possibility is that one or more of these genes are involved in the development of schizophrenia, but that the effect of the causative mutation is too small to detect in this study. We screened for association with polymorphic microsatellite markers in close proximity to, or within, candidate genes, not for particular functional variants. It is therefore possible that the microsatellite markers that we used are not in linkage disequilibrium with causative mutations within the dopaminergic genes of interest. This question is difficult to address. To date, in contrast to information about LD of numerous SNPs available from the HapMap project, little is known about disequilibrium between microsatellite loci, although there is evidence that microsatellite markers may detect LD over larger distances (of over 0.5 cM and even up to 2 cM [107, 120]) than SNPs.

Determining the power of a genetic association study requires several assumptions to be made, such as the frequency of marker alleles and disease-related alleles, linkage disequilibrium between them, and the effect size of the risk allele. Consequently, the exact power for a genetic association study is difficult to indicate [121], although a rough indication can be provided. If, for example, we assume that the studied markers are in complete linkage disequilibrium with disease-related mutations, and that DNA pooling does not influence the sensitivity of the tests, then the power to detect a high-risk allele with a relative risk of 1.3, or with a relative risk of 1.5, is 73% and 91%, respectively, in the pooled stage of the study, for both a marker and high-risk allele frequency of 0.2. Moreover, with our sample of 282 patients and 585 control subjects typed individually in the second stage of the study, under these same assumptions the power to detect a high-risk allele with a relative risk of 1.3 (1.5) is more than 90% (99%). Even after changing the assumption of complete linkage disequilibrium to one of considerably less linkage disequilibrium (for example D'= 0.65), the power of the second stage of this study would still be 81%, for a high-risk allele with a relative risk =1.5 [122].

In the first stage of this study, we found a significant association of a microsatellite marker close to the DRD5 gene, when we compared allele frequencies of pools of schizophrenic patients and controls. This marker, D4S615, was also associated with schizophrenia in a combined sample of Scottish and Irish schizophrenic patients [123]. However, when we subsequently extended our sample of patients, used an independent control group and genotyped all DNA samples individually to verify the association, we could not confirm our initial finding of association with this marker. There are at least three possible explanations for our different findings in stages 1 and 2. Firstly: the pooling method used might not be accurate enough in detecting association. This explanation seems unlikely, for we compared the allele frequencies of the patients that were typed in pools and the genotypes of these same patients when genotyped individually. Comparisons were made for markers D4S615 and DRD5_Dwn, and no significant differences between allele frequencies derived from pools and those from individual genotyping were found. Secondly: the different results may be due to population stratification: different control groups were used in the two stages of the study. However, for various other markers (not shown in this study) we tested for differences in allele frequencies between these two control groups and we detected small, but non-significant differences in allele frequencies. We should therefore not rule out a third possibility: the significant association between D4S615 and schizophrenia in stage one of the study was due to chance findings, and by increasing the power of the study in stage 2 we failed to replicate our initial findings.

Although substantial evidence points to altered dopaminergic function in schizophrenia, the combined data of our efficient and powerful two-stage study show no support for association with any of the dopaminergic genes we examined in a large Dutch sample. One must take into account that the polymorphisms we tested may not have been in linkage disequilibrium with disease-related alleles, but the results of this study could also imply that these genes are not directly involved in the pathophysiology of schizophrenia. However, the hypothesis of dopaminergic imbalance in schizophrenia still warrants further research. It is possible that it is not the genes implicated in the maintenance of dopaminergic transmission, but rather those involved in the development of dopamine neurons, or those important for dopamine neuron survival that influence susceptibility to schizophrenia.

Acknowledgements

The authors thank Jackie Senior for critically reading the manuscript.

Am J Med Genet B Neuropsychiatr Genet (2005) 134(1):6-9 (reprinted with permission).

3 NEUREGULIN 1: GENETIC SUPPORT FOR SCHIZOPHRENIA SUBTYPES

S.C. Bakker, M.L.C. Hoogendoorn, J.-P. Selten, W. Verduijn, P.L. Pearson, R.J. Sinke and R.S. Kahn

Abstract

The neuregulin 1 gene (NRG1) has repeatedly been shown to be associated with schizophrenia. We hypothesized that the heterogeneous disease course in schizophrenia is related to different susceptibility genes. Therefore, we have genotyped three previously reported markers at the NRG1 locus in 130 Dutch patients with deficit schizophrenia (characterized by enduring, idiopathic negative symptoms) and 130 patients with non-deficit schizophrenia, and compared these with 585 control individuals.

Single nucleotide polymorphism (SNP) SNP8NRG221533 ($p=0.004$) as well as two- and three-marker haplotypes ($p=0.001$-0.041) were associated with non-deficit schizophrenia, but not with the deficit syndrome ($p>0.31$). The associated SNP allele is different from the one in the previously reported at-risk haplotype in Iceland. Our results further support the evidence for NRG1 as a schizophrenia susceptibility gene across populations, and suggest that the course of illness in schizophrenia is influenced by different sets of genes.

SIR-The deficit syndrome is a form of chronic schizophrenia characterized by enduring, idiopathic negative symptoms [37]. Patients with this clinically distinct disease subtype have been suggested to differ in several biological aspects from patients with a more benign course with exacerbations and remissions [38]. We hypothesized that part of the heterogeneity of disease course in schizophrenia is related to different contributing genetic factors. The neuregulin 1 gene (NRG1), assumed to be involved in glutamate neurotransmission and neurodevelopment, was recently reported to be associated with schizophrenia in several, but not all populations [66, 124-129]. We have investigated the association of NRG1 with schizophrenia in patients with and without the deficit syndrome.

In order to enrich the sample for patients with the deficit syndrome, 282 schizophrenia patients were mainly recruited from psychiatric hospitals. Patients had at least three Dutch-born Caucasian grandparents. DSM-IV diagnosis of schizophrenia, excluding schizoaffective disorder, was made using the Comprehensive Assessment of Symptoms and History (CASH) [117] and information from medical records. The Schedule for the Deficit Syndrome (SDS) [36] was completed for 260 patients (92.2%), 130 of whom met deficit criteria. Twenty-nine patients had severe negative symptoms, which may have been secondary to factors such as substance abuse. Following the SDS, they were classified as non-deficit schizophrenia. The Medical Ethical Committee of the UMC Utrecht approved the study and all patients gave written informed consent. The control panel (n=585) consisted of 472 DNA samples from random Dutch individuals, obtained from the Immunogenetics and Transplantation Immunology Section of the Department of Immunohaematology and Blood Transfusion, LUMC, Leiden and 113 healthy controls from our department. Under optimal conditions, these samples have approximately 90% power to detect a locus with a relative risk (RR) of 1.5, and still 75% to detect a locus with a RR of 1.25.

All 5 single nucleotide polymorphisms (SNPs) from the previously reported at-risk haplotype [66] were first screened in DNA pools from an unselected subset of patients (n=208) and independent controls (n=179), using the SNaPshot technique (Applied Biosystems, Foster City, Ca, USA). SNP8NRG221533 was genotyped individually on a 7900HT TaqMan system (Applied Biosystems). Microsatellites 478B14-642 and 478B14-848 [66] were analyzed on an ABI 3700 sequencer (Applied Biosystems), and genotyped by two independent raters. GENEPOP [119] software was used to verify Hardy-Weinberg equilibrium (HWE). Haplotypes, linkage disequilibrium (LD) and likelihood ratios were calculated using UNPHASED software [130]. Alleles and haplotypes with frequencies <1% were combined. Global *P* values were calculated using n-1 degrees of freedom (n= number of alleles or haplotypes). No correction of *P* values for testing several markers and haplotypes was applied, because the markers were in LD and tests are therefore not independent.

SNP8NRG221533, which was the most strongly associated single marker in previous studies, showed a significantly higher frequency of the T allele in the schizophrenia pool than in controls (p<0.01). This marker was then genotyped individually in the extended sample, as was microsatellite 478B14-848, which is part of the reported at-risk haplotype [125]. Before testing for association, we calculated LD between the markers, and found it to be low (D'=0.20), which is in agreement with a recent study [128]. Therefore, we genotyped microsatellite 478B14-642, instead of the even more distant third microsatellite from the reported at-risk haplotype. This marker is located much closer to the SNP, and was recently found to be in a different LD block than 478B14-848 [128]. Indeed, in our sample, 478B14-642 was in considerable LD with SNP8NRG221533 (D'=0.60), but not with 478B14-848. All three markers were in HWE.

Interestingly, SNP8NRG221533 was associated with non-deficit schizophrenia (p=0.004), but not with deficit schizophrenia (p=0.542). Likewise, haplotypes of SNP8NRG221533 and 478B14-642 were significantly associated with non-deficit schizophrenia only (Table 3.1). It is not very likely that population stratification caused the association, because there were no differences in allele frequencies between patients originating from different areas of the Netherlands, or between the two control samples. Given the relatively low prevalence of deficit schizophrenia (15-30% of patients [38]), it is perhaps not surprising that we found NRG1, first identified in samples unselected for deficit schizophrenia, to be associated with non-deficit schizophrenia.

Table 3.1. Results of association analysis of three markers at the NRG1 locus

Marker[a]	**kb**[b]	**Variant**[c]	**Control**	**Schizophrenia**			
				Non-deficit		**Deficit**	
			%[d]	% [d]	*P* value[e]	% [d]	*P* value[e]
SNP8NRG221533	-	C	38.7	29.2	0.004	36.6	0.542
(M1)		T	61.3	70.8	0.004	63.4	0.542
		Overall			**0.004**		0.542
478B14-642	17	1	30.8	35.6	0.147	31.8	0.767
(M2)		2	4.9	5.9	0.528	4.3	0.670
		5	51.6	50.0	0.639	50.8	0.805
		6	5.0	3.5	0.304	4.7	0.828
		7	7.4	4.7	0.105	8.5	0.549
		Overall			0.362		0.959
478B14-848	97	3	14.1	13.4	0.775	16.1	0.409
(M3)		4	33.1	33.5	0.911	34.3	0.722
		5	13.6	15.0	0.582	12.9	0.761
		6	17.6	23.6	0.030	14.5	0.233
		7	20.9	13.4	0.005	21.0	0.977
		Overall			**0.048**		0.749
Haplotype M1-M2		C-1	1.3	0.0	0.027	1.6	0.793
		C-5	33.7	25.0	0.006	31.2	0.479
		T-5	17.8	25.0	0.010	19.6	0.539
		Overall			**0.001**		0.965

[a] for marker details see Stefansson *et al*. [66].

[b] distance in kilobases from SNP8NRG221533, which is located upstream of NRG1-isoform GGF2, at 23 kb from the start of exon 1.

[c] single alleles and haplotypes with frequencies >1% are shown. Specific haplotypes are shown if the *P* value for the haplotype was below 0.05. Overall= global *P* values for all alleles or haplotypes. NOTE: *P* values for single alleles/ haplotypes are presented without correction for the number of alleles/ haplotypes tested. These values should therefore be regarded as explorative results, indicating the largest differences.

[d] frequencies as percentage of all alleles and haplotypes.

[e] *P* value of likelihood test of 130 deficit and 130 non-deficit patients vs. 585 controls. *P* values <0.05 for global tests are shown in boldface. Post-hoc, the difference between deficit and non-deficit groups was significant (p=0.01), and SNP8NRG221533 was associated with schizophrenia in the entire sample (p=0.02). NOTE: haplotypes including M3 (overall *P* values= 0.009-0.041) are not presented. Since this marker was hardly in LD with the other markers, it is not likely to contribute much information to the two-marker haplotype.

Genotypes of CEPH individuals for M2 were: 133101=1/5; 133102=5/7; 134701=5/5; 134702=1/5, and for M3: 133101=3/3; 133102=4/4; 134701=4/7; 134702=4/6. The 4-allele of M3 corresponds to the associated 219-bp allele in the Icelandic at-risk haplotype (V. Steinthorsdottir, personal communication).

The putative functions of NRG1 in gliogenesis, neuronal migration and synaptic plasticity [66] seem to indicate a neurodevelopmental cause in non-deficit schizophrenia. Our results support previous suggestions that deficit schizophrenia may represent a distinct biological entity [38]. The rapid and persistent clinical deterioration seen in many deficit patients could be the result of a degenerative process, dominated by other than the recently suggested gene systems, and environmental factors. The identification of susceptibility genes for deficit schizophrenia may be very valuable, since negative symptoms are a cause of much disability, and often resistant to the current treatment. Compared with the findings in Iceland, different SNP alleles and haplotypes are associated in the Dutch population. Different associated NRG1 haplotypes have been described [126, 129] and, as pointed out, it is not unexpected that mutations in different populations have a different haplotype background [131]. Likewise, different haplotypes were found to be associated for dysbindin (DTNBP1), another schizophrenia susceptibility gene [132]. Additional fine mapping studies are required to identify the causative variation(s), and to establish if NRG1 or a nearby gene is involved [129].

In conclusion, our results further support the evidence for NRG1 as a schizophrenia susceptibility gene across populations, and suggest that the course of illness in schizophrenia is influenced by different sets of genes.

Mol Psychiatry (2004) 12: 1061-63 (reprinted with permission).

4 ASSOCIATION OF DEFICIT AND NON-DEFICIT SCHIZOPHRENIA WITH THE PIP5K2A AND RGS4 GENES, NOT WITH THE DYSBINDIN AND G72/G30 GENES

S.C. Bakker, M.L.C. Hoogendoorn, J. Hendriks, K. Verzijlbergen, S. Caron, H.G. Otten, W. Verduijn, J.-P. Selten, P.L. Pearson, R.S. Kahn and R.J. Sinke

Abstract

The clinical presentation of schizophrenia patients is very heterogeneous, suggesting the existence of disease subtypes with different underlying causes. Previously, we have reported an association of the neuregulin 1 gene with non-deficit schizophrenia, but not with deficit schizophrenia (characterized by enduring, idiopathic negative symptoms). This study investigates the role in both clinical subtypes of the dysbindin, G72/G30 and RGS4 genes, repeatedly reported to be associated with schizophrenia, and the PIP5K2A gene, which has hardly been studied in schizophrenia thus far. Selected single nucleotide polymorphisms (SNPs) were genotyped in 273 Dutch schizophrenia patients, 146 with deficit schizophrenia, and 580 controls. SNPs in the dysbindin and G72/G30 genes were not significantly associated with (non-) deficit schizophrenia. In the RGS4 gene, however, one SNP (p=0.03) and a two-marker haplotype (global p=0.02) were associated with the non-deficit type only. Finally, in the PIP5K2A gene, we detected strong association with a SNP that leads to a change in protein structure (p=0.00005). There was no apparent difference between deficit and non-deficit patients. In conclusion, our data suggest that the RGS4 gene is associated with non-deficit schizophrenia, and that the PIP5K2A gene is a new susceptibility gene for schizophrenia in general.

Introduction

Schizophrenia is a mental disorder that affects approximately 1% of the population worldwide. With a peak onset in late adolescence and early adulthood, it causes much individual suffering and large health care expenses for society. Family studies have consistently shown that the susceptibility to schizophrenia is largely determined by heritable factors (approximately 80 percent). Most likely, multiple genes and environmental factors are involved, which may explain why the identification of schizophrenia susceptibility genes has proved to be difficult. Recently, however, systematic fine mapping efforts in regions of replicated linkage have resulted in the identification of several genetic variations that are associated with schizophrenia. Evidence is now accumulating for the involvement of genes near these variations in schizophrenia and related psychiatric disorders [65].

One of these genes, dysbindin (DTNBP1), was identified as a schizophrenia susceptibility gene after fine mapping of a linkage region on chromosome 6p22. The first reported association in Irish multiplex families [67] has been replicated in schizophrenia cohorts from Ireland [133] [134] and the UK [132], Germany/Israel [135], Sweden [136], Bulgaria [137] and China [138]. No association was found in German and a Polish samples [136]. There seems to be extensive molecular heterogeneity, with different markers and their combinations (haplotypes) being associated across populations.

G72 and G30 are two overlapping genes, on complementary DNA strands, at chromosome location 13q32-33, a linkage region for schizophrenia and bipolar disorder. Fine mapping of a 5 Mb critical interval resulted in detection of significant association of schizophrenia with several single nucleotide polymorphisms (SNPs) and SNP haplotypes in a 65 kilobase (kb) region that comprised both genes [68]. These results, from a French-Canadian and a Russian sample, have been replicated in samples from Germany [139], China [140] and in an American sample that also included psychosis NOS [141]. Interestingly, SNPs in this region appeared to be also associated with bipolar disorder in several studies, further suggesting a shared genetic background for both disorders [142, 143].

The first indications for involvement of the RGS4 gene came from gene expression studies, which showed a decreased expression in brains of schizophrenic patients [69]. In addition, the gene is located in a schizophrenia linkage region on chromosome 1q21-q22 [144]. A comprehensive association study of the region around the gene resulted in a group of four associated SNPs in three samples with a Caucasian, a mixed and an Indian background [144]. Involvement of the same SNPs was subsequently confirmed in two Irish [145, 146] and a UK/Irish sample [147]. RGS4 is involved in the regulation of G protein-coupled receptors, and may modulate dopamine and glutamate receptors [144].

Chromosome region 10p12 was also repeatedly linked to both schizophrenia and bipolar disorder. This region harbours the PIP5K2A gene, the product of which synthesizes phosphatidylinositol 4,5 bisphosphate, a membrane phospholipid that plays a central role in signal transduction and trafficking of synaptic vesicles. Lithium, which is used to treat bipolar disorder, is known to block inositol monophosphatases, enzymes that are part of the same phosphoinositide pathway [148]. Different lines of evidence therefore indicate that PIP5K2A could be involved in both schizophrenia and bipolar disorder. Recently, an intragenic CT repeat polymorphism was reported to be more frequent in patients with bipolar disorder than in controls [148]. In a different study, several SNPs in the PIP5K2A gene were shown to be associated with schizophrenia [149].

These recently identified genes may provide insights into the biological origin of schizophrenia. However, their reported effect is typically modest, with estimated relative risks (RRs) of less than two, and the variations or haplotypes that increase the disease risk are present in some patients only. Moreover, some genes appear to be associated with both schizophrenia and bipolar disorder. These observations suggest that specific genes are associated with specific symptoms, which may be unique to, or shared by, clinically defined diseases.

Schizophrenia symptoms can be broadly divided into two classes. So-called positive symptoms include hallucinations and delusions, while lack of initiative, social withdrawal and flattened emotions are known as negative symptoms. The presence of both types of symptoms in individual patients can vary considerably, with a virtual absence of either positive or negative symptoms in some patients. Deficit schizophrenia, which is diagnosed using the Schedule for the Deficit Syndrome (SDS) [36], is characterized by enduring, idiopathic negative symptoms and a generally poor prognosis. Approximately 15% of first episode patients and 25-30% of chronic patients fulfil deficit criteria [38]. A number of observations indicate that deficit schizophrenia has a high heritability [43], and may represent a distinct biological entity [38].

Recently, we found indications that the neuregulin 1 gene is associated with non-deficit schizophrenia, but not with deficit symptoms, which supports the hypothesis that specific gene variants may predominate in particular disease subtypes [150]. In the present study, we have investigated whether the dysbindin, G72, PIP52A, and RGS4 genes are associated with deficit and non-deficit schizophrenia, in a sample of Dutch schizophrenic patients that was enriched for the deficit form of the disorder.

Materials and methods

Sample collection. In order to enrich the sample for patients with the deficit syndrome, 308 schizophrenia patients were mainly recruited from psychiatric hospitals. Patients had at least three Dutch-born Caucasian grandparents. DSM-IV diagnosis of schizophrenia, excluding schizoaffective disorder, was made using the Comprehensive Assessment of Symptoms and History (CASH) [117] and information from medical records. The Schedule for the Deficit Syndrome (SDS) [36] was completed for 273 patients (89%), 146 (53%) of whom met deficit criteria. Patients without a completed SDS were excluded. In twenty-nine patients with severe negative symptoms, it could not be ruled out that these symptoms were secondary to factors such as substance abuse. Following the SDS, they were classified as non-deficit schizophrenia. The Medical Ethical Committee of the UMC Utrecht approved the study and all patients gave written informed consent. The control panel for DNA pooling studies (n=179) consisted of DNA from blood bank donors obtained from the Department of Immunology of the UMC Utrecht. The independent control panel for individual genotyping (n=580) consisted of 467 DNA samples from random Dutch individuals, obtained from the Immunogenetics and Transplantation Immunology Section of the Department of Immunohaematology and Blood Transfusion, LUMC, Leiden and 113 healthy controls from the Department of Biomedical Genetics, UMC Utrecht.

Genotyping. In the dysbindin and G72/G30 genes, previously described SNPs were first screened in DNA pools. Pools were made from a random subset of patients (n=204) and controls (n=179) as described elsewhere [118]. SNPs (see Table 4.1) were analyzed in pools with the SNaPshot technique of single base extension (Applied Biosystems, Foster City, Ca., USA, further called ABI) on a 3700 capillary sequencer (ABI).

Individual genotyping of SNPs in dysbindin and G70/G30 was performed on a 7900HT TaqMan system (ABI). Microsatellite Dys_2 was identified in human sequence data from the Ensembl database using Tandem Repeats Finder software [151], and the following primers were designed with the Primer3 program [152]: CACCAACAACATTCAATCTGAG and TGTTTTTCCATTCGTGTCATC. The marker is located between P1325 and P1765 [67], at 13 kb from P1325. It was analyzed on an ABI 3700 sequencer as described elsewhere [118].

Single nucleotide polymorphisms with flanking sequences for RGS4 were obtained from a previous report [144], and for PIP5K2A from the Celera database [2], based on previously reported data [149]. SNPs were genotyped using a novel technique, which was recently developed at our department (unpublished data). Basically, the PCR involves two allele-specific labelled primers with a 3' locked nucleic acid (LNA), which is a modified SNP-binding nucleotide with a high binding affinity [153, 154]. Pooled PCR products from different SNPs were separated by size on a 3700 capillary sequencer (ABI). We have previously validated this SNP detection technique by comparing genotypes for multiple SNPs with those obtained from a TaqMan system (ABI), with perfect agreement of results (data not shown). To further ensure reliability, all four RGS4 and PIP5K2A SNPs were also blindly genotyped by direct sequencing, in a random selection of 24 samples. Each 96-well PCR plate contained 9 blind duplicate samples.

Data analysis. DNA pools: relative peak heights for each allele, without correction for possible differential signal intensity, were compared with a chi-square test. With individual genotypes, Hardy-Weinberg Equilibrium (HWE) was tested using a chi-squared test for SNPs, and GENEPOP software for the microsatellite [119]. Linkage disequilibrium (LD) analyses were performed using GOLD software [155] and compared with data from the HapMap project using the Haploview program [156], with haplotype blocks defined according to Gabriel *et al.* [20]. Likelihood ratio tests for single markers and marker haplotypes were performed using the UNPHASED program [130]. Reported *P* values are two-tailed, without correction for multiple testing.

Results

Previously described SNPs at the G72/G30 locus and the dysbindin locus were first screened in DNA pools from 204 patients and 179 controls. Markers were selected for individual genotyping based on the results of pooling experiments and reported associations in previous studies. All individually tested markers were in HWE; re-sequencing of 24 random individuals for the PIP5K2A and RGS4 SNPs confirmed genotyping results obtained by allele-specific PCR.

Dysbindin. Allele frequencies for 5 out of 17 dysbindin SNPs tested in DNA pools were significantly different at the p=0.05 level (Table 4.1a). These markers, as well as a new microsatellite, Dys_2, in the centre of previously reported associated haplotypes, were individually genotyped in all persons that contributed to the pools. The differences remained significant for SNPs rs2619528, rs760761 and rs2619522 (data not shown, *P* values 0.02-0.03). However, as shown in Table 4.2, these three markers formed one haplotype block. Therefore, only SNP rs2619528, which 'tagged' all haplotypes with a frequency >0.5%, was analyzed in the entire sample. Marker Dys_2 had three alleles with a frequency of >5% and was in strong LD with all the other markers, but it showed no significant allele frequency differences between cases and controls (data not shown). Table 4.3 lists allele frequencies and the resulting *P* values of likelihood tests for single markers and marker haplotypes that were tested in the entire sample. None of the markers was significantly associated with deficit or non-deficit schizophrenia, or the combined samples. We also tested two- to five-marker haplotypes, none of which was significantly associated (data not shown).

G72/G30. None of the 11 SNPs at the G72/G30 locus showed significant differences between cases and controls in pooled DNA (Table 4.1b). Four SNPs that have repeatedly been associated with schizophrenia were also genotyped individually in the entire sample. SNPs M14-M15, and M23-M24 formed two haplotype blocks, but there was hardly any LD between the blocks (Table 4.2). There was no significant association of single markers (Table 4.3), or any 2-, 3- or 4-marker haplotype, single or global, with deficit or non-deficit schizophrenia, or with schizophrenia in general (data not shown).

RGS4. LD between the two tested markers was significant in controls (D'=0.85; r^2=0.55) as well as in patients (D'=0.95; r^2=0.75). The G-allele of SNP RGS4-1 was more frequent in non-deficit patients than in controls (p=0.033), as were two-marker haplotypes (global p=0.022 in the non-deficit patients; combined samples p=0.025).

Table 4.1. DNA Pooling results and published studies for dysbindin (a) and G72/G30 (b)

a				**Ref**[e]	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
				Origin[f]	Irl	Ger	Irl	Swe	Irl	UK	Bul	Chi
DYSBINDIN				Case[g]	270	203	219	142	268	708	488	233
				Contr[g]	fam	fam	231	272	fam	711	trios	trios
Marker[a]	**dbSNP**[b]	**kb**[c]	**p**[d]	**Pheno**[h]	sa	sa	sa	sa/sf	sa	s	sa	s
[P]	rs1047631	0					-			-		
P1328	rs742106	1379	0.636		-				-			-
P1333	rs742105	48594	0.785		X				-			
[M]		13058									-	
P1655	rs2619539	34723	0.610		X		-		-	-	-	-
[J]		6916								X		
P1635	rs3213207	331	0.010		X	X	-	-	-	-	X	
P1325	rs1011313	5330	0.899		-	X		X	-	-	-	
P1765	rs2619528	3395	0.010		-	X			X			
P1757	rs2005976	973	0.004		-		-	-	X	-	X	
P1320	rs760761	330	0.006		X	X	-	-	X			
P1763	rs2619522	2517	0.008		-	X			-			-
P1578	rs1018381	3421	0.315		-	-		-	-			-
P1792	rs1474605	1142							X			
P1583	rs909706	2659	0.226		-				-	-	-	-
[C]	rs2743852	3893								X	-	
[A]	rs2619538	445								-	-	
b				**Ref**[e]	**IX**	**IX**	**X**	**XI**	**XII**	**X**	**XIII**	**XIV**
				Origin[f]	Can	Rus	Ger	USA	Chi	Ger	USA	USA
G72/G30				Case[g]	213	183	299	98	537	300	174	139
				Contr[g]	241	183	300	trios	538	300	fam	113
Marker[a]	**dbSNP**[b]	**kb**[c]	**p**[d]	**Pheno**[h]	s	s	s	nos	s	bp	b/a	b/a
M12	rs3916965	0	0.421		X	-	X		X	-		
	rs1935058	7990						-			X	-
	rs1341402	4159					-			-	X	
M14	rs3916967	1839	0.277		X	-		X	-			
M15	rs2391191	2098	0.156		X		X	X	X	-	-	-
	rs1935062	8690					-	-	-	-	X	X
M18	rs947267	11526			-	-					-	X
M19	rs778294	2573			-		-	-		-	X	
	rs954581	10031									-	-
M22	rs778293	16933	0.256		X	-		-				
M23	rs3918342	16550	0.991		X	X	X		-	X	-	
M24	rs1421292	12486	0.936		X	-	X			-		

NOTE: Only markers with significant association (p<0.05) as single marker, or as part of associated haplotypes in any study are shown. Consequently, pooling results for eleven additional SNPs, none of which were associated, are not listed. (-), no significant result as single marker; (X), significant result as single marker; light grey background, part of any associated haplotype; dark grey background, part of most strongly associated haplotype.

[a] Between brackets: marker names used by Williams *et al*. [132], P-numbers used by Straub *et al*. [67], M-numbers used by Chumakov *et al*. [68].

[b] SNP names in dbSNP.

[c] Distance in kb from the previous marker.

[d] *P* value of chi-squared test of cases and controls in DNA pools.

[e] References: I [67], II [135], III [133], IV [136], V [134], VI [132], VII [137], VIII [138], IX [68], X [139], XI [141], XII [140], XIII [142], XIV [143].

[f] Origin of study samples. Irl, Ireland; Ger, Germany; Swe, Sweden; UK, United Kingdom and Ireland; Bul, Bulgaria; Chi, China. No association with dysbindin was detected in additional Polish and German samples in the study by Van den Bogaert *et al*. [136].

[g] Numbers of cases and controls. Trios, nuclear families; fam, extended families. The study by Hattori *et al*. [142] involved two samples of 152 and 22 families.

[h] Phenotypes, according to DSM-criteria. sa, schizophrenia plus schizoaffective disorder; sf, schizophreniform disorder; s, schizophrenia; nos, schizophrenia plus psychosis not otherwise specified; b, bipolar disorder; a, schizoaffective disorder.

PIP5K2A. The two tested SNPs in PIP5K2A were in strong LD, in cases (D'=0.93; r^2=0.63) and in controls (D'=0.78; r^2=0.55). The A-allele of coding SNP hCV11558870 was approximately 10% more frequent in the combined sample (p=0.00005), with a comparable contribution of the deficit and non-deficit groups. Haplotypes with SNP hCV9591220, which was not significantly associated as a single marker, were less strongly associated (global p=0.009).

Table 4.2. Linkage disequilibrium in dysbindin and G72/G30

	DYSBINDIN								**G72/G30**			
	P1655	**P1635**	P1325	Dys_2	**P1765**	P1320	P1763	**SNP A**	**M14**	**M15**	**M23**	**M24**
P1655		0.15	0.11	n.a.[a]	n.s.	n.s.	n.s.	0.06				
P1635	0.94		n.s.	n.a.	0.49	0.5	0.47	0.08				
P1325	1	0.34		n.a.	0.03	0.03	0.03	n.a				
Dys_2	0.64	0.88	1		n.a	n.a	n.a	n.a.				
P1765	n.s.[b]	0.96	1	0.97		0.97	0.98	0.14				
P1320	n.s.	0.97	1	0.94	0.98		0.98	n.a.				
P1763	n.s.	0.94	1	0.98	1	1		n.a.				
SNP A	0.21	0.80	n.a.	n.a.	0.76	n.a.	n.a.					
M14										0.99	0.02	0.03
M15									1		0.02	0.03
M23									0.14	n.s.		0.91
M24									0.20	0.19	1	

NOTE: D' values for controls are shown below the diagonal, while r^2 values are shown above the diagonal. Haplotype blocks are shown on a grey background. Marker names in boldface were genotyped individually in the entire sample (see Table 4.3); pair wise LD between these markers was calculated in the entire sample.

[a] n.a., not available.

[b] n.s., not significant at the p=0.05 level.

Discussion

Several genes have repeatedly been shown to be associated with schizophrenia, but at present, the relevance of particular gene variants in different populations, or in clinical subtypes of schizophrenia, remains unclear. We have investigated the association of variations in the dysbindin, G72, RGS4 and PIP5K2A genes with deficit or non-deficit forms of schizophrenia.

Table 4.3. Individual genotyping results

				Control	Schizophrenia					
Gene	**Marker**[a]	**dbSNP**[b]	**Var**[c]		NDef[d]	*P*[e]	Def[f]	*P*[e]	Total[g]	*P*[e]
Dysbindin	P1655	rs2619539	C	53.3	53.9	0.873	57.9	0.183	56.1	0.307
	P1635	rs3213207	G	87.4	87.0	0.861	89.3	0.371	88.2	0.630
	P1765	rs2619528	G	21.8	22.6	0.766	21.5	0.932	22.0	0.898
	SNP A	rs2619538	A	0.58	0.62	0.24	0.55	0.38	0.58	0.918
			Haplo			n.s.		n.s.		n.s.
G72	M14	rs3916967	C	40.5	37.7	0.407	37.9	0.421	37.8	0.292
	M15	rs2391191	C	61.0	62.6	0.649	62.4	0.677	62.5	0.574
	M23	rs3918342	T	46.9	44.4	0.499	48.6	0.610	46.7	0.940
	M24	rs1421292	A	53.9	55.2	0.747	54.9	0.782	55.0	0.696
			Haplo			n.s.		n.s.		n.s.
RGS4	RGS4-1		G	37.8	45.4	**0.033**	38.7	0.790	41.9	0.131
	RGS4-18		A	55.8	50.8	0.155	54.0	0.580	52.5	0.209
			A_A	52.2	49.0	0.336	54.5	0.568	51.9	0.832
			G_G	35.1	43.5	**0.020**	38.6	0.339	40.9	**0.038**
			A_G	9.2	5.2	**0.047**	6.5	0.192	5.9	**0.034**
			Global			**0.022**		0.269		**0.025**
PIP5K2A	hCV11558870		A	61.0	70.3	**0.007**	72.5	**0.0004**	71.5	**0.00005**
	hCV9591220		G	38.8	34.3	0.195	34.9	0.236	34.6	0.109
			A_G	7.6	7.7	0.884	8.8	0.408	8.3	0.514
			A_A	54.3	63.6	**0.012**	63.9	**0.006**	63.8	**0.001**
			G_G	33.2	26.9	0.052	26.4	**0.025**	26.6	**0.007**
			Global			0.084		**0.040**		**0.009**

NOTE: alleles and haplotypes with frequencies >5% are shown.

[a] Marker names in different studies. P-numbers used by Straub *et al*. [67]; M-numbers used by Chumakov *et al*. [68]; SNP A used by Williams *et al*. [132]; RGS4-1 and RGS4-18, names according to Chowdari *et al*. [144]; hCV11558870 and hCV9591220, names in Celera database.

[b] SNP names in dbSNP.

[c] Alleles and haplotypes. Haplo: haplotypes, single and global; global, global *P* value for haplotype.

[d] NDef, non-deficit patients (n=127).

[e] *P*, *P* value of likelihood test of patient versus control groups.

[f] Def, deficit patients (n=146).

[g] Total, combined deficit and non-deficit samples (n=273).

We first screened most of the previously reported SNPs in dysbindin and G72/G30 in DNA pools. Several SNPs in dysbindin, but not in G72/G30, had significantly different allele frequencies in DNA pools. Individual genotyping of the pools confirmed this association, but with marginally significant results. It should be noted that only single SNPs can be compared in DNA pools, and that negative findings do not exclude association with marker haplotypes. In order to reduce the risk of missing associations, the selection of markers for individual genotyping in the entire sample was based on the pooling data, as well as previously reported associations and LD analyses in a subset of cases and controls. Patterns of LD were very similar to those in previous reports, and LD data from the HapMap project.

For the dysbindin gene, we could not confirm the marginally significant association in DNA pools in the entire sample, which suggests that the pooling results in the smaller samples were chance findings. Haplotypes also failed to show any association with schizophrenia. It is possible that our markers were not in strong LD with disease-related variants, since there appears to be substantial molecular heterogeneity, even in geographically related patient samples (see Table 4.1a for a summary of published studies). In a recent study, association was only detected after including a new SNP, which was part of all associated haplotypes [132]. The same markers were associated in a Scottish sample that was originally reported as a non-replication [133]. This marker, nor haplotypes that included it (p>0.37), were associated in our sample. Our set of markers included SNPs that have repeatedly shown association as single markers or as (part of) associated haplotypes, but we cannot exclude that yet other haplotypes are associated in the Dutch population [132].

The specific character of our sample provides another explanation for the lack of association. As pointed out, the expected relatively small effects can easily be missed by chance in the sample sizes that have been used so far [136]. Our sample did not include patients with schizoaffective disorder, and was enriched for patients with a poor disease outcome. In contrast, all recently reported positional candidate genes were identified in unselected samples, probably with relatively few deficit patients. If there is a genetic difference between the two disease subtypes, these genes are therefore less likely to be associated with deficit schizophrenia. Our non-deficit group may have been too small to detect minor genetic effects, if these were confined to this subtype. However, the data did not even show a trend towards a difference between the clinical subgroups and, accordingly, it is less likely that dysbindin plays a substantial role in schizophrenia in the Dutch population, unless different, yet unknown, variations in the gene are involved. In different reports, associations involved different markers and haplotypes, but they all cluster in a region of approximately 140 kb (Table 4.1a). Despite negative findings in some samples, including ours, the evidence for dysbindin as a schizophrenia susceptibility gene is accumulating, and the combined studies seem to provide sufficient ground for a thorough search for schizophrenia-related variants.

At the G72/G30 locus, none of 11 previously reported markers showed significantly different allele frequencies when tested in DNA pools. We genotyped four markers individually that have been repeatedly associated with schizophrenia in previous studies. LD patterns were in agreement with previous reports and HapMap data, but there was no association with single markers or their haplotypes. The general explanations for negative findings as discussed above also hold for G72/G30, and it seems premature to exclude the involvement of G72/G30 in our population based on these results. Several studies have now reported association with schizophrenia and, interestingly, also with bipolar disorder. Table 4.1b shows that associations were detected with different markers and haplotypes in a region spanning approximately 100 kb. Although the first reports implicated markers around M23-M24, most recently reported associations seem to cluster in the region around M14-M15. Taken together, the evidence for involvement of G72/G30 in both schizophrenia and bipolar disorder seems convincing.

In RGS4, four markers within 10 kb around the start of exon 1 have been implicated in schizophrenia. In previous studies, the four markers were in strong LD, in particular combinations RGS4-1 with -7, and RGS4-4 with -18. We selected the two most distant markers, RGS4-1 and RGS4-18, which we still found to be in strong LD. The G-allele of SNP RGS4-1, as well as the two-marker G-G haplotype had a higher frequency in schizophrenia patients, which could almost completely be attributed to the non-deficit group. Chen *et al.* recently found the strongest association in the most narrowly defined schizophrenia group [145], and our data suggest that the gene is specifically involved in a further refined subgroup of schizophrenia patients without prominent negative (deficit) symptoms. After correction for testing four genes and two phenotypes, *P* values are no longer significant at the p=0.05 level. Likewise, the evidence in previous studies was modest. However, the associated haplotypes in all Caucasian populations seem consistent, with the G-allele being associated with schizophrenia at each SNP, and the combined genetic and gene expression studies suggest that RGS4 is involved in schizophrenia, or in specific clinical aspects of the disorder. The role of RGS4 in modulating dopamine and glutamate neurotransmission, as well as its altered expression in the brains of schizophrenic patients, make it a plausible functional candidate gene. Interestingly, it was recently shown in mice that a transient influenza virus infection caused an altered RGS4 expression, and persistent changes in emotional and cognitive functions [157], which may provide an example of interacting environmental and genetic factors in schizophrenia. Replication studies are necessary, however, preferably in samples with different ethnic backgrounds, in order to evaluate the presence of allelic heterogeneity between populations, as suggested by the results of Chowdari *et al.*

The most convincing evidence for association with schizophrenia in our sample was found for PIP5K2A. This gene is located in a linkage region that is common to schizophrenia and bipolar disorder, and interestingly, it is part of the pathway that is directly modulated by lithium [148]. PIP5K2A may therefore predispose to both disorders. The reported association of a CT-repeat marker in intron 9 with both schizophrenia and bipolar disorder [148], and an independent report of association of several SNPs in the same 3' region of the gene with schizophrenia [149], guided our choice of markers. SNP hCV11558870, located in exon 7, is one of the very few coding SNPs in the gene, causing an amino acid change from serine to asparagine. Interestingly, we found this SNP to be strongly associated as a single marker, with both deficit and non-deficit schizophrenia. Haplotypes with the second SNP, with which it is in strong LD, were less significantly associated. This observation implicates that the coding SNP could actually be a causal variant. If so, it would be a polymorphism with a high frequency in the general population. Schizophrenia and bipolar disorder share affective (mood) symptoms, but since our sample did not include patients with schizoaffective disorder, one would perhaps expect the gene to be associated with psychotic symptoms, rather than with mood symptoms. This hypothesis could be tested in samples that were selected for psychotic or affective symptoms, instead of clinically defined diseases. Taken together, the few available studies suggest that PIP5K2A is a new susceptibility gene for psychiatric disorders, which deserves more attention.

In a case-control design, it cannot be excluded that population stratification caused differences between the groups. However, in the current and previous studies, we have tested many different markers in both samples, with mostly negative results. Moreover, we found no allele frequency differences between groups of individuals from different regions in the Netherlands (data not shown). Therefore, stratification seems a less likely cause of the detected differences.

In summary, we have studied a selection of previously reported variations in the dysbindin, G72/G30, RGS4 and PIP5K2A genes for association with deficit and non-deficit schizophrenia. Support was found for a contribution of RGS4 to non-deficit schizophrenia, while a coding SNP in PIP5K2A was strongly associated with schizophrenia, regardless of clinical subtype. In combination with our previous findings in the neuregulin 1 gene [150], these data suggest that some genes increase susceptibility to schizophrenia in general, while others are more relevant to disease subtypes or specific symptoms. Studies in stratified samples may facilitate the identification of new schizophrenia susceptibility genes, and may be necessary to identify susceptibility genes that are unique to deficit schizophrenia. Identifying such genes would be very valuable, since this may lead to much-needed new forms of treatment. Deficit patients represent a minority of about 20% of all schizophrenics, but their symptoms are often treatment-resistant and severely disabling.

Acknowledgements

The authors thank N.M. Williams for kindly providing details of SNP A, and associated haplotypes prior to their publication [132]. This project was supported by the UMC Utrecht's Genvlag program, ABC Neurogenomics and the Makaria Foundation.

Submitted for publication.

SECTION II

GENETIC STUDIES IN ATTENTION-DEFICIT HYPERACTIVITY DISORDER (ADHD)

5 DAT1, DRD4 AND DRD5 POLYMORPHISMS ARE NOT ASSOCIATED WITH ADHD IN DUTCH FAMILIES

S.C. Bakker, E.M. van der Meulen, N. Oteman, H. Schelleman, P.L. Pearson, J.K. Buitelaar, and R.J. Sinke

Abstract

Recent meta-analyses have indicated that the dopamine transporter gene (*DAT1*) and the dopamine receptor genes D4 (*DRD4*) and D5 (*DRD5*) are associated with attention-deficit hyperactivity disorder (ADHD), although single studies frequently failed to show significant association. In a family-based sample of 236 Dutch children with ADHD, we have investigated the previously described variable number of tandem repeat (VNTR) polymorphisms and two additional microsatellites at the DAT1 and DRD4 loci. DRD5 was investigated using the microsatellite that was previously found to be associated. Transmission disequilibrium tests (TDTs) did not show preferential transmission of alleles or two-marker haplotypes to affected offspring. These data suggest that DAT1, DRD4 and DRD5 do not contribute substantially to ADHD in the Dutch population.

Introduction

Attention-deficit hyperactivity disorder (ADHD) is the most common child psychiatric disorder, affecting 4-5% of children in western countries [72]. The estimated contribution of genetic factors is approximately 80 percent [77, 158], and it is likely that multiple genetic and environmental factors interact in causing the disease. Genetic research has focused on genes involved in neurotransmission, and in particular the dopaminergic system, since effective medication was reported to block the reuptake of dopamine by the dopamine transporter molecule (DAT1) [159]. Three dopaminergic genes have recently been reported to be associated with ADHD in meta-analyses of data from multiple studies. The DAT1 gene has a variable number of tandem repeat (VNTR) polymorphism in the 5' untranslated region (UTR). A recent meta-analysis concluded that having a 10-repeat allele increased the risk to develop ADHD by a factor of approximately 1.3 [81]. The VNTR may change DAT1 function, since it has been suggested to regulate gene expression [160, 161].

The dopamine receptor D4 gene (DRD4) has a VNTR polymorphism in the third exon, which is part of the third intracellular loop of the receptor, and may therefore have functional relevance [162, 163]. Recent meta-analyses confirmed that the 7-repeat allele increased the risk of developing ADHD 1.4 to 2.0 times [81, 82]. Children with the 7-repeat allele were found to have a more inaccurate, impulsive response style on neuropsychological tasks [164].

In a meta-analysis of data from 14 different centres, the common 148 base pair (bp) allele of a compound microsatellite located 18.5 kb from the dopamine receptor D5 gene (DRD5) was shown to be significantly associated with ADHD (odds ratio 1.24) [83].

The three polymorphisms mentioned above, as well two additional microsatellites near the VNTRs in DAT1 and DRD4, were genotyped in a sample of 236 Dutch children from 144 families.

Materials and Methods

Most children (n=198) were part of a previously described sample of sib pair families [165]. Children were only included if they had ADHD of the inattentive, hyperactive or combined subtype, according to DSM-IV criteria. Children with autism spectrum disorders were excluded. This sample was extended with 38 families with only one affected child, diagnosed using the same criteria. In 5 families, no DNA from the father was available.

DNA was isolated as described [165]. The VNTR polymorphisms in the DAT1 and DRD4 genes were amplified using previously described PCR primers [166, 167]. Reactions were performed in 50 µl, containing 50 ng of genomic DNA, 100 ng of forward primer and 100 ng of reverse primer, 150 mM of each dNTP, 67 mM TrisHCl, 6.7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 6.7 µM EDTA, 16.6 mM $(NH_4)_2SO_4$, 10% DMSO, 7.5 µg BSA and 0.4 units of AmpliTaq polymerase (Applied Biosystems, Foster City, CA, U.S.A.). PCR reactions were performed on a ABI 9600 GeneAmp PCR system (Applied Biosystems) using the following conditions: 2 min at 94°C, followed by 33 cycles of 30s at 94°C, 30s at 60°C (DAT1) or 54°C (DRD4), 2 min at 72°C and a final extension of 4 min at 72°C. Subsequently, 10 µl of PCR product was analyzed on a 3% agarose gel, by applying 125V for a duration of 2 hours. Fragments were stained with ethidium bromide and sizes were determined using a PGEM DNA size marker (Promega, Leiden, The Netherlands). In order to determine the repeat numbers of the different alleles of both genes, sequence analysis of the repeat regions were performed in several individuals. The DRD4 mononucleotide repeat, located in intron 1, was amplified with primers ACAGGCCCTGAGGTTTCC and GTGGGGAAGGGGTGTTTC [168]. Primers for dinucleotide repeat D5S2005, which is located at 50 kilobases (kb) from the DAT1 VNTR, were obtained from the Ensembl database. These microsatellites, as well as the DRD5 repeat [169], were amplified and analyzed as described elsewhere [165]. Two independent raters scored all genotype data, and when they disagreed genotyping was repeated. Inheritance consistency was verified using the Pedcheck program [170], and in case of inconsistencies the entire family was 'zeroed out' for one marker. Hardy-Weinberg equilibrium (HWE) was investigated using the GENEPOP program for multi-allelic markers [119].

Table 5.1 Results of TDTs for single markers and 2-marker haplotypes

Haplotype	Allele[a]	Transmitted allele		Non transmitted allele	
		number	frequency	number	frequency
DRD4 VNTR	2	42	0.10	52	0.13
p=0.50	4	277	0.68	265	0.65
	7*	71	0.17	67	0.16
	Other	19	0.05	25	0.06
DRD4 Mono	2	103	0.23	105	0.24
P=0.71	4	274	0.62	282	0.64
	5	63	0.14	55	0.12
	Other	2	0.00	0	0.00
Haplotype DRD4	2_2	15	0.04	24	0.06
p=0.65	2_4	26	0.07	25	0.06
	4_4	206	0.52	204	0.51
	4_5	49	0.12	44	0.11
	7_2	54	0.14	58	0.15
	Other	46	0.12	41	0.10
DAT1 VNTR	9	96	0.22	92	0.21
P=0.73	10*	341	0.77	346	0.78
	Other	5	0.01	4	0.01
D5S2005	4	221	0.52	214	0.50
P=0.54	5	99	0.23	96	0.23
	6	60	0.14	55	0.13
	7	23	0.05	33	0.08
	Other	21	0.05	26	0.06
Haplotype DAT1	9_4	33	0.08	32	0.08
P=0.21	10_4	178	0.44	172	0.42
	10_5	75	0.18	79	0.19
	10_6	45	0.11	38	0.09
	10_7	11	0.03	25	0.06
	Other	66	0.16	62	0.15
DRD5	4	27	0.07	26	0.06
P=0.32	5	19	0.05	25	0.06
	8	34	0.08	25	0.06
	9*	183	0.45	186	0.46
	10	48	0.12	41	0.10
	11	29	0.07	38	0.09
	12	11	0.03	21	0.05
	Other	55	0.14	44	0.11

NOTE Overall *P* values of likelihood tests are shown in italic print below the marker names.

[a] Allele numbers of the VNTRs indicate repeat numbers; previously reported at-risk alleles are indicated with an asterisk (*); Two-marker haplotypes are indicated by the alleles of the respective markers, separated by an underscore (_); Other, combined numbers/ frequencies of all alleles and haplotypes with frequencies< 0.05.

Likelihood ratios for transmissions of marker alleles and haplotypes from parents to affected offspring, as well as linkage disequilibrium (LD) between markers, were calculated using the TDTPHASE program [130]. This estimates missing haplotypes using the Expectation Maximization (EM) algorithm, and uses unconditional logistic regression on the full likelihood of parents and offspring. Only alleles and haplotypes with frequencies higher than 0.05 were taken into account. *P* values were not corrected for multiple testing.

Results

All five markers were in HWE in parents as well as in children. Results of overall likelihood tests for single markers and marker haplotypes are shown in Table 5.1. There were no signs of distorted transmissions of any of the single polymorphisms, or of single alleles. As genetic markers, both VNTRs are not very informative, with heterozygosity values below 0.5, and this could be one explanation for the replication problems in previous studies. In order to increase the information content, we also analyzed two-marker haplotypes with polymorphic microsatellites for DAT1 and DRD4. LD between the DRD4 markers was substantial for the DRD4 polymorphisms (D'=0.59), but lower for the DAT1 polymorphisms (D'=0.20), which is in agreement with a previous study [171]. Overall, no haplotypes were significantly associated.

Discussion

In this Dutch family-based sample, which is one of the largest described so far, no association was found between the DAT1, DRD4 or DRD5 genes and ADHD. The DRD5 data presented in detail here were part of the recent multi-centre analysis for this gene, in which the Dutch sample was one of the two studies that did not contribute to the overall detected association [83]. These findings seem to be in agreement with the results of our recent whole-genome scan, in which there were no indications for linkage in the chromosomal regions that contain the three genes [165]. The power of the linkage study, however, may have been too low to detect small effects. Likewise, we cannot rule out that our negative findings, as well as those by others, are due to chance, given the low relative risks attributed to the individual genes, or due to insufficient LD with an unknown disease-related variant. It is also possible, however, that these genes do not play an equally important role in all populations, or in multiplex families as compared to sporadic cases [169]. In another recent study in a large sample of multiply affected families, the DAT1 and DRD4 VNTRs also did not show a distorted transmission, although a different DRD4 polymorphism, as well as the DRD5 microsatellite, were positively associated with ADHD [172]. Now that combined studies have quite convincingly suggested a small but significant role for dopaminergic genes in ADHD, further studies in different populations and different samples seem to be required to assess their role as general risk factors across populations.

Am J Med Genet B Neuropsychiatr Genet. (2005) 132(1):50-2 (reprinted with permission).

6 A WHOLE-GENOME SCAN IN 164 DUTCH SIB PAIRS WITH ATTENTION-DEFICIT HYPERACTIVITY DISORDER: SUGGESTIVE EVIDENCE FOR LINKAGE ON CHROMOSOMES 7P AND 15Q

S.C. Bakker*, E.M. van der Meulen*, J.K. Buitelaar, L.A. Sandkuijl†, D.L. Pauls, A.J. Monsuur, R. van 't Slot, R.B. Minderaa, W.B. Gunning, P.L. Pearson and R.J. Sinke

*Both authors contributed equally

†4 December 2002

Abstract

A genome scan was performed on 164 Dutch affected sib pairs (ASPs) with attention-deficit hyperactivity disorder (ADHD). All subjects were of Dutch Caucasian descent and phenotyped according to DSM-IV criteria. Initially, a narrow phenotype was defined in which all the sib pairs met the full ADHD criteria (117 ASPs). In a broad phenotype, additional sib pairs were included in which one child had an autistic spectrum disorder but also met the full ADHD criteria (164 ASPs). A set of 402 polymorphic microsatellite markers with an average intermarker distance of 10 cM was genotyped and analyzed using the Mapmaker/sibs program. Regions with multipoint maximum likelihood scores (MLS) > 1.5 in both phenotypes were fine mapped with additional markers.

This genome scan indicated several regions of interest, two of which with suggestive evidence for linkage. The most promising chromosome region was located at 15q, with an MLS of 3.54 under the broad phenotype definition. This region was previously implicated in reading disability and autism. In addition, MLSs of 3.04 and 2.05 were found on chromosome regions 7p and 9q in the narrow phenotype. Except for a region on chromosome 5, no overlap was found with regions mentioned in the only other independent genome scan in ADHD reported to date (Ogdie *et al*, submitted).

Introduction

Attention-deficit hyperactivity disorder (ADHD [MIM 143465]) is a highly heritable psychiatric disorder that affects about 4-5% of children in western countries [72]. The syndrome persists into adulthood in about one-third of the cases [173] and affects approximately 0.5-2% of adults [72].

According to the criteria of the Diagnostic and statistical manual of mental disorders, 4th edition (DSM-IV), children with ADHD should meet six out of nine criteria for inattention (inattentive subtype), or six out of nine criteria for hyperactivity/impulsivity (hyperactive/impulsive subtype), or both (combined subtype). Furthermore, the behavioural problems must have started before the age of 7 years and have persisted for at least 6 months in at least two different settings (e.g. at school and at home) [71].

There is a substantial gender difference in the prevalence of childhood ADHD, with boys being affected three times more often in the general population [174] and ten times more often in clinical settings [175]. In adults this ratio becomes 2:1 in the general population [176]. High levels of co-morbidity with other psychiatric disorders are common in both genders [73]. When viewed either categorically or on a continuous scale, ADHD has an estimated heritability of 75-91% [158]. The relative risk for siblings of affected children (λ_s) is increased approximately five-fold [177]. Based on the relative risk for first (λ=5-8) and second-degree family members (λ=2), a model of multiple genes interacting in an additive manner was proposed [78]. Despite much research, the exact aetiology of ADHD has still not been clarified.

Since methylphenidate, the most widely prescribed drug for treating ADHD, mainly blocks the dopamine transporter, most genetic research has focused on the dopaminergic system. Several studies have found an association with the 7-repeat allele of a 48 base pair repeat in the dopamine 4 receptor (DRD4) gene [178]. Similar results were reported for the 10-repeat allele of a 40 base pair repeat in the dopamine transporter (DAT1) gene [166, 169, 179-181]. A number of studies have failed to replicate these findings, possibly due to their small sample size or to the fact that ADHD is a genetically heterogeneous disorder [182-186]

The relative risk for developing ADHD for a carrier of a DAT1 or DRD4 risk allele is approximately 1.3 and 1.4 respectively [81, 82]. When the syndrome is viewed as the extreme of a continuous trait, these genes explain only 2-3% in symptom severity [181, 187]. Accordingly, it is assumed that there are other, as yet unidentified, genes that play a major role in ADHD. It is possible that some of these genes are involved in serotonergic and noradrenergic neurotransmission [188]. Whole-genome linkage analysis provides a means of identifying chromosome regions containing susceptibility genes, without an a priori hypothesis about their function. In addition, studies involving affected sib pairs (ASPs) require no assumptions about the mode of inheritance. Recently, the results of the first whole-genome screen in American affected sib pairs were published [189]. However, under a broad as well as a narrow definition of the ADHD phenotype, no chromosomal regions were found that were suggestive for linkage, according to recently suggested criteria for the interpretation of linkage results [190]. In a follow-up study in an extended sample, significant linkage was found on chromosome 16p13, in a region already implicated in autism [191]. It has been suggested that ADHD and autistic spectrum disorders, which also have a high heritability [192], have common genetic factors [191, 193]. Although in the DSM-IV classification an autistic spectrum disorder rules out the diagnosis of ADHD, a substantial number of ADHD patients have mild problems in social interactions and communication that are rather similar to the symptoms of autism [112, 193]. A subgroup of autistic children also showed high levels of inattention, hyperactivity and impulsivity [194-196].

Our study reports the results of a whole-genome scan in 164 Dutch Caucasian affected sib pairs with ADHD using a broad and a narrow phenotype. We detected three loci with MLSs of 3.54, 3.04 and 2.05 on chromosome regions 15q, 7p and 9q, respectively.

Materials and methods

Ascertainment. The current sample consists of 238 children in 106 families, who were clinically diagnosed according to DSM-IV criteria. The families were recruited from three academic child psychiatric outpatient clinics in Utrecht (N=24), Groningen (N=25) and Amsterdam (N=3). Other families (N=54) were recruited through a patient organization and by placing advertisements in journals and on the Internet. The Medical Ethical Committee of the Utrecht University Medical Centre approved the study and all parents gave written informed consent.

Instruments and procedures. The children and their parents were invited to participate in extensive diagnostic evaluations that lasted approximately one day. After evaluation, ASPs were included in the study if they met the following five criteria: (1) at least one member of each ASP met the full criteria for the DSM-IV defined combined, inattentive or hyperactive/impulsive type of ADHD; (2a) the other member of the ASP met the same criteria; or (2b) the other member met at least five out of nine DSM-IV criteria for inattention and/or five out of nine criteria for hyperactivity/impulsivity ('sub threshold' ADHD); or (2c) the other member met the full DSM-IV criteria for ADHD but had been diagnosed with an autistic spectrum disorder, which excludes an ADHD diagnosis; (3) both members were at least 3 years old, but not older than 18 (in our sample of 238 children, 5% were under 6 years of age); (4) for those children with a history of educational problems, only children with an estimated full-scale IQ above 80 on the Wechsler Intelligence Scale for Children-Revised (WISC-R) or the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) were included in the analysis [197, 198]; (5) All four grandparents were of Dutch Caucasian descent, with the exception of two families, each with two affected siblings, in which one of the parents was not of Caucasian descent.

Patients who suffered from handicaps (e.g. deafness), other psychiatric disorders (e.g. schizophrenia) or genetic syndromes (e.g. fragile X-syndrome) that could be related to behavioural problems were excluded from the analysis.

Relationships between siblings were verified using the program GRR [199]. This resulted in the identification of two half-sibs, who were also excluded. In two cases of families with monozygotic twins and additional affected siblings, one member of the twins was excluded.

In five families these twins comprised the only sib pair, which resulted in exclusion of the whole family.

Despite the fact that all the children in this analysis had been previously diagnosed with ADHD by child psychiatrists or paediatricians, the diagnosis was verified in clinical interviews with the parents and the children. In addition, for all the patients the DSM-IV version of the Diagnostic Interview Schedule for Children (DISC) [200] was conducted with both parents by trained graduate students in medicine or child psychology. This instrument was also used to measure the presence of mood disorders, anxiety disorders (other than simple phobias), oppositional defiant disorder (ODD) and conduct disorder (CD). Finally, the Conners and Childhood Behaviour Checklist/ Teacher Report Form were collected from teachers and parents [201, 202].

The final diagnosis of ADHD, which served as the primary basis for inclusion in the study, was determined using a 'best-estimate procedure' [203]. To this end, the results of the medical history, clinical interview, DISC interview, information about social skills, and the scores on the Conners and Childhood Behaviour Checklist/ Teacher Report Form, as rated by the parents and teachers, were summarized by EvdM into a patient report [200-202]. This report was discussed in regular case reviews with an experienced child psychiatrist (JKB). The resulting diagnosis was a consensus diagnosis.

If a child showed severe social deficits, the possibility of an autistic spectrum disorder was considered after collecting additional information on their language and social development, and their repertoire of activities and interests.

A narrow phenotype of probands and siblings with full ADHD symptoms according to the DSM-IV criteria was defined. The sample comprised 117 ASPs according to an unweighted analysis that calculated all possible pairs in families with more than two affected siblings. In the same manner, a broad phenotype was defined comprising the narrow phenotype plus an additional 47 ASPs with a broad phenotype (i.e. in which one member had full ADHD and the other met criteria 2b or 2c of the selection criteria). The broad and narrow samples are described in Table 6.1.

The mean age of the children was 10 years (SD 3). Their clinical characteristics are shown in Table 6.2. The highest educational level of either one of the parents defined the social economic status of the family unit. The following classification was used: I no or uncompleted high school; II high school completed; III some college education; IV college completed.

Children with an autistic spectrum disorder were divided into pervasive developmental disorder not otherwise specified (PDD-NOS) and autism/ Asperger syndrome, according to the DSM-IV criteria. PDD-NOS is generally viewed as a less severe form of autism. Children with Asperger syndrome have normal language skills, contrary to those with autism.

Table 6.1. Numbers of affected sib pairs by phenotype

	Broad phenotype			Narrow phenotype		
Children per family	Families	ASP[a]	Children[b]	Families	ASP[a]	Children[b]
2	85	85 (85)	170	62	62 (62)	124
3	17	51 (34)	51	9	27 (18)	27
4	3	18 (9)	12	3	18 (9)	12
5	1	10 (4)	5	1	10 (4)	5
Total	106	164 (132)	238	75	117 (93)	168

[a] Total number of affected sib pairs, calculated in two different ways. In the unweighted (all pairs) method, a family with n sibs contributes (n2-n)/2 sib pairs. The number of independent sib pairs, in which a family with n sibs contributes (n-1) pairs, is shown between brackets.

[b] Total number of children who make up the relevant sib pairs.

The DSM-IV criteria distinguish three subtypes of ADHD: inattentive, hyperactive/impulsive and combined [71]. In accordance with most other studies, the majority of children suffered from the combined subtype. In two families with minor and adult siblings with ADHD, the adult siblings were also included in the analysis.

Phenotypic characteristics of study samples were compared with a chi-squared test.

Genotyping and analysis. DNA was extracted from peripheral blood lymphocytes or buccal mucosa using established procedures. Samples obtained from buccal mucosa were purified with an additional phenol extraction step. DNA concentration was measured with a spectrophotometer (Tecan, Männedorf) and samples were diluted with distilled water to a concentration of 15 ng/µl.

The Mammalian Genotyping Service of the Marshfield Medical Research Foundation performed the genotyping for the initial screen. The marker set was based on Marshfield Screening set 10 and consisted of 402 polymorphic microsatellite markers with an average spacing of 10 cM and an average heterozygosity of 0.75.

In the second stage of the screen, chromosomal regions were fine mapped with additional microsatellite markers from the Marshfield database. Marker positions in these regions were verified using the Ensembl, Celera and Decode [204] human sequence databases.

Either the forward or the reverse oligonucleotide primer was labelled with 6-FAM, HEX, or NED fluorescent dyes (Biolegio, Malden, the Netherlands, and Applied Biosystems, Foster City, USA.). PCR reactions were performed in a 10 µl volume containing 30 ng template DNA, 25 ng of each oligonucleotide primer, 200 mM of each dNTP, and 0.4 units Amplitaq Gold (Applied Biosystems), in 1 × PCR buffer II with 2.5 mM $MgCl_2$ (Applied Biosystems). DNA was initially denatured at 94°C for 7 min and was then subjected to 33 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, followed by a final extension step of 30 min at 72°C. PCR products were diluted 4 times with distilled water and one microliter of diluted product was mixed with 4 µl HiDi (Applied Biosystems) and 0.1 µl GS-500 size standard

Table 6.2. Clinical characteristics of the individual children

	No. (%) of children with:	
Characteristic	Narrow phenotype	Broad Phenotype
Sex:		
Male	167 (83.9)	30 (76.9)
Female	32 (16.1)	9 (23.1)
Social economic status:		
I	0	0
II	50 (25.1)	3 (7.6)
III	73 (36.7)	18 (46.2)
IV	76 (38.2)	18 (46.2)
Co-morbidity[a]:		
Anxiety	27 (13.8)	10 (27.8)
Mood	15 (7.7)	2 (5.6)
Oppositional defiant disorder (ODD)	82 (41.8)	9 (25.0)
Conduct disorder (CD)	14 (7.1)	1 (2.8)
Broad phenotype:		
Sub threshold ADHD	…	13 (33.3)
PDD-NOS[b]	…	18 (46.2)
Autism/Asperger	…	8 (20.5)
ADHD subtype:		
Combined	170 (85.4)	…
Inattentive	25 (12.6)	…
Hyperactive/impulsive	4 (2.0)	…
Overall Total	199 (83.6)	39 (16.4)

NOTE: in contrast to Table 6.1, in which sib pairs are the point of reference, in this table the individual child is the point of reference. Thus, a child with ADHD who has an autistic sibling is now counted in the narrow phenotype. The total number of children in the narrow phenotype therefore differs from Table 6.1.

[a] In the narrow as well as in the broad phenotype, co-morbidity data from three children are missing.

[b] PDD-NOS, pervasive developmental disorder not otherwise specified.

(Applied Biosystems) and separated in POP6 polymer on an ABI 3700 capillary DNA sequencer (Applied Biosystems). Data were analyzed using Genescan 3.6 and Genotyper 3.5 for Windows NT (Applied Biosystems). Two independent raters genotyped the additional markers and checked results for inconsistency. If they disagreed, the genotypes were set to unknown. In the initial screen, genotypes with a reported Marshfield quality score below 0.99 were also recoded as unknown.

Inheritance within families was verified using the Pedcheck program [170]. If there were inheritance errors, the complete family was excluded from the analysis for that marker. Allele frequencies were calculated from the parental genotypes.

Data were analyzed using the Mapmaker/sibs program [205]. Maximum-likelihood-scores (MLS) were determined in single-point and multipoint analyses and calculated using the possible triangle method [206], which makes no assumptions about the mode of inheritance. Dominance variance was allowed for in all analyses.

In multipoint analysis, MLS were calculated at ten intervals between two adjacent markers, with off-end ranges of 10 cM at both ends of the chromosome. All possible pairs within each family were used in the calculations. In those regions with MLS > 2, data were also analyzed with a weighted procedure. According to this method, all possible sib pair combinations in families with more than two affected siblings were taken into account, but weighted with a factor (2/n), where n was the number of patients in the family. In the analysis of the X-chromosome the separate LOD scores for sister-sister, sister-brother and brother-brother pairs were summed [207].

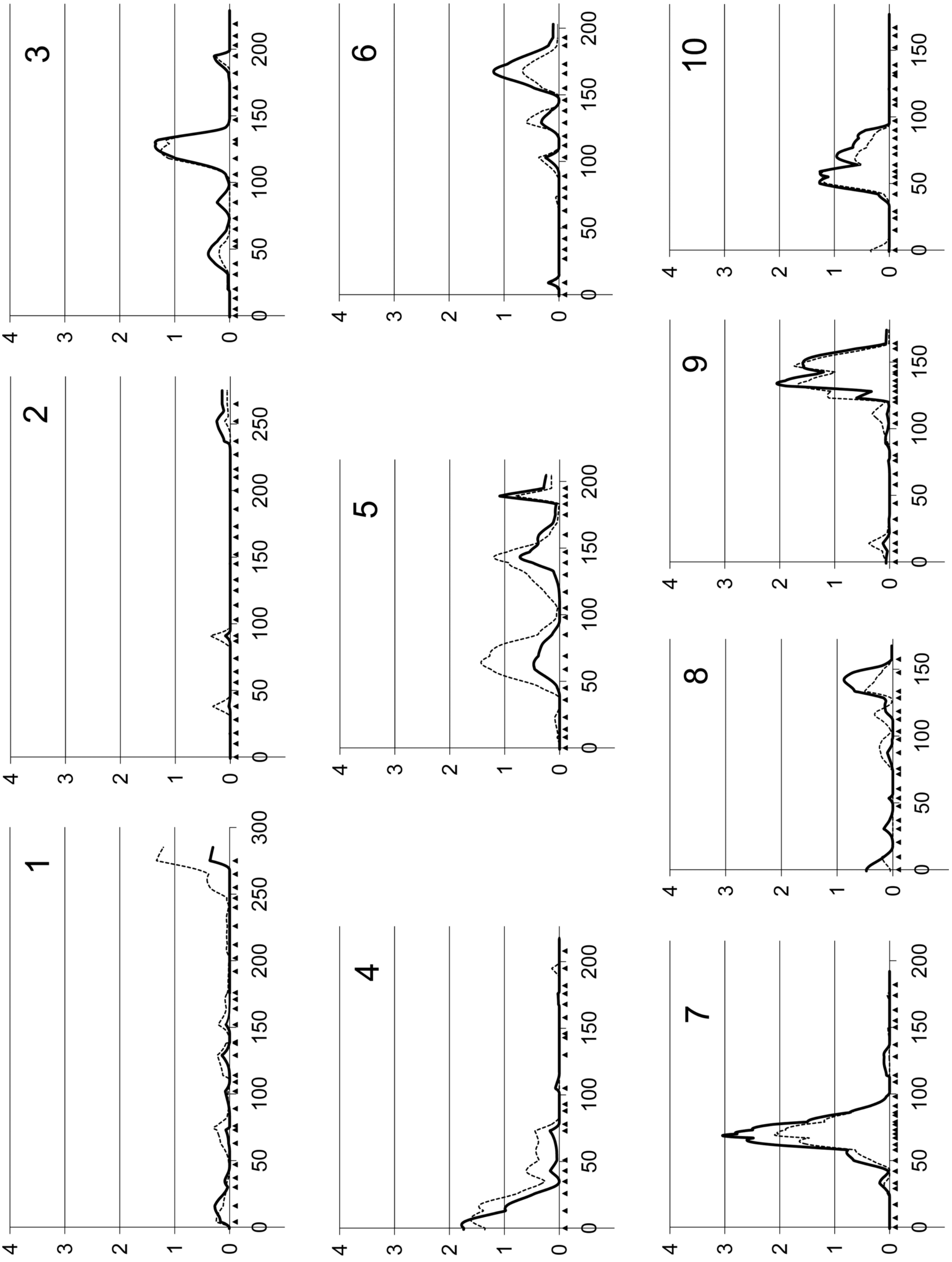
Ninety-five percent support intervals under linkage peaks were determined by taking the maximum MLS minus one, and determining the map position of this point. Locus specific (sib values were calculated by dividing the observed Z_0 value at the point of the maximal MLS by the expected value of 0.25 [205]. Overall information content of the markers was obtained by calculating the average multipoint information content for all markers, including the 10 cM off-end scales. Exclusion mapping was performed for different relative risks using the 'exclude' option in Mapmaker/sibs [205].

Results

DNA from all the parents was available, except for one missing father. In the initial screen all families were genotyped using 402 polymorphic microsatellite markers with an average inter-marker spacing of 10 cM. Average genotyping completeness was 97.3% and the average marker information content across the genome was 0.72. The most promising regions on chromosomes 7, 9, 10 and 15 were fine mapped with additional markers, while genotyping in these areas was also repeated for markers from the initial screen. This resulted in a resolution finer than 5 cM, and a marker information content of 90-95%. Multipoint MLS scores for the broad and narrow phenotype groups are shown in Figure 6.1.

Table 6.3 summarizes all the genomic regions with multipoint MLS higher than one. Single-point MLS scores were in agreement with multipoint results (data not shown).

On chromosome region 7p the MLS was 3.04, with a 95% support interval of 16 cM, or 26.6 Mb. The estimated sibling relative risk of this locus would be 1.19. The relative risk of a locus on chromosome 15 was estimated to be 1.60 (with a lower limit of 1.32 at the borders of the support interval). Here, the MLS was 3.54, while the 95% support interval measured 18 cM, or 16 Mb.


1
2
3
4
5
6
7
8
9
10
0
1
2
3
4
0
50
100
150
200
250
300

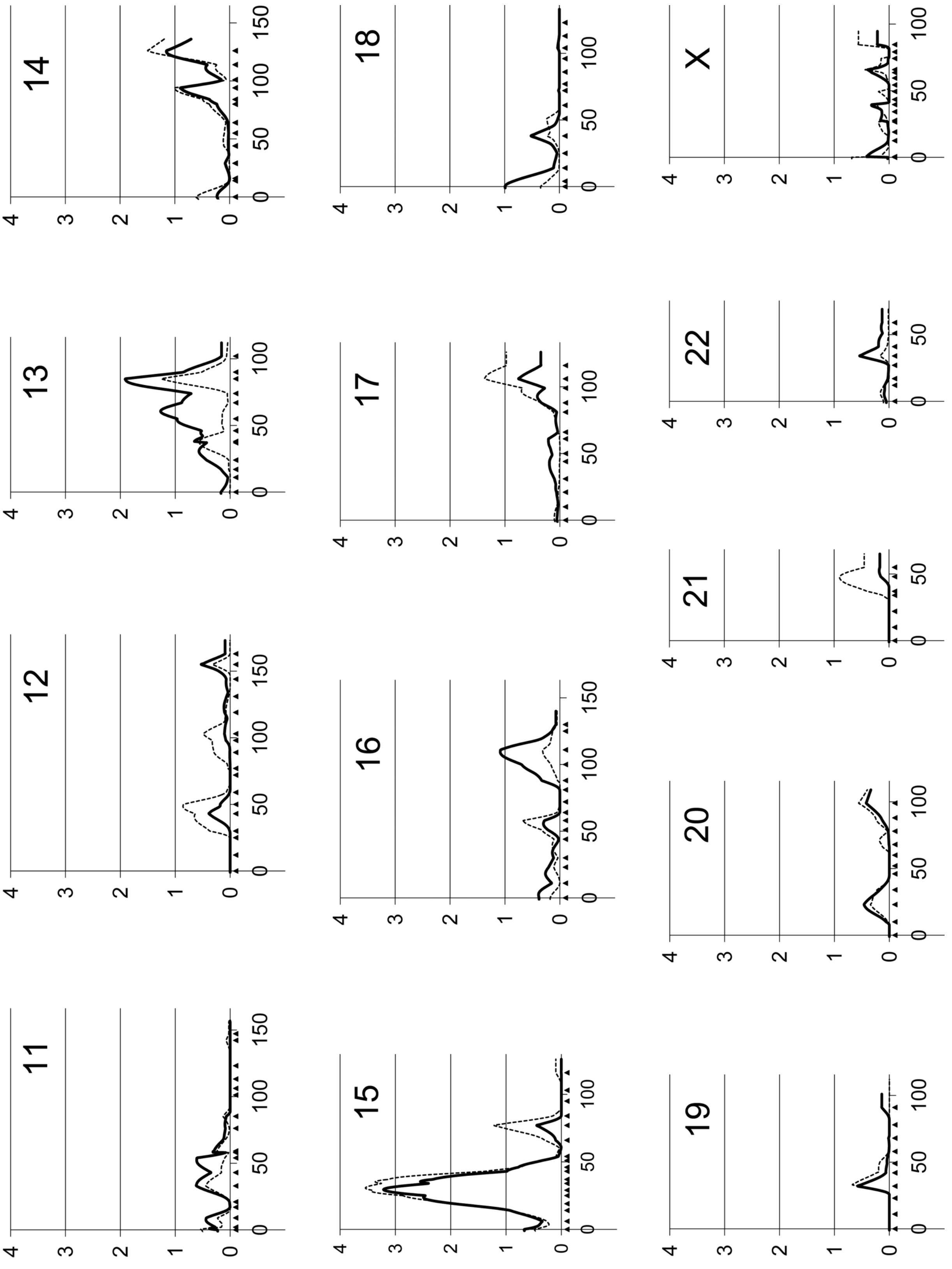


Figure 6.1. Multipoint MLS for all chromosomes (y-axis). MLS for the small phenotype are indicated with a solid line, while the dotted line represents MLS for the broad phenotype. Genetic distance is given in cM, with marker positions indicated by solid triangles (x-axis). Chromosome numbers are listed in the upper right corner of each graph.

Table 6.3. Chromosome regions with multipoint maximum likelihood scores > 1

			Multipoint MLS[a]	
Chromosome	**Marker**	**Location[b]**	Narrow Phenotype	Broad Phenotype
3q13.32	D3S2460	134.6	**1.36**	1.25
4p16.3	D4S3360	0	**1.78**	1.62
5p13.1	D5S2500	69.2	0.47	**1.43**
6q26	D6S305	166.4	**1.19**	0.66
7p13	D7S1818	69.6	**3.04**	2.09
9q33.3	D9S1825	136.5	**2.05**	1.68
10cen	D10S1426	59.0	**1.26**	1.25
13q33.3	D13S796	93.5	**1.91**	1.25
15q15.1	GATA50C03	34.8	3.21	**3.54**

[a] Multipoint maximum likelihood scores (MLSs) for both phenotypes are given, with the highest score printed in bold. The marker closest to the maximum MLS is shown. Only those multipoint MLSs are listed that were supported by multiple markers in single-point analysis.
[b] Locations in cM correspond with the Marshfield genetic map, sex-averaged distances.

Since 1/5 of the families in our sample had more than two affected siblings, we investigated the effect of weighting multiple sibships in the regions with MLS > 2. Following this procedure, the MLS on chromosomes 15q and 7p decreased to 2.49 and 2.27, respectively, while on chromosome 9q the MLS increased to 2.34. Exclusion mapping showed that the existence of a locus with a relative risk of 1.5 could only be excluded for 8.3% of the total genome. However, 53.9% and 87% of the genome could be excluded for loci with relative risk values of 2 and 3, respectively.

Discussion

Our genome scan in 164 affected sibling pairs with ADHD has identified regions on chromosomes 7p and 15q, with maximum MLS scores of 3.04 and 3.54 respectively, which can be regarded as suggestive for linkage according to the criteria proposed by Lander and Kruglyak [190]. The study was based on a narrow phenotype of sib pairs in which both children had ADHD according to the DSM-IV criteria. An additional broad phenotype was distinguished, in which the probands met the full ADHD criteria and the siblings had sub threshold ADHD or had been diagnosed with an autistic spectrum disorder and also met the full criteria for ADHD. The results for the broad and narrow phenotype groups were very similar. There were no unique areas of linkage restricted to one phenotype only. In most peak regions MLS were higher in the analysis of the narrow phenotype, even though the number of contributing sib pairs was reduced by almost 30%.

Following the method used in the first reported genome screen in ADHD [189], the results of this study were analyzed using all possible sib pairs in each family. Since pairs in families with more than two affected siblings are not independent, the resulting MLS may be inflated. On the other hand, assigning a lower weight to sib pairs from families with more than two affected siblings is probably too conservative [208]. Even in the weighted analysis the regions on chromosomes 15q and 7p can be classified as suggestive for linkage.

In their genome scan in an extended sample of 270 ASPs, Ogdie *et al.* identified regions with LOD scores > 1 on chromosome regions 5p13, 6q14, 11q25, 16p13, 17p11 and 20q13 (for details see Ogdie *et al.*, submitted). Comparing these data with our results, it is remarkable that none of these regions coincide with the regions with LOD scores > 1 in our study, with exception of the 5p region, that showed modest LOD scores in both studies. It is not yet clear to what extent these differences are due to spurious findings, or to clinical or genetic heterogeneity. Both groups are now investigating the possibility of performing a combined analysis of their samples. However, such a study will require a detailed comparison of marker sets as well as phenotypic characteristics.

Several differences between our sample and the one described by Ogdie *et al.* are apparent. Firstly, the Dutch sample was specifically selected for a homogeneous ethnic background. Families, with two exceptions, were included only if all four grandparents of the affected sib pairs were Caucasian and of Dutch descent. Since there was relatively little immigration to the Netherlands until the second half of the twentieth century, our sample is ethnically more homogeneous than the American sib pair sample, which included 20% children of non-Caucasian descent.

Other differences are the gender ratio (17% girls in the Dutch sample compared with 28% in the American sample, p=0.01), and the prevalence of the inattentive subtype of ADHD (13% in the Dutch sample versus 55% in the American sample, p<0.01).

Moreover, the Dutch sample is characterized by a lower level of Conduct Disorder (6% versus 15%, p<0.01) and a higher social economic status (0% versus 22% in the lowest social economic class, p<0.01). Social security and health care coverage levels in the Netherlands are higher, while inner city problems are less pronounced than in the US. It is therefore possible that there were different environmental factors affecting the two studies and, assuming there are at least several genes involved, different, as yet partly unknown, environmental components could have led to the identification of different gene sets. Genetic linkage studies in a wide variety of complex disorders have generally demonstrated a lack of replication [209], and comparison of the genome screens conducted in ADHD indicates that linkage studies in ADHD could face similar difficulties.

ADHD is not only associated with autistic features, but also with dyslexia and language disorders [112, 210-212]. It is interesting that in both genome screens performed thus far the chromosome regions of maximum linkage coincide with linkage regions found in autism studies. In our study, the highest MLS was obtained on chromosome 15q in the broad phenotype group, which also included siblings with an autistic spectrum disorder. In autism, the chromosome region 15q11-13 has repeatedly been suggested to harbour disease susceptibility genes. Autism was shown to co-segregate with different chromosomal abnormalities in this region [213-215] and several authors have reported evidence for linkage in this area [216-220]. The highest MLS in our study was located approximately 20 cM distal to the region most frequently reported, which falls just outside our 95% support interval. One study, however, found the highest MLS at marker D15S118, which lies exactly between the markers that gave the highest MLS in our study [218]. This sample of autistic children had a similar, mainly Western European ethnic background.

The results of studies on the genetics of reading disability are also relevant for our findings in the 15q region. Two genome scans reported a susceptibility locus for reading disability in areas that fall within the 95% support interval of our linkage peak on chromosome 15 [221, 222]. Moreover, reading disability was found to be associated with a three-marker haplotype that also included marker D15S994 [223]. This marker had the highest single point LOD score (3.37) in our study and contributed directly to the MLS peak. Interestingly, the findings in reading disability prompted an association study with markers in this region in ADHD families. Significant association was found with marker D15S146, which is located at 0.5 cM distance from D15S994 [224]. Given these findings, we plan to collect detailed information on the reading abilities of the participants in our sample, and to analyze the region on chromosome 15 in a further independent sample of families collected for genetic studies on reading disability.

The results presented in this study seem to lend further support to the hypothesis that a chromosome 15q locus plays a role in the aetiology of genetically overlapping developmental disorders including ADHD, autism and reading disability.

Most of the well-known ADHD candidate genes, including DAT1, DRD4 and DRD5, lay outside areas with elevated LOD scores, a point also raised by Fisher *et al.* [189]. It should be noted, however, that our study had limited power to detect loci with a low locus-specific relative risk. Recent meta-analyses suggest relative risks values for DAT1, DRD4 and DRD5 of 1.3, 1.4 and 1.6, respectively [81, 82]. Although our study had a 97% power of finding a LOD score of 2.6 for a locus with a relative risk of 2.0, the chances of finding a similar LOD score for loci with a relative risk of 1.6 and 1.3 would have been 63% and 11%, respectively. These loci would therefore have gone largely unnoticed.

Two regions in this study harbour previously suggested ADHD candidate genes, both involved in dopaminergic neurotransmission. Dopa decarboxylase (DDC), or Aromatic L-amino acid decarboxylase (AADC) is the enzyme that converts dopa into dopamine. Positron emission tomography (PET) showed an accumulation of [18F] fluorodopa in the right midbrain of children with ADHD, indicating an impaired function of DDC [225]. The gene is located on chromosome region 7p13, within 600 kilobases of the D7S2422 marker, which gave a MLS of 3.04 in our study. In an Irish population of ADHD patients, a haplotype consisting of the D7S2422 marker and a 4 base pair deletion in exon 1 was transmitted more often to affected offspring, although these results were only marginally significant [226]. In our study we found evidence of an increased marker sharing in the DDC region relative to the ADHD phenotype. Taken together, the involvement of DDC in the aetiology of ADHD deserves future investigation using fine mapping techniques to define possible regions of linkage disequilibrium.

Another enzyme in the dopamine pathway is Dopamine beta-hydroxylase (DBH), which converts dopamine to norepinephrine. The gene on chromosome region 9q34 is located within the MLS peak found in our study. Several studies have reported an association of restriction site polymorphisms in this gene with ADHD [169, 227, 228]. A relationship between low DBH activity and symptoms of hyperactivity was reported several years ago [229].

In conclusion, this whole-genome scan in ADHD has located several susceptibility loci, two of which - on chromosome regions 7p and 15q - are suggestive for linkage. The chromosome 15 region is particularly interesting since it has been implicated in autism and reading disability. These results may provide new directions in the search for specific genetic determinants of ADHD.

Acknowledgements

This study is dedicated to the memory of L.A. Sandkuijl (1953 - 2002), who died shortly after its completion. We thank him for his invaluable help as a statistician and a teacher.

We gratefully acknowledge the support of the Mammalian Genotyping Service of the Marshfield Medical Research Foundation. We would also like to thank C. Wijmenga for helpful discussions, J.L. Senior for critically reading the manuscript, and all the families for their willingness to participate in this study.

This project was supported by the Makaria Foundation, the UMC Utrecht's Genvlag program and Catharijne Foundation (the Netherlands). Dr. E. van der Meulen was supported by the Drie Lichten Foundation, the Ter Meulen Foundation and the Koningsheide Foundation (the Netherlands) and the Yale Child Study Centre and the Harvard Unit of Neurodevelopment (USA).

Am J Hum Genet (2003) 72(5): 1251-60 (reprinted with permission).

7 ASSOCIATION ANALYSIS OF DOPA DECARBOXYLASE, A FUNCTIONAL AND POSITIONAL CANDIDATE GENE FOR ATTENTION-DEFICIT HYPERACTIVITY DISORDER

S.C. Bakker, K. Kusters, E.M. van der Meulen, B.P.C. Koeleman, W. Verduijn, J.K. Buitelaar, P.L. Pearson and R.J. Sinke

Abstract

Dopa decarboxylase (DDC) is a candidate gene for attention-deficit hyperactivity disorder (ADHD) and other psychiatric disorders, since this enzyme mediates the final step in the synthesis of dopamine. Association of markers close to the DDC gene with ADHD was recently reported. In addition, the gene is located in a region on chromosome 7p, in which we recently detected suggestive evidence for linkage with a multipoint LOD score of 3.04. In the same 106 sib pair families, we have genotyped 5 single nucleotide polymorphisms, a 4 base pair insertion/deletion polymorphism and two microsatellite markers spanning DDC and its promoter region. Patterns of linkage disequilibrium throughout the gene were investigated and marker haplotypes were constructed for association analysis. In addition, the five SNPs were genotyped in a case-control study, involving one affected child from each family and 253 unrelated controls. All markers were in strong LD, but no evidence for association of single markers or haplotypes with ADHD was found. Therefore, DDC is not likely to play a major role in ADHD in Dutch families.

Introduction

Attention-deficit hyperactivity disorder (ADHD) is a childhood psychiatric disorder characterized by inattention and hyperactivity. Family studies have indicated a substantial hereditary susceptibility to the disease (approximately 80%) [77], but the responsible genes have not been identified to date. Genes involved in dopamine neurotransmission have received much attention, since methylphenidate, the most widely used drug to treat ADHD, is known to block the dopamine transporter (DAT1) [159]. The net effect is an increased availability of dopamine in the synaptic cleft, due to reduced re-uptake by the presynaptic neuron. This observation makes all genes involved in dopamine synthesis, breakdown and neurotransmission good candidate genes in ADHD. DAT1 as well as the dopamine receptor 4 (DRD4) and 5 (DRD5) have been studied in numerous association studies, and meta-analyses of these studies showed a slightly increased risk for developing ADHD in subjects who possess the risk alleles [81-83]. These genes, however, only explain a small proportion of the total variance of the ADHD phenotype. This indicates that other, as yet unknown genes are involved.

Dopa decarboxylase (DDC), also known as aromatic L-amino acid decarboxylase (AADC) is an enzyme that mediates the final step in the synthesis of dopamine and several other neurotransmitters. It is encoded by a gene with 15 exons, that span 108 kilobases (kb). Neuronal and non-neuronal transcripts have been described [230, 231], and the gene possesses tissue-specific promoters [232-235]. Although it is located in a region with reported imprinting, the expression of DDC is probably bi-allelic [236]. In ADHD, an association was reported with a haplotype of a 4-basepair (bp) insertion/deletion (in/del) polymorphism, located in the untranslated region of exon one, and microsatellite marker D7S2422, which is located at 0.5 megabase (Mb) distance from the gene [226]. Interestingly, in a recent genome scan in 164 sib pairs with ADHD, we found the gene to be located in a linkage region on chromosome arm 7p, with a maximum multipoint LOD score of 3.04 [165]. The marker with the highest single point LOD score in our ADHD genome scan was the previously mentioned microsatellite D7S2422.

The fact that DDC is now both a functional and a positional candidate gene for ADHD led us to investigate the possible involvement of the gene in more detail. We therefore genotyped the distant marker D7S2422, the 4-bp in/del, and 6 other markers spanning DDC and its promoter region in both family-based and case-control association studies.

Materials and methods

Families and control samples. For a detailed description of inclusion criteria and the family sample, see Bakker and Van der Meulen *et al.* [165]. Briefly, the total sample consisted of 106 families with 238 affected children (a broad phenotype that included autistic spectrum disorders), 199 of which with full ADHD symptoms according to the DSM-IV criteria (narrow phenotype). The control panel consisted of 253 DNA samples from random Dutch individuals, obtained from the Immunogenetics and Transplantation Immunology Section of the Department of Immunohaematology and Blood Transfusion, LUMC. The Medical Ethical Committee of the UMC Utrecht approved the study and all subjects gave written informed consent.

Sample preparation. DNA was isolated as described elsewhere [165]. Each 96-well DNA plate contained 4 CEPH reference samples and 5 random duplicates.

Markers. Primer sequences for marker D7S2422 were obtained from the Genome database. The physical location of the marker was verified in the Ensembl human sequence database. A new intragenic dinucleotide repeat (DDC_intra) was found using the Tandem Repeat Finder program [151], for which the following primers were designed using the Primer3 program [152]: TAGCTTATTGCTAGGATATTAGG and CTTTCCCAGCTATCTCTCTC.

Either the forward or the reverse oligonucleotide primer was labelled with 6-FAM or HEX fluorescent dyes (Biolegio, Malden, the Netherlands). SNP probes were obtained from the Assay-on-Demand service (Applied Biosystems, Foster City, USA).

PCR conditions. The microsatellites and the 4-bp in/del were analyzed as previously described [165]. SNPs were analyzed on a 7900HT TaqMan system, according to the recommendations of the manufacturer (Applied Biosystems) and genotyped with the SDS package (Applied Biosystems).

Analysis. Two independent raters verified genotypes and checked results for inconsistency. If they disagreed, the genotypes were set to unknown. Hardy-Weinberg equilibrium for bi-allelic markers was calculated using a chi-square test with one degree of freedom. For the microsatellite markers the GENEPOP program was used [119]. Inheritance within families was verified using the Pedcheck program [170]. If there were inheritance errors, the complete family was excluded from the analysis for that marker. The presence of hidden population stratification was investigated in the parents (n=211) with the Structure program [30], using genotype data from thirty-six random microsatellites on different chromosome arms that were previously genotyped in this sample for a whole-genome scan [165]. Two, three or four population strata were modelled using a burn-in of 10,000 repeats, followed by 1,000,000 repeats, assuming admixture and correlated allele frequencies among populations. Results of three independent runs were compared.

Linkage disequilibrium between markers was calculated from family data with the GOLD program [155], using founders only.

Preferential transmission of alleles and haplotypes in families was tested with the TRANSMIT program. Case-control studies involved one random child from each family; association tests were performed with UNPHASED software [130]. Marker alleles and haplotypes with frequencies <5% were grouped.

Results

All markers were in Hardy-Weinberg equilibrium and duplicate genotypes were in agreement. As displayed in Table 7.1, there was significant linkage disequilibrium between all marker pairs, except for marker D7S2422, located at 0.5 Mb from DDC, which was not in significant LD with any of the other markers.

For single markers, family-based association tests did not show evidence for preferential transmission of marker alleles to affected children (Table 7.2). No haplotypes of two and three adjacent markers were significantly associated. Haplotypes of D7S2422 and the 4-bp in/del polymorphisms, which have previously shown association to ADHD, were also not significantly associated in our sample.

Table 7.1. Pair wise linkage disequilibrium (D') for the narrow phenotype group

	C11998157	C8320346	DDC_di	C1333607	C1333592	4bp_del	C1333584
C11998157							
C8320346	0.93						
DDC_di	0.67	0.70					
C1333607	0.92	0.86	0.80				
C1333592	0.91	0.75	0.76	0.94			
4bp_del	0.85	0.59	0.64	0.72	0.95		
C1333584	0.79	0.74	0.71	0.80	0.92	0.91	
D7S2422	0.08	0.15	0.14	0.10	0.09	0.06	0.08

The five SNP markers were also tested for association in a case-control study. The group of cases consisted of one affected child from each family, which was tested against 253 unrelated controls. No association with any of the markers was found, nor with 2- and 3-marker haplotypes.

Table 7.2. Results of family-based and case-control tests for single markers

Marker	Allele	T[a]	NT[a]	*P* value	Cases[b]	Controls[b]	*P* value
C11998157	1	0.16	0.19	0.40	0.16	0.18	0.58
C8320346	1	0.39	0.42	0.44	0.35	0.42	0.14
DDC_di	All			0.41	n.t.	n.t.	n.t.
C1333607	1	0.53	0.53	0.94	0.49	0.52	0.58
C1333592	1	0.51	0.47	0.33	0.46	0.50	0.42
4bp_del	1	0.20	0.26	0.08	n.t.	n.t.	n.t.
C1333584	1	0.63	0.63	1.00	0.60	0.64	0.37
D7S2422	All			0.97	n.t.	n.t.	n.t.

[a] Frequency of transmitted and non-transmitted marker alleles in TRANSMIT analyses.
[b] Frequency in cases and controls; n.t.: not tested.

Discussion

Dopa decarboxylase (DDC) is a plausible functional candidate gene for ADHD, since it is involved in the synthesis of dopamine. In addition, suggestive evidence for linkage was recently detected in the chromosomal region that contains DDC.

Therefore, we have investigated DDC for association with the disorder, in a both family-based and a case-control study. Our aim was to cover the gene with markers that were in LD with at least one of the other markers, in order increase the chances to detect association with an unknown disease locus, and to form marker haplotypes. We also included marker D7S2422, since this marker was reported to be part of an ADHD-associated haplotype [226]. Except for this marker, which is located at 0.5 Mb distance from the gene, all two-marker pairs throughout the gene were in significant LD. Based on our data, no recombination hot spots seem to be present in the gene.

No association was detected with single markers, or with marker haplotypes. The previously reported association of a haplotype of the 4-bp in/del polymorphism and marker D7S2422 could not be replicated in our sample, which was considerably larger than the sample of the original finding [226]. One would probably not have expected useful haplotypes of these markers to be present, since there was no significant LD between them in our sample. Our results seem to indicate that DDC does not play a significant role in the aetiology of ADHD in our population, but this conclusion should be drawn with caution. Recently published associations of dopaminergic genes with ADHD only became apparent in meta-analyses of combined studies. The estimated relative risks of carrying the at-risk alleles of the DAT1, DRD4 and DRD5 genes were modest, ranging from 1.3 to 2.0, and the frequent non-replication in single studies may have been the result of insufficient power. The estimated relative risk for DDC based on our genome scan data was less than 1.5, and weak associations may therefore have been missed due to insufficient power. Power calculations assume the presence of only one risk allele or haplotype. In case of extensive allelic heterogeneity, detecting association with specific alleles in the region of linkage would be almost impossible. In conclusion, our results suggest that if the previously reported linkage on chromosome 7 is real, DDC is not likely to be the causal gene in the region. Before large-scale fine mapping studies to find alternative genes are started, however, independent replication of the linkage findings seems necessary.

Acknowledgements

This project was supported by the UMC Utrecht's Genvlag program, ABC Neurogenomics, the Makaria Foundation and De Hersenstichting.

SECTION III

NEW GENOTYPING METHODS

8 DIFFERENCES IN STUTTER INTENSITIES BETWEEN MICROSATELLITES ARE RELATED TO LENGTH AND SEQUENCE OF THE REPEAT

S.C. Bakker, R.J. Sinke and P.L. Pearson

Abstract

PCR-induced stutter artefacts can complicate the analysis of microsatellite markers in genetic and forensic studies. Knowledge about the factors underlying differences in stutter intensity between different markers may allow correction of these artefacts, but at present this knowledge is incomplete. We measured the height of stutter peaks in existing genome screen genotypes from 28 di- and 10 tetranucleotide repeat markers, and investigated whether differences in stutter intensity were related to the absolute length and sequence of the repeat. Dinucleotide markers showed higher stutter peaks than tetranucleotides. Average stutter height in the dinucleotide markers correlated well with the absolute length of the uninterrupted CA repeat. Additional CT repeats increased stutter, while an interruption of the repeat resulted in lower stutter. Extrapolation of the data suggested that a minimum length of 5-6 repeat units is required for stutter to occur. Results from existing genotypes were confirmed by amplifying newly constructed PCR templates into which different repeats were introduced. Together, our findings indicate that, under similar PCR conditions, stutter intensity is mainly determined by the length and sequence of the repeat. However, the occurrence of unnoticed mutations in the repeat may hinder the accurate correction of stutter. Finally, it is suggested that stutter height can be used as a measure for microsatellite mutation rate in studies of microsatellite evolution and microsatellite instability (MSI).

Introduction

Microsatellites or short tandem repeats (STRs) are repeated DNA sequences with a motif length of one to five nucleotides. These repeats are widely used as markers in genetic studies, since they are present in high numbers throughout the genome and often show large variation in length between different individuals, making them highly polymorphic [14]. Microsatellite length can be measured using size separation by electrophoresis, after DNA amplification with the polymerase chain reaction (PCR). After electrophoresis of fluorescently labelled PCR products, peaks are visible with a length of one or several repeat units less than the original alleles (see Figure 8.1).

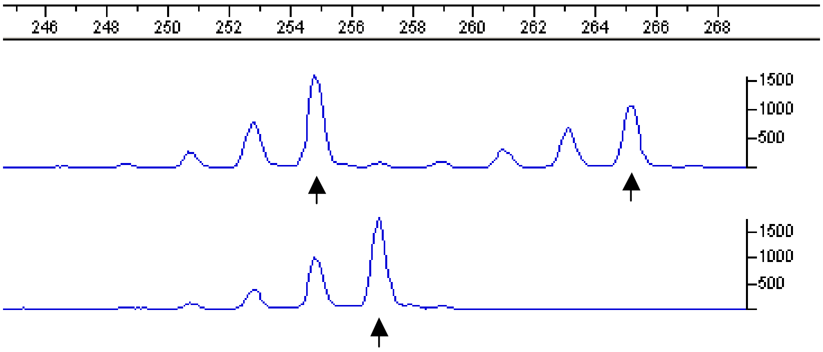


Figure 8.1. Typical stutter patterns for a dinucleotide marker, as shown in the Genotyper program. The upper graph represents the genotype of an individual heterozygous for the 255 and 265-bp alleles (arrows). The genotype in the lower graph is from a subject that is homozygous for the 257 base pair allele (arrow). All peaks not indicated by arrows are the result of stutter.

These so-called 'stutter artefacts' can seriously hinder the automated allele calling in genetic studies, since stutter peaks in heterozygous persons tend to overlap with the 'true' allelic peaks. The problem becomes worse when DNA from several people has mixed, for example in forensic identification of persons, or after deliberate pooling of DNA from large numbers of individuals for efficient genetic case-control studies. When alleles of different lengths are present in a DNA pool, stutter peaks will contribute to neighbouring peaks, which makes it impossible to directly derive allele frequencies from the electrophoresis pattern. It has been shown that accurate peak height correction for stutter is possible if the stutter characteristics for each allele are known [99]. However, to obtain marker specific stutter correction models, additional genotyping of individual persons is necessary, which decreases the efficiency of DNA pooling studies. If general rules determining stutter height in individual markers could be identified, this could allow correction of genotyping patterns for this artefact without additional individual genotyping.

Knowledge about the factors that determine stutter height is limited at present. The pattern of stutter peaks can be modelled by assuming a certain chance of losing (and a much smaller chance of gaining) one repeat unit during each PCR cycle [237]. Sequence analysis showed that PCR products isolated from stutter bands lacked one or more repeat units, and that these shorter fragments were formed during PCR [238-240]. Stutter for a particular marker was highly reproducible [241], but increased with the number of PCR cycles [242, 243]. Dinucleotide repeats were reported to have higher stutter than tri- and tetranucleotide repeats [99]. Although some investigators found constant stutter height for different alleles of the same marker [237], others reported stutter to increase with the measured allele length [90, 99, 244]. Observations in single tetranucleotide markers suggested that stutter intensity correlates with the length of the uninterrupted repeat [245-247]. A recent study confirmed that stutter height increased with increasing repeat length, when A and CA repeats inserted into vectors were amplified [248]. The same study suggested that a minimum length of 5 CA repeats is required before stutter can occur [248].

Stutter is thought to be the result of errors in the replication of the repeat, probably due to extension of a 'slipped' strand by the Taq polymerase. For a review of proposed mutation mechanisms see Kunkel *et al.* [249].

We have investigated if the differences in stutter height between dinucleotide markers could be related to the length and sequence of the microsatellite repeat. Stutter height was determined for 38 different microsatellite markers in existing genotyping data from a diabetes genome screen, that was previously performed at our department [250]. Using dinucleotide marker sequences obtained from the public CEPH/Généthon database, stutter height was related to length and sequence of the repeat, and the presence of additional repeats in the PCR product.

In addition, the findings from these existing data were verified by amplifying newly synthesized PCR templates with dinucleotide repeats of different lengths and sequence. This was done by a novel procedure using primers that included different repeats, as well as a tail sequence to which a second labelled primer could anneal.

The possibilities of using repeat sequence information to develop general stutter correction models for DNA pooling data will be discussed, as well as the potential use of stutter height as an indicator for microsatellite instability in diagnostics and population genetics.

Materials and methods

Marker selection. Genotype data were obtained from a diabetes genome screen previously performed at our department. From this set of tetranucleotide and dinucleotide markers, which was based on Marshfield screening set 6, ten tetranucleotide markers were chosen at random. Of the dinucleotide markers used in this screen, 105 were present in the CEPH/Généthon database. Three groups were selected on the basis of repeat sequence characteristics, without prior knowledge of stutter height: 18 markers with an uninterrupted CA repeat surrounded by non-repetitive sequence, 6 markers with an interruption of the CA repeat and 4 markers with an additional CT repeat in the sequence surrounding the CA repeat (see Table 8.1).

Sequences containing unknown or polymorphic nucleotides within the repeat sequence (indicated by N) were discarded. All markers used in the genome scan were genotyped under similar conditions (for details of PCR and analysis see Van Tilburg *et al.* [250]).

Generation of PCR products with different repeats. Primers rs909706F en rs909706R ('Repeat primers') were designed to amplify a single nucleotide polymorphism, and the product does not include a microsatellite repeat.

Table 8.1. Characteristics of stutter and repeat sequence in 38 different microsatellite markers

Repeat type	Marker	# CA[a]	Stutter[b]	Increase[c]	CEPH 134702[d]
CA only	D10S189	11.9	0.27	3.1	$(CA)_{12}$
	D9S1853	17.5	0.32	2.8	$(CA)_{15}$
	D2S2166	16.3	0.33	3.5	$(CA)_{20}$
	D10S579	14.6	0.34	3.4	$(CA)_{12}$
	D18S465	17.7	0.35	3.3	$(CA)_{13}$
	D3S1311	17.9	0.39	3.1	$(CA)_{21}$
	D17S1856	18.6	0.42	3.0	$(CA)_{22}$
	D4S403	18.4	0.44	3.9	$(CA)_{20}$
	D6S1654	17.1	0.46	3.3	$(CA)_{17}$
	DXS1214	19.0	0.49	4.5	$(CA)_{19}$
	D2S338	21.5	0.52	3.1	$(CA)_{16}$
	D11S4131	22.0	0.56	2.4	$(CA)_{24}$
	D6S273	19.6	0.59	4.5	$(CA)_{20}$
	DXS1068	20.5	0.60	4.5	$(CA)_{19}$
	DXS1003	27.6	0.68	2.7	$(CA)_{29}$
	D7S550	23.3	0.68	3.1	$(CA)_{25}$
	D2S2242	22.3	0.69	3.6	$(CA)_{23}$
	D3S1594	n.a.	0.89	3.0	$(CA)_{25}$
Interrupted CA	D13S164	8.2	0.16	2.0	$(CA)_{13}CTA(CA)_{3}CTA(CA)_{4}$
	D17S1843	15.9	0.29	2.0	$(CA)_{4}GA(CA)_{15}$
	D1S2612	17.1	0.43	3.3	$(CA)_{6}TA(CA)_{14}$
	D16S422	18.0	0.45	n.a.	$(CA)_{19}TATA(CA)_{6}$
	D13S1250	17.0	0.47	3.0	$(CA)_{15}TGCG(CA)_{2}CG(CA)_{7}$
	D5S1981	20.1	0.51	5.3	$(CA)_{18}GAGA(CA)_{2}TACA$
CA plus CT	D4S408	11.8	0.32	2.1	$(CT)_{14}(CA)_{17}TA(CA)_{6}$
	D14S70	15.2	0.47	3.8	$(CT)_{7}TT(CT)_{7}(CA)_{14}$
	D17S947	24.0	0.83	1.9	$(CA)_{20}(CT)_{7}CC(CT)_{13}$
	D14S77	17.9	0.85	2.6	$(GTCT)_{6}(CT)_{15}(CA)_{16}(CT)_{14}$
	average	18.2	0.49		
	st.dev	4.0	0.18		
Tetranucleotide	D3S1766		0.06		
	D7S1799		0.09		
	D6S1270		0.11		
	D12S372		0.14		
	D4S2417		0.06		
	D8S1132		0.13		
	D6S1007		0.08		
	D6S1056		0.10		
	D19S400		0.10		
	GAAT1A4		0.05		
	average		0.09		
	st.dev		0.03		

[a] Average number of uninterrupted CA repeats of all observations for a marker.

[b] Ratio of the highest stutter peak and the allelic peak.

[c] Percentage increase with each additional repeat unit of the highest stutter peak relative to the allelic peak.

[d] Sequence of CEPH individual 134702, obtained from the CEPH/Généthon database.

Table 8.2. Primers for repeat amplification experiments

Primer	Primer sequence
rs909706_CA10	TGGTAAAACGACGCCGA(CA)$_{11}$GTCAGTTTCCAAGGGGTTCTAACT
rs909706_CA15	TGGTAAAACGACGCCGA(CA)$_{16}$GTCAGTTTCCAAGGGGTTCTAACT
rs909706_CA20	TGGTAAAACGACGCCGA(CA)$_{21}$GTCAGTTTCCAAGGGGTTCTAACT
rs909706_CA30	TGGTAAAACGACGCCGA(CA)$_{31}$GTCAGTTTCCAAGGGGTTCTAACT
rs909706_CA20_T	TGGTAAAACGACGCCGA(CA)$_{11}$TA(CA)$_{9}$GTCAGTTTCCAAGGGGTTCTAACT
Reverse primer	ACAAGAGCCCATCTTGTAGTTA
M13HEX[a]	* TGGTAAAACGACGCCGACA

[a] In the M13HEX primer an asterisk (*) indicates the fluorescent HEX label.

NOTE: the effective number of CA repeats is increased by one in each repeat primer, since the M13-derived tail sequence adds one CA to the repeat.

One primer was ordered with different repeat sequences added to the 5' end, which were further extended with a 19 base pair (bp) tail that was derived from the M13-21 sequence. A third primer consisted of the same 19-bp tail sequence, which was labelled with the HEX fluorescent dye ('M13HEX'). Primer details are shown in Table 8.2.The principle of product labelling during PCR, which allows the use of less expensive unlabelled primers, has been described before [251]. During the initial cycles of the PCR the forward primer and the reverse primer will form an unlabelled product that includes the repeat sequence. This product will be amplified, again resulting in an unlabelled complementary product, to which the forward primer, but also the M13HEX primer can anneal. Since the forward primer is present in a much lower concentration, relatively more of the HEX labelled product will form, which can then be detected.

PCR conditions. Samples were amplified simultaneously on an ABI 9600 (Applied Biosystems, Foster City, USA). PCR reactions were performed in a 10 μl volume containing 25 ng of template DNA, 0.25 ng Repeat primer, 25 ng Reverse primer, 25 ng M13HEX primer, 200 mM of each dNTP, and 0.4 units Amplitaq Gold (Applied Biosystems, Foster City, USA), in 1 × PCR buffer II with 2.5 mM $MgCl_2$ (Applied Biosystems, Foster City, USA). DNA was denatured at 94°C for 7 min and was then subjected to 27 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s, followed by a final extension step of 30 min at 72°C. PCR products were diluted 4 times with distilled water. One microliter of diluted product was then mixed with 4 μl HiDi (Applied Biosystems, Foster City, USA) and 0.1 μl GS-500 size standard (Applied Biosystems, Foster City, USA). After 1 min denaturation at 96°C samples were scanned on an ABI 3700 capillary DNA sequencer (Applied Biosystems, Foster City, USA).

Analysis. Sample files were imported into Genotyper 3.5 for Windows NT (Applied Biosystems). For each allele, the allelic peaks as well as the stutter peak immediately preceding the allelic peak were labelled with allele length and peak height. Genotypes with stutter peak heights of less than 50 or with any peaks higher than 6000 were not labelled. The shorter allele of heterozygous individuals was not included unless the two allelic peaks were clearly non-overlapping, i.e. separated by more repeat lengths than the highest number of observable stutter peaks for that particular marker. Stutter height for each allele was expressed as the ratio of the heights of the labelled stutter peak and the allelic peak. Average stutter height for single alleles was determined from at least three observations, while average stutter height for all alleles of a marker was calculated from 20 or more stutter peaks.

Determination of CA repeat length. Sequences of marker alleles for CEPH individual 134702 were obtained from the CEPH/Généthon database. Since this individual was also typed in all our analyses allele lengths could be compared. Absolute length of the allele (A) was defined as the length of the sequence between the 5' end of the forward and reverse primers. The length of the CA repeat (CA) was defined as the uninterrupted occurrence of alternating C and A nucleotides (starting with either C or A). For all labelled genotypes of a marker the absolute length of the repeat sequence in a given individual was calculated as follows:

$CA_{individual} = CA_{134702} + (A_{134702} - A_{individual})$

in which CA_{134702} is the length in bp of the CA repeat for CEPH individual 134702, obtained from the in the CEPH/Généthon database, and A_{134702} and $A_{individual}$ are the respective allele lengths in bp of CEPH individual 134702 and a given individual, as measured in the genome screen data.

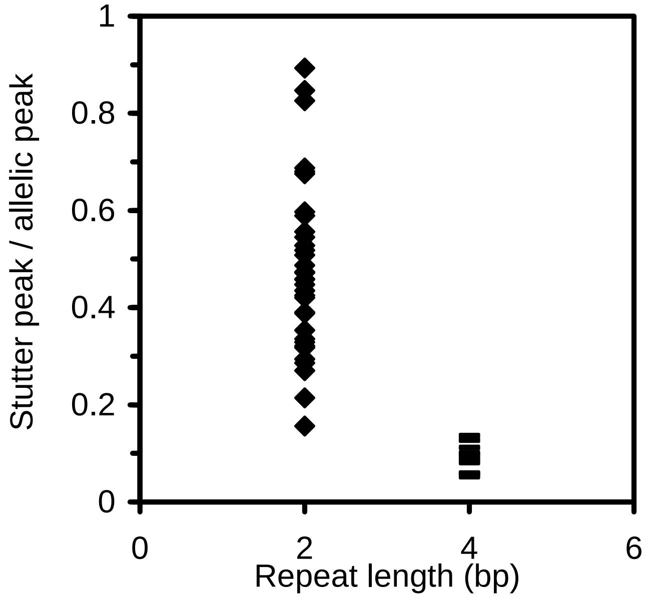


Figure 8.2. Average stutter height for 28 dinucleotide markers and 10 tetranucleotide markers. Each data point represents the average stutter height of multiple genotypes for a single marker.

Results

In 28 dinucleotide markers and 10 tetranucleotide markers, the stutter ratio was calculated by dividing the height of the first stutter peak (the peak that is one repeat unit shorter than the allelic peak) by the height of the allelic peak. Figure 8.2 shows that tetranucleotide markers had lower stutter ratios (average 0.09 +/- 0.03) than dinucleotide markers (average 0.49 +/- 0.18), and that average stutter between dinucleotides varied from 0.16 to 0.89 of the height of the allelic peak.

For each marker separately, the stutter ratio was plotted against the relative length of the alleles. A representative example is shown in Figure 8.3. In general, average stutter height correlated well with the length of the allele (correlation coefficients for all markers 0.38-0.94, average = 0.75). In tetranucleotide markers stutter increased 0.8% - 2.4% with each additional repeat unit, compared with 2.0% - 5.2% (average 3.4 +/- 0.8) in dinucleotides.

The only exception was marker D10S189, with stutter peaks that were almost half of the expected value for the 185-bp allele, of which seven copies from unrelated persons were present (see Figure 8.4). Three persons that were homozygous for the 185-bp allele were sequenced, as well as three control individuals that were homozygous for alleles with a different length. The repeat of all three individuals with the 185-bp allele was interrupted by a C to T transition, which was not present in the control individuals, or in the sequence of CEPH individual 134702 (see Table 8.3).

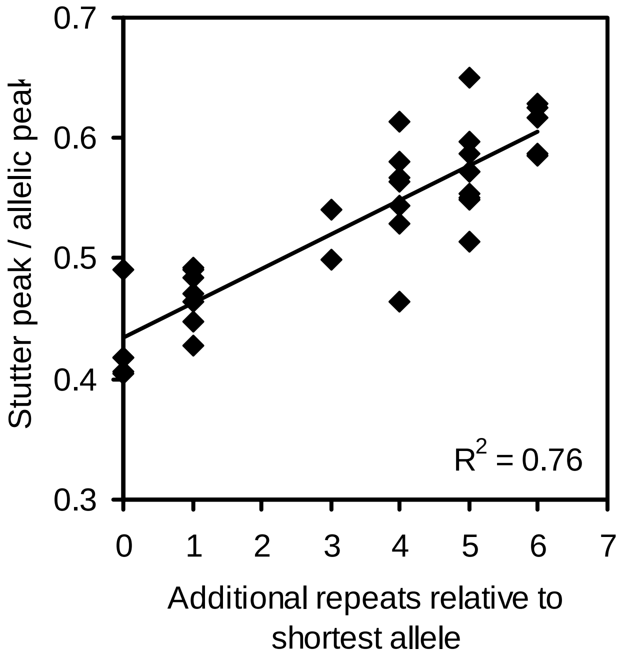


Figure 8.3. Representative example of the relation between stutter height and relative allele size in a dinucleotide marker.

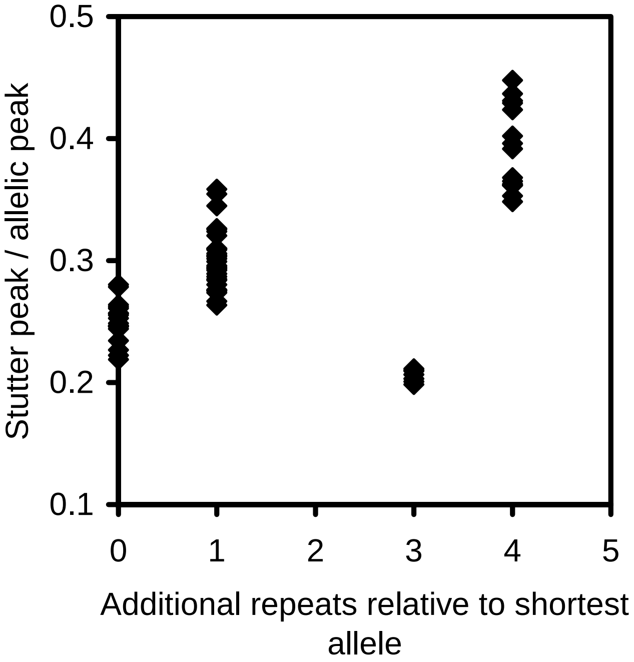


Figure 8.4. Aberrant stutter height in marker D10S189, present in seven copies of the 185-bp allele (which is three repeat units longer than the shortest detected allele).

Using sequence data from the CEPH/Généthon database, the length of the uninterrupted CA repeat, as well as stutter height were determined in multiple genotypes for 28 different markers, and values for all alleles were averaged. Figure 8.5 shows a correlation of 0.78 between the average lengths of the uninterrupted CA repeats and the average height of the first stutter peak in the 24 different dinucleotide markers without additional CT repeats.

Table 8.3. Repeat sequences and 10 bp of flanking non-repetitive sequence for marker D10S189 in several individuals who were homozygous for alleles of different lengths

DNA[a]	Length[b]	Sequence[c]
000243	179	TTTCACTATTTCTCTCTCTCTCCACACACACACACACACACACAGTTANCGATC
900009	181	TTTCACTATTTCTCTCTCTCTCCACACACACACACACACACACACAGGTANCGATC
134702	181	TTTCACTATTTCTCTCTCTCTCCACACACACACACACACACACACAGTTATCGATC
133102	185	TTTCACTATTTCTCTCTCTCTCCACACACACACACA**T**ACACACACACACAGGTATCGATC
980774	185	TTTCACTATTTCTCTCTCTCTCCACACACACACACA**T**ACACACACACACAGGTATCGATC
990329	185	TTTCACTATTTCTCTCTCTCTCCACACACACACACA**T**ACACACACACACAGGTATCGATC
980554	187	TTTCACTATTTCTCTCTCTCTCCACACACACACACACACACACACACACANAGGTANCGATC

[a] DNA-ID number is shown.

[b] Allele length in base pairs.

[c] The sequence of reference CEPH individual 134702 was obtained from the CEPH/Généthon database. The C>T substitution in the 185-bp alleles is printed in bold.

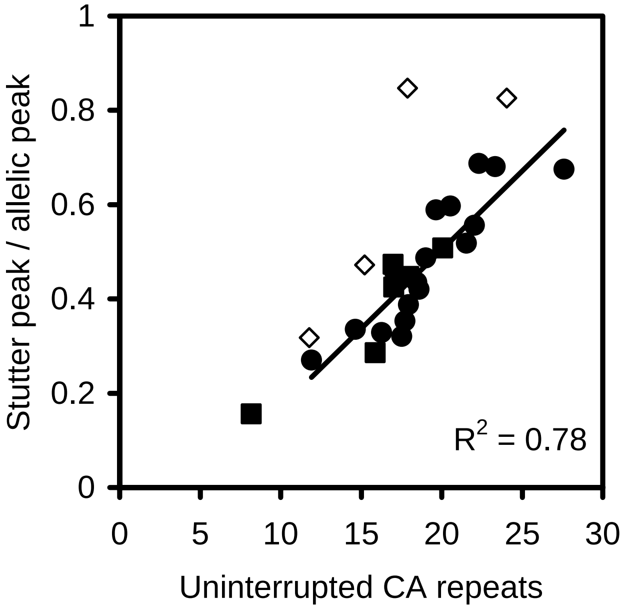


Figure 8.5. Average stutter height versus average length of the uninterrupted CA repeat sequence for 28 dinucleotide markers. Black circles indicate markers with uninterrupted repeats, black squares indicate markers with interrupted repeats and open diamonds indicate markers with additional CT repeats. The regression line for the markers without additional CT repeats was y = 0.0335x - 0.1651, with correlation coefficient 0.78.

The regression line indicates an increase of 3.4 % for each repeat unit increase in size, which is identical to the increase per repeat unit for alleles of a single marker, mentioned above. Furthermore, the line intersects the x-axis at 5.6 repeat units, suggesting a threshold before stutter can occur. The four markers with an additional CT repeat had higher stutter peaks than the CA-only markers. The highest stutter was measured in marker D14S77, which was the only marker with two additional CT repeats and a short tetranucleotide motif.

The influence of repeat length and repeat interruption on stutter height, suggested by these observations, was then tested by synthesizing PCR products that incorporated different repeats, but were otherwise identical. As can be seen in Figure 8.6, stutter height increased in a linear fashion with repeat length. Interestingly, as with the regression line of the 28 combined microsatellites the regression line in this different setting intersects the x-axis at 5.6 repeat units. An interruption of the repeat by a C to T transition, comparable to the 185-bp allele of marker D10S189, decreased stutter height in a repeat of 21 CA units to the level expected for a repeat with only 14.5 CA units.

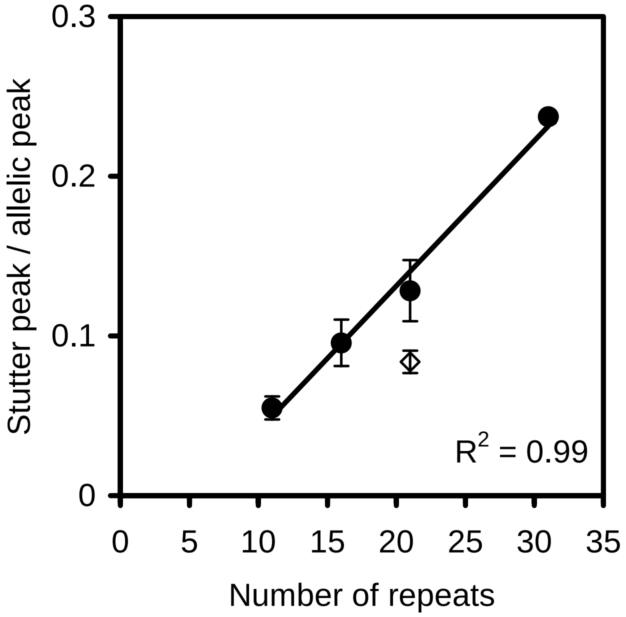


Figure 8.6. Stutter characteristics after introduction of different CA repeats into otherwise identical PCR templates. Averaged stutter heights in triplicate measurements are shown, with error bars indicating 1 standard deviation. Solid symbols: uninterrupted CA repeats of different lengths. Open symbol: $(CA)_{11}TA(CA)_9$ repeat. A regression line (y = 0.0091x - 0.0506) is shown for the four uninterrupted repeats (correlation coefficient is shown in the lower right corner of the graph).

Discussion

PCR induced stutter artefacts can seriously hinder genotyping of individual and pooled DNA samples with microsatellite markers. We investigated if differences in stutter intensity between microsatellite markers could be related to the absolute length and the sequence of the repeat. Although in vitro DNA amplification by PCR is an artificial system, the replication mechanism is similar to *in vivo* DNA replication, with a dependency on a DNA polymerase in both systems. Therefore, one would expect similar mechanisms to determine both stutter intensity after PCR and microsatellite mutation rates *in vivo*.

We determined stutter height in large series of existing genotypes from 38 different markers that were typed for a diabetes genome screen. As a measure, the height of stutter peaks was preferred over peak area, since we previously found peak height to be more reproducible (unpublished results), which was also noticed by others [252]. Differences in stutter height between markers, as well as in different alleles of a single marker, were found to be related to characteristics of the repeat.

Motif of the repeat unit. Average stutter was considerably higher in dinucleotides than in tetranucleotides, which confirms previous studies of stutter [253]. In publications that addressed microsatellite mutation rate instead of stutter, mutation rates were reported to be inversely related to motif size, with dinucleotides mutating at a 1.5 - 2 times higher rate than tetranucleotides [254]. Microsatellite instability in prostate cancer was reported to be higher for dinucleotides than for tri-, tetra-, and pentanucleotides [255]. Finally, when equal numbers of dinucleotide and tetranucleotide repeats were introduced into cell lines, mutation rates were considerably higher for dinucleotides than for tetranucleotides [256]. However, not all studies found higher mutation rates for dinucleotides [257, 258].

Number of repeated units. In CA repeat markers, a general increase of stutter height of almost 3.5% was noted for each additional repeat unit. A comparable increase with allele length was reported previously for sporadic markers [90, 99, 244], although not by all authors. When inserted into bacteria, constructs containing longer repeats showed higher stutter after PCR [248]. Our results generalize these observations by showing that differences in average stutter height between different markers can also be explained by the length of the uninterrupted repeat. Several *in vivo* studies of mutation rates support this finding. A $(GAA)_{10}$ repeat was found to expand less rapidly than a $(GAA)_{17}$ repeat [259], and in cultured cells as well as in Drosophila longer dinucleotide repeats displayed higher mutation rates [258, 260, 261]. A length dependence of mutation rate was also found in several human population studies [262, 263]. It was shown that both polymerase selectivity and proofreading efficiency decrease as the length of the repeat increased [264].

Both the results from the existing genotyping data and from the artificial PCR templates with different repeats suggested a minimum length of 5 to 6 repeat units before stutter could occur. These findings provide *in vitro* support for the reported finding that a threshold of 4-5 CA repeats had to be exceeded before mutations during PCR can be detected [248]. Other studies also suggested a threshold value before mutation of the repeat could be measured. A theoretical model of microsatellite evolution showed that polymerase slippage was not likely to occur in microsatellites of up to four repeats [263]. Others found a threshold of about 8 base pairs before polymerase slippage mutation could occur in S. cerevisiae [265].

Presence of repeat interruptions. In single markers, we found that stutter increased in a regular fashion with allele length. An exception was marker D10S189, which showed much lower stutter for the 185-bp allele. Sequence analysis showed that a C>T substitution was present in the 185-bp alleles, but not in alleles with a different length. A similar reduction of stutter height was noted when we generated otherwise identical PCR products repeats with and without a C>T substitution. Our results are in agreement with several sporadic observations in tetranucleotide markers. For instance, an interruption of a tetranucleotide sequence by a repeat with a different motif [245, 246] or with one hexanucleotide unit [247] reduced the amount of stutter. Moreover, at the tetranucleotide FGA locus, a correlation was found between stutter height and the length of the uninterrupted core repeat [245]. Studies of microsatellite mutation showed a similar stabilizing effect of repeat interruptions. In trinucleotide expansion diseases, an interruption with a CAA motif was reported to stabilize CAG repeats [266], while in mismatch repair (MMR) deficient cells mutation rates were decreased when a CA repeat was interrupted with a TA motif [267]. Reconstruction of the evolution of the DQCAR dinucleotide locus in large population samples indicated that alleles containing nucleotide substitutions had very low mutation rates [268].

Presence of additional repeat sequences. In the four markers with an additional CT repeat in the sequence surrounding the CA repeat, stutter was higher than would be expected from the length of the CA repeat alone. The marker with the highest stutter peaks had two additional long CT repeats and a short tetranucleotide repeat in its sequence. The fact that this marker also had a very high heterozygosity of 94% suggests that the additional repeats contribute to total stutter height and mutation rate. Indeed, dinucleotide repeats with motifs other than CA have also been reported to show slippage mutation, although at different rates. In Drosophila CA repeats had higher mutation rates than TA repeats [269]. In addition, CT repeats inserted into human lymphoblastoid cells mutated more frequently than CA repeats [270], whereas *in vitro* GAA repeats expanded more rapidly than CAG repeats of similar length [259]. It is likely that additional long CA repeats also increase stutter and marker polymorphism. Although examples are present in the database, unfortunately we did not have stutter information for these markers. We speculate that measured stutter height will be determined by an additive effect of all repetitive sequences above a certain minimum length. However, this hypothesis can only be tested by analyzing more markers with compound repeats. We did not do so, but the presented method to incorporate different repeats into PCR templates would allow such studies.

Our results suggests that stutter height in a particular microsatellite marker and in its different alleles can be predicted when the repeat motif and length of the uninterrupted repeat are known. The sequence surrounding the repeat seems to be of minor importance, unless additional repeat motifs are present. These observations suggest the possibility of a general correction of the stutter artefact in individual as well as pooled DNA samples, without the need for genotyping several individuals to derive marker-specific stutter correction models. There may be limitations to such an approach, however. It is likely that sequence-based correction models will only be valid when PCR conditions such as type of polymerase and number of PCR cycles are fixed. Under identical conditions, we and others found stutter height to be highly reproducible between different experiments, but changing a parameter such as the number of PCR cycles profoundly affected stutter height [242, 243]. This is a likely explanation for the observation that stutter for similar repeat sizes was higher in our genome scan data, amplified with 33 cycles, than in the artificial PCR templates, which were amplified using 27 cycles. The general behaviour of stutter under specific conditions must therefore be determined at least once, which would still require individual genotyping, although not for each marker. Furthermore, the presence of additional repeat sequences is likely to increase total stutter, but at present it is unclear to what extent. Finally, the use of sequence data from public marker databases carries the risk that subgroups in the population under investigation will have mutations in the repeat sequences, similar to marker D10S189, with a resulting aberrant stutter pattern. Without the availability of several individual genotypes to verify stutter height, this is likely to go unnoticed. The alternative, sequencing an individual for each marker to determine repeat length, does not fully protect from this risk either, while it requires additional work. Therefore, it remains to be investigated if the theoretical possibility of sequence-based general stutter correction has advantages over deriving marker specific correction models from a limited set of individual genotypes [118].

The characteristics of stutter described in this study could be useful for the selection of microsatellite repeats to be used as genetic markers. For example, selecting marker loci with interrupted repeats, and choosing PCR primers close to the repeat, thus excluding additional repeats from the PCR product, may reduce stutter. On the other hand, incorporating additional repeat sequences in the PCR product may increase marker heterozygosity, although at the cost of increased stutter.

From a different perspective, it is suggested that stutter height can be used as a simple indicator of mutation rate of specific markers, in studies of microsatellite evolution, or population migration. In clinical settings, the diagnostic sensitivity of microsatellite instability in detecting deficient mismatch repair in malignancies may be increased by selecting markers with high stutter, which are likely to be mutation-sensitive. However, if stutter in a particular patient would be lower than usual for that marker, one should suspect mutations in the repeat sequence and the possibility of not detecting MSI in that particular marker due to a stabilized repeat.

In summary, this study shows that stutter height in microsatellites is mainly determined by the repeat motif and the absolute length of the uninterrupted repeat sequence. The feasibility of obtaining general stutter correction models from sequence databases remains to be investigated, however, since population-specific repeat variations may change stutter characteristics.

Acknowledgements

The authors thank J.H. van Tilburg and E. Strengman for providing genotype data, and R. van 't Slot for technical assistance. This project was supported by the UMC Utrecht's Genvlag program and the Makaria Foundation.

9 ACCURATE DETERMINATION OF MICROSATELLITE ALLELE FREQUENCIES IN POOLED DNA SAMPLES

H.G. Schnack*, S.C. Bakker*, R. van ’t Slot, B.M. Groot, R.J. Sinke, R.S. Kahn and P.L. Pearson

*Both authors contributed equally

Abstract

Pooling of DNA samples instead of individual genotyping can speed up genetic association studies. However, for microsatellite markers the electrophoretic pattern of DNA pools can be complex, and procedures for deriving allele frequencies are often confounded by PCR-induced stutter artefacts. We have developed a mathematical procedure to remove stutter noise and accurately determine allele frequencies in pools. A stutter correction model can be reliably derived from one standard “training set” of the same 10 individual DNA samples for each marker, which can also include heterozygous patterns with partially overlapping peaks. Compared with earlier methods, this reduces the number of genotypes needed in the training set considerably, and allows standardization of analyses for different markers. Moreover, the use of a procedure that fits all data simultaneously makes the method less sensitive to aberrant data. The model was tested with 34 markers, 18 of which were newly defined from human sequence data. Allele frequencies derived from stutter-corrected DNA pool patterns were compared with the summed individual genotyping results of all the individuals in the pools (n=109 and n=64). We show that the model is robust and accurately extracts allele frequencies from pooled DNA samples for 32 of the 34 microsatellite markers tested. Finally, we performed a case-control study in celiac disease and found that weakly associated disease alleles, identified by individual genotyping, were only detectable in pools after stutter correction. This efficient method for correcting stutter artefacts in microsatellite markers enables large-scale genetic association studies using DNA pools to be performed.

Introduction

It has been cogently argued that population-based genetic association studies will have a greater power than linkage studies to localize genes contributing moderately or only a little to the phenotype of complex diseases [271]. However, the detection of association, or linkage disequilibrium between a genetic marker and a disease locus in outbred populations is only possible over small genetic distances [13, 272-275]. For screening large genomic regions, or even comprehensive whole-genome association studies, this implies that hundreds or thousands of markers have to be genotyped for each subject. Such studies are barely feasible using currently available genotyping technology.

Pooling of DNA samples for genetic marker analysis is a method to reduce the amount of genotyping required in allelic association studies [85, 90, 93-96, 100, 244, 276]. This technique involves combining equal amounts of DNA from patients and controls into separate pools, and comparing the pools for differences in allele distributions of genetic markers. In the absence of haplotype information, which is the situation encountered in a typical association study based on pooled case-control comparisons, the biallelic variation of single nucleotide polymorphisms (SNPs) contains far less polymorphic information than microsatellite markers. Therefore, microsatellites provide a more powerful tool on a marker-by-marker basis than SNPs [85, 277]. However, in the case of microsatellite markers, the overall genotype patterns of pooled samples are often distorted by PCR artefacts such as stutter and preferential amplification, which prevent an accurate determination of the allele frequencies by simple procedures. Several methods have been proposed to handle these artefacts. Some studies compared summed differences in patterns between two pools without correction for PCR artefacts, and without allotting the individual allelic contributions to the differences [93-96].

A fundamentally different way to compare pool patterns is to correct the pool signal for predicted PCR artefacts, in order to derive more accurate estimates of the allele frequencies. Advantages of this approach are that it allows the comparison of frequencies for individual marker alleles, and that results from different experiments can be summated and analyzed using regular statistics such as chi-squared tests, since the entire pool signal is deconvoluted into individual allele counts [90]. All recent correction methods use information derived from a training set of individual genotype patterns to obtain information about the stutter behaviour of the marker under investigation. One approach is to build a matrix of stutter patterns for individual alleles [99, 100, 244]. This requires a set of well distributed homozygous or well-separated (non-overlapping) heterozygous individual genotype patterns, and interpolation or extrapolation has to be invoked to complete the matrix for missing alleles. These methods are sensitive to one or more non-representative patterns caused by, e.g., measurement errors.

Alternatively, a stutter model can be derived from individual genotypes, which is used to correct for stutter and permits interpolation of stutter for allele sizes not encountered in the training set [90]. The advantage of a model is that it partly removes the influence of aberrant patterns. On the other hand, it interprets the stutter peaks according to a fixed behaviour, which can yield a less accurate description. The model approaches presented thus far also require well-distributed homozygous or well-separated heterozygous individual patterns for each marker to define the model parameters. In both types of correction procedure, a rather large set (at least 20 [100] to 50 [96]) of individual patterns has been considered necessary to provide sufficient data to obtain the necessary stutter information [90]. The search for and analysis of informative marker data often make these approaches tedious and highly interactive.

We have developed a stutter correction method that fits a model to one small set of genotype data from 10 individuals. This standard training set is identical for all markers, and can be of any allelic composition, since it does not need to include particularly defined homozygous or heterozygous individuals. The accuracy of the stutter correction model has been tested on 34 different microsatellite markers and used in a case-control study for celiac disease.

Materials and methods

Definitions

Uncorrected pool: allele frequencies derived from pool signals uncorrected for stutter.

Corrected pool: allele frequencies derived from pool signals corrected for stutter.

True pool: allele frequencies obtained by individual genotyping of all samples present in a pool, and summing the allele counts.

Preparation of DNA pools, marker selection, PCR, and analysis

Genomic DNA was obtained from peripheral blood lymphocytes using established procedures. Stock solutions were diluted to approximately 25 ng/µl, vortexed gently, and measured with Pico Green (Molecular Probes, Leiden, the Netherlands) on a Genios plate reader (Tecan, Männedorf). Subsequently, samples were diluted to 10 ng/µl and final concentrations were measured in triplicate. Each sample was tested for adequate PCR amplification. Volumes containing 100 ng of DNA from individual samples were pooled. Pools, as well as a set of 10 random individual samples, were purified by phenol extraction, and diluted with water to 10 ng/µl. Characterized microsatellite markers were obtained from the Genome Database (GDB) and Marshfield database. New microsatellite markers were identified by searching a 4 Mb ADHD linkage region on chromosome 15 for microsatellite repeats [278] using the Tandem Repeat Finder program (TRF). PCR primers flanking the repeats were designed with the Primer3 program (sequences are available on request). A so-called pig-tail sequence extension was added to one of the primers in order to reduce plus-A artefact during PCR [165]. The other primer was labelled with 6-FAM, HEX, or NED fluorescent dyes (Biolegio, Malden, the Netherlands, and Applied Biosystems, Foster City, Ca., USA).

Individual samples and triplicate pools were amplified simultaneously as described elsewhere [278], but with 27 instead of 33 cycles. Up to three products were pooled, and analyzed on an ABI 3700 sequencer [278]. Sample files were analyzed using Genescan 3.5 and Genotyper 3.6 for Windows NT and the heights of all peaks were labelled. Samples with allelic peak heights below 200 or above 6000 were not labelled. A computer program called PoolFitter (freely available from our web site), which is a user interface invoking our stutter correction algorithm, then processed the tables with allele sizes and peak heights (see below). The pool patterns were corrected for stutter by applying the model parameters derived from the individual genotypes (see below). For marker D7S2422 only, preferential amplification of shorter alleles was compensated in the PoolFitter program, by dividing the peak heights of both individual data and pooled data before model fitting by a function fitted to the corrected heterozygous patterns without compensation for preferential amplification. Estimates from corrected and uncorrected pool patterns (averages of triplicate measurements) were compared with true pools using the program CLUMP [116].

The model

The basic concept is that, for pooled DNA, any electrophoretic microsatellite marker pattern (See Figure 9.1a) is the sum of its constituent parts comprising a mixture of homozygous and heterozygous individual patterns. Peaks may represent individual alleles, or individual alleles plus a stutter component, or only stutter. We describe a pattern by Y(a), where Y is the height of the signal at fragment length a. In all figures the signal height has been scaled to facilitate comparison with calculated quantities later on.

The length a can assume discrete values differing by multiples of the repeat length Δa. For a dinucleotide marker, $\Delta a = 2$ base pairs. Looking at a pattern for a single allelic peak at length a_0, one expects stutter peaks at $a_0 - \Delta a$, $a_0 - 2\Delta a$, ... and possibly "up-stutter" peaks at $a_0 + \Delta a$, ... In general, peaks are located at $a = a_0 - m\Delta a$, with m integer. The modelled peaks are described by y(a), representing the peak height at length a. This peak y(a) can have contributions from an allelic peak $y_0(a)$, and from stutter peaks of alleles located m base pairs away: $y(a) = y_m(a + m\Delta a)$. Thus, the index m refers to the order of the (stutter) peak: $y_0(.)$ is the main, allelic peak, $y_1(.)$ is the first stutter peak, and so on. The argument of $y_m(.)$ refers to the location of the main, allelic peak.

Our aim is to describe the total set of peaks by a model with as few parameters as possible. The values of the parameters will depend on the marker, PCR conditions, settings of the electrophoresis apparatus, etc. Knowing the stutter parameters, it is possible to deconvolve the measured signal Y(a) of a DNA pool into the contributions of individual alleles and hence calculate the frequency of each allele in the pool.

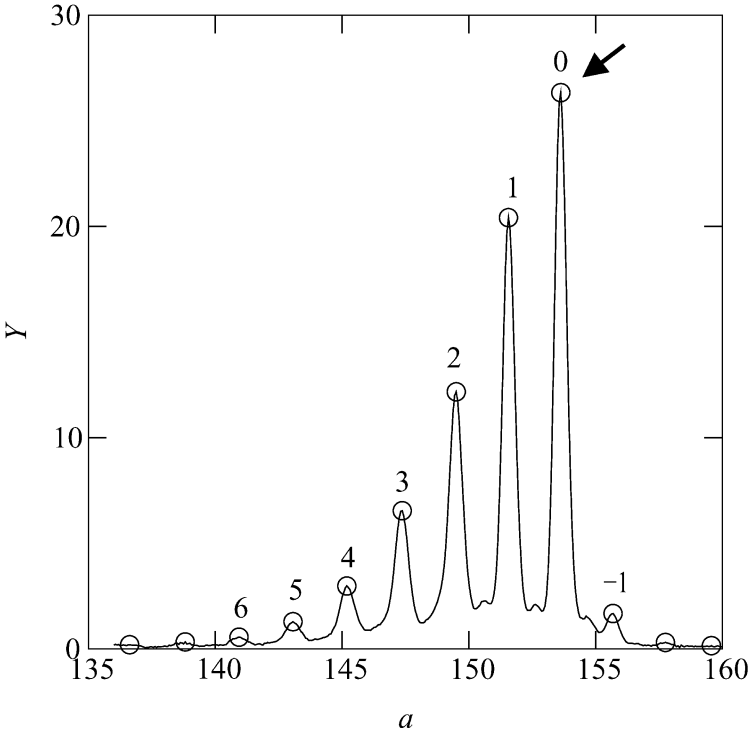


Figure 9.1a. Typical individual electrophoretic pattern after removal of background signal (line). The marker is the dinucleotide repeat marker DRD5. The individual is homozygous for the allele of size *a*=153 base pairs (indicated by the arrow). The circles indicate the tops of the peaks that are used as measures of the amount of stutter and allele signal present. The numbers refer to the indexing of the peaks: 0 is the allelic peak, 1 the first stutter peak, and so on. Since the absolute heights of the peaks are not important, *Y* has been scaled to a 100% scale in most figures.

The ratio between allele and stutter signal appears to be marker specific, and since pool patterns do not contain enough information to obtain reliable estimates for the stutter parameters used in the model, these parameters have to be derived by fitting the model to a number of individual test patterns for each marker.

Stutter pattern of a homozygous individual

It can be demonstrated that the heights of stutter peaks decay exponentially with the number of stutters, as clearly shown in a logarithmic plot (Figure 9.1b), in which a straight line can be drawn through the tops of the stutter peaks. A few simple theoretical assumptions about the nature of DNA amplification predict this exponential behaviour [237]. From many of such plots we found empirically that the ratios of the heights of successive stutter peaks are roughly constant for all samples of the same marker and amplification condition, but that the constant differs between markers and conditions. We denote this constant by the ratio *r*.

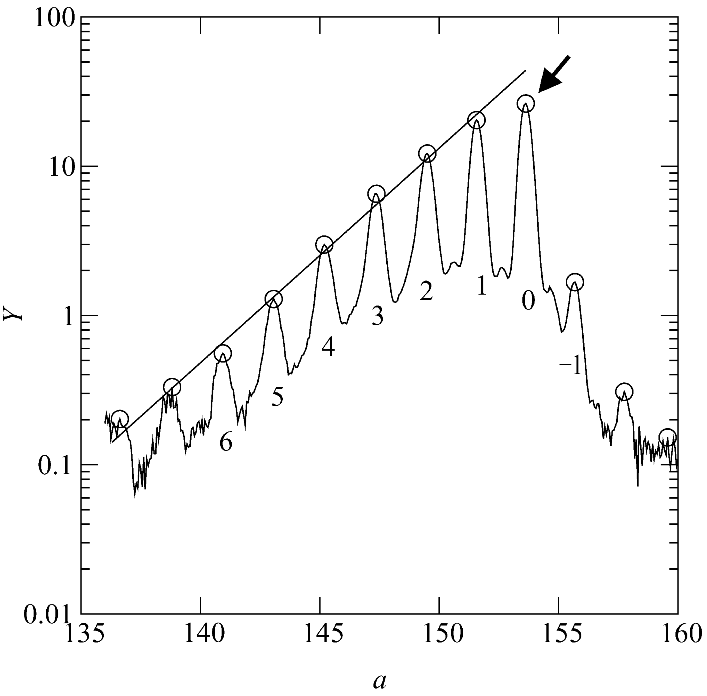


Figure 9.1b. The same electrophoretic pattern as in Figure 9.1a, plotted on a logarithmic scale (line and circles; the numbers refer to the indexing of the peaks). The straight line indicates the exponential relationship between successive stutter peaks. The arrow indicates the main, allelic, peak at *a*=153.

The first stutter peak is usually found to be proportionally higher compared to the main peak; in Figure 9.1b this is observed as a deviation of the stutter straight line with the top of the allelic peak. We therefore use a different ratio to describe the relationship between the first stutter peak and the allele peak:

$$y_m(a) \,/\, y_{m-1}(a) = r, \qquad m=2,3,\ldots \tag{1a}$$

$$y_1(a) \,/\, y_0(a) = \lambda r, \qquad m=1 \tag{1b}$$

with 0<*r*<1, and λ>1, normally. This is for the "normal" downward stutter. For the upward stutter, we take only one peak into account, as it is rare to see more up-stutter peaks; however, the model can easily be extended to more, if necessary.

$$y_{-1}(a) \,/\, y_0(a) = \mu, \qquad m= -1 \tag{1c}$$

with 0<μ≪1, normally.

For all other positions, i.e., positions at larger lengths:

$$y_m(a) = 0, \qquad m = -2,-3,.. \tag{1d}$$

We can combine and rewrite eqs. (1a-1d) as follows:

$$y_m(a) = \begin{cases} y_0(a), & m = 0 : \text{main peak} \\ \mathrm{y}_0(a)\lambda r^m, & m = 1,2,3... : \text{stutter peaks} \\ \mathrm{y}_0(a)\mu, & m = -1 : \text{up - stutter} \\ 0, & m = -2,-3,... \end{cases} \tag{2}$$

The fragment length of (stutter) peak $y_m(a)$ is:

$$a_m = a - m\Delta a. \tag{3}$$

It is usually observed that stutter is more severe for longer alleles than shorter alleles. This can be understood, at least qualitatively, by realizing that a larger number of repeats offers more chances for the PCR process to stutter. We therefore introduce an *a*-dependence of the stutter ratio *r*:

$$r = \exp(b_0 + b_1 a). \tag{4}$$

For positive values of b_1 this formula yields an increasing stutter for increasing *a*.
The true amount of signal at the allelic peak, i.e., if no allele signal had been dissipated into stutter peaks, is represented by:

$$y_t(a) = \sum_{m=-1}^{\infty} y_m(a). \tag{5}$$

We have now described the set of stutter peaks by four parameters: b_0, b_1, λ, μ. In the trivial case of a pattern of a homozygous individual, eq. (2) can be fitted directly to the $Y(a)$ data, with $y_0(a_0)$ as a fifth fit parameter. The length of the allele is directly read from the pattern: a_0; $y_0(a_0)$ is just the height of the measured main peak $Y_0(a_0)$; μ is the ratio of the up-stutter peak at $a_0+\Delta a$ and the main peak. The factor *r* is determined by the logarithm of the heights of the stutter peaks $y_1(a_0)$, y_2,(a_0) ... the heights of which are directly taken from the measured peaks $Y(a_0-\Delta a)$, $Y(a_0-2\Delta a)$, ... Then λ follows from $y_1(a_0)/y_0(a_0)$, with the value of r inserted in eq. (2). In this example r is kept constant; in a more realistic situation involving alleles of several lengths (such as in a pooled DNA sample) b_0 and b_1 can be fitted instead of a constant r.

Stutter pattern of a heterozygous individual

In the case of a heterozygous individual, there are more measured peaks to fit, and there is one extra fit parameter, namely the y_0 of the second allelic peak. We will refer to the two y_0s as: 0S and 0L, located at a_{0S} and a_{0L}, respectively, with $a_{0S} < a_{0L}$ (see inset of Figure 9.2). Heterozygous patterns often overlap to a large extent, and pool patterns always do. For two alleles close together, the measured peak heights in the overlapping region are the sum of two contributing peaks, one for the (shorter) S-allele, and one for the (longer) L-allele. For instance, the peak at the left arrow in the inset of Figure 9.2 is made up of the first stutter peak of the S-allele and the third stutter peak of the L-allele and is represented by:

$$y(a{=}151) = y_1(a_{0S}) + y_3(a_{0L}) = y_0(a_{0S})\lambda r + y_0(a_{0L})\lambda r^3, \qquad (6)$$

with $a_{0S}{=}153$ and $a_{0L}{=}157$.

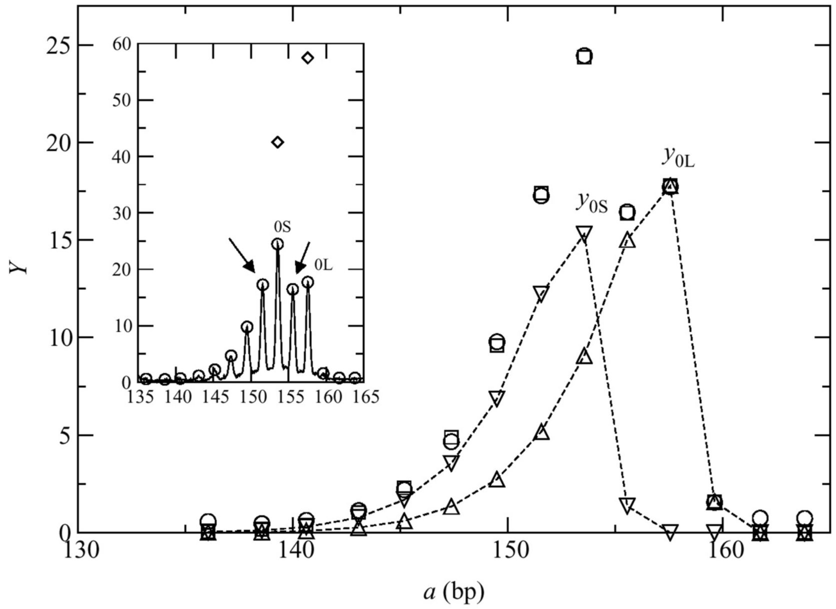


Figure 9.2. Model fitted to a heterozygous individual electrophoretic pattern (see inset for the original pattern (line) with peaks (circles); allelic peaks 0S and 0L are indicated). The marker is DRD5. The circles represent the data peaks; the squares show the fit. The two types of triangles and the dashed lines indicate the individual peak patterns of the two alleles (y_{0S} and y_{0L}) that comprise the measured signal. The diamonds in the inset represent the corrected, i.e., estimated, frequencies of the two alleles. These values are obtained by summing all peaks for each separate allele and normalizing the total sum of the two alleles to 100.

This effect makes the fit procedure more challenging and real solution algorithms have to be invoked. We used the Levenberg-Marquardt method [279]. The result of such a fit is shown in Figure 9.2. The model fits the data well, and the relatively large contribution of the stutter peaks of the L-allele to both the allelic peak and stutter peaks of the S-allele is clearly seen.

Pattern of a pooled sample

The generalization to fitting a pattern of a pooled DNA sample containing alleles of *n* individuals is straightforward. At every measured fragment length *a* the following peaks can contribute to $y(a)$, depending on the presence of alleles in the pooled sample:

(a) the allelic peak $y_0(a)$ of the allele at $a_0=a$;

(b) the up-stutter peak $y_{-1}(a–\Delta a)$ of the allele just left of it, at $a_0=a–\Delta a$;

(c) the first-order stutter peak $y_1(a+\Delta a)$ of the allele just right of it, at $a_0=a+\Delta a$;

(d) higher-order stutter peaks $y_m(a+m\Delta a)$ of alleles more to the right (m=2,3,...).

In a formula, this can be written as:

$$y(a) = \sum_{m=-1}^{\infty} y_m(a + m\Delta a)\,, \qquad (7)$$

with $y_m(a + m\Delta a) = y_0(a+m\Delta a)\lambda\exp((b_0 + b_1a)m)$ for $m\geq1$ the m-th stutter peak of the allele at $a_0=a+m\Delta a$, and $y_{-1}(a–\Delta a) = y_0(a\text{-}\Delta a)\mu$ the up-stutter peak of the allelic peak just left, at $a_0 = a–m\Delta a$. The arguments of $y(.)$ in eq. (7) must lie in the measured range of allele lengths $(a_{\min},a_{\max})$.

The $y_t(a)$s are now calculated as follows

$$y_t(a) = \sum_{m=-1}^{\infty} y_m(a)\,. \qquad (8)$$

These values are proportional to the number of individuals *n* contributing to that allele *a*. To obtain estimates of the true allelic frequencies $F(a)$ in the pool, one calculates:

$$F(a) = 2ny_t(a)/\sum_{a'} y_t(a')\,, \qquad (9)$$

where the summation is carried out over the full range $(a_{\min},a_{\max})$ of the pool pattern.

To correct a pattern of a pooled DNA sample one has to fit eq. (7) to the measured data $Y(a)$. Values for the four model parameters b_0, b_1, λ, μ could be found from fitting the model to the genotype patterns of a small number of representative individuals one at a time, and deriving n_i values for each of the fit parameters. These n_i values could then be averaged to obtain a good estimate for each of the parameters. A much more efficient way is to perform the model fitting to all individual patterns simultaneously. The total number of data points is $n_i m_i$, with m_i the average number of measured peaks per individual. The total number of fit parameters is $4 + n_i(1+h)$, with h the calculated heterozygote frequency of the marker ($0 \leq h \leq 1$). A set of $n_i = 10$ individuals, each with on average $m_i = 7$ data points and a heterozygote frequency of $h = 0.5$, requires fitting a model with 19 parameters to a combined data set of 70 data points, which, as shown in Figure 9.3 for marker D6S273, yields a very stable fit.

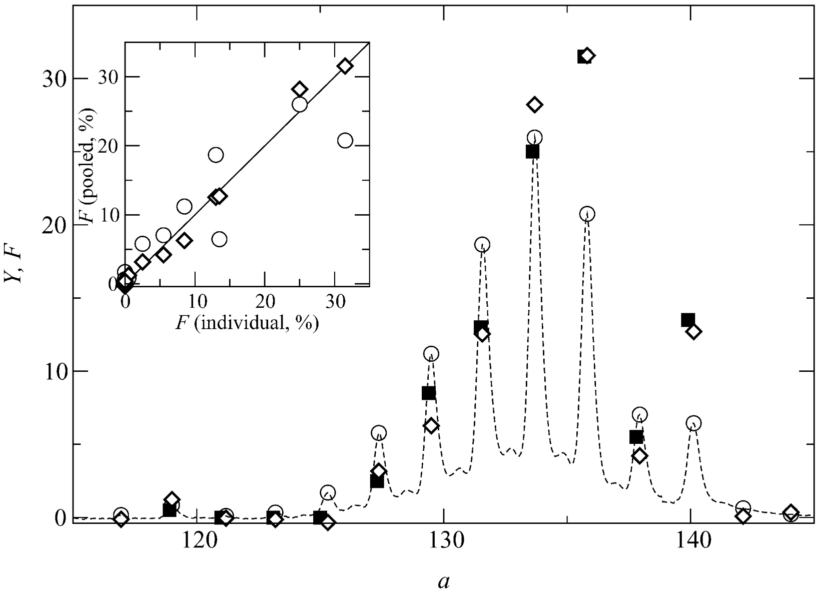


Figure 9.3. Electrophoretic pattern of a pool of 100 individuals for marker D6S273 (dashed line). The circles mark the peaks used in the analysis. Diamonds indicate the estimated true allelic frequency *F* after compensation for stutter, as calculated from the model. The model parameters were derived from ten randomly chosen individual marker patterns. The fit parameters (b_0, b_1, λ, μ) were inserted into the pool fit, from which the y_is were derived. For most alleles quite a big difference is observed between the original and corrected peak heights. The squares represent the summed individual genotypes. The inset shows the relationship between summed individual frequencies and uncorrected (circles) and corrected (diamonds) frequencies from the pool. The straight line represents the identity line: symbols on this line represent alleles for which the frequencies estimated from the pool equal the summed individual frequencies, showing perfect agreement.

Comparison with a deconvolution method

The most robust method previously published is the deconvolution method described by Perlin *et al.* [99]. Like our method, it uses a set of individual patterns to obtain the stutter behaviour. The main difference between our method and Perlin *et al.*'s is the fact that we fit a model to the data to describe the stutter behaviour, which makes our method potentially much more robust, and thus requiring fewer individual patterns to train the method. This has been tested below.

Results

Minimum number of genotypes required in the training set

For marker D6S273, we investigated the influence of training set size on the reproducibility of the results by fitting models based on sets varying in size from 2 to 30 individuals, that were taken at random from the n individuals in the pool. For each chosen set size, a random selection of individuals was taken 20 times to derive the model parameters and to correct the pool data. Figure 9.4 shows the effect of training set size plotted against the spread in the corrected peak height of one of the alleles (a=127; see Figure 9.3) in the pool. We chose to show this allele because of its low frequency (3%), in which adequate correction is crucial. For sets smaller than about 5 individuals the variation in the results was relatively large, but for 10 individuals or more, the gain in reproducibility was limited. The effect of training set size was also tested for two other dinucleotide markers, with similar results (not shown). A set of n_i=10 was found to give reliable results with a coefficient of variation of about 1%.

This test was also carried out for Perlin *et al.*'s method [99]. For all alleles and values of n_i, the variation in estimated frequencies from this method was at least 3 times as large. For n_i=30 the variation was still twice as high as our method's variation for n_i=10.

Robustness to atypical training sets and measurement errors

Using data for marker D6S273, the robustness of the algorithm was checked in the following way: various sets of n_i=10 individuals were used as a training set to fit the model to the pooled data depicted in Figure 9.3: (i) a set of 10 homozygous individuals; (ii) a set of 10 heterozygous individuals; (iii) a set of 10 individuals whose alleles were closely packed together in a certain region, leaving part of the allelic range of the pool uncovered; (iv) a set of 8 regular individuals plus 2 measurement errors: patterns containing an allele exhibiting a completely different stutter behaviour to the others (but with peaks in the same molecular weight range). All tests yielded good results that hardly differed from the "normal" pool fit results of Figure 9.3.

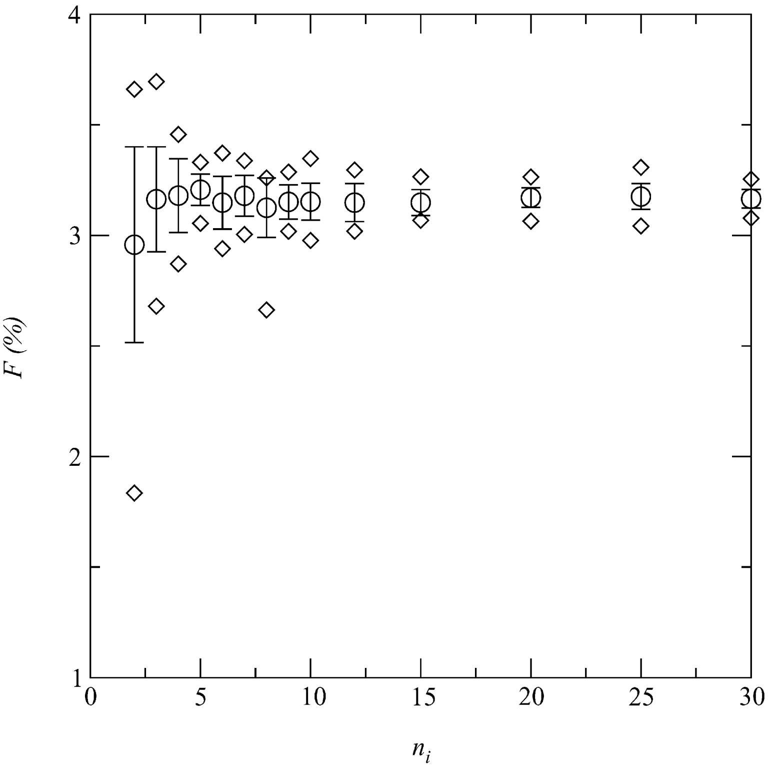


Figure 9.4. The peak at *a*=127 of Figure 9.3, calculated from fits with individual sets of n_i=2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25 and 30 individuals. For each value of n_i, 20 fits with randomly comprised sets were carried out. The values of the allele frequency *F* derived following correction are plotted on the vertical axis: mean (circles) and standard deviation (error bars) and the smallest and largest values (diamonds). Small sets already provide reliable frequencies with mean values of 3.2. The frequency found from summing the individual genotypes in the pool is 2.5. The frequency read from the uncorrected pool pattern was 5.8.

A comparison of test (ii) with test (i), shows that no pre-chosen homozygous (or well-separated heterozygous) individual patterns are needed to derive good parameter estimates. Further, test (iii) shows that there is no need for training data to cover the full molecular range of alleles. Only in the extreme case of having only data points at one extreme of the molecular range in the training set, do the pool results at the other end become less reliable. Test (iv) simulates the presence of measurement errors. If one or two of the 10 individual patterns are dissimilar to the others, e.g. because of an artefact in the PCR process or a measurement error, the fit procedure does not appear to be misguided. The test showed that the fits derived from a training set of 8 normal and 2 abnormal patterns were nearly as good as those based on 10 good patterns.

Validation of the model

For 34 different microsatellite markers, correction models were derived from the same training set of 10 individuals, and both uncorrected and corrected pools were compared

Table 9.1. Statistical comparison of allele frequencies obtained by individual genotyping and frequency estimates from uncorrected and corrected pool patterns

				Uncorrected		**Corrected**	
Marker	**Type**	**Alleles**[e]	**Heterozygosity**	***r***[f]	***P value***[g]	***r***	***P value***
D11S1338	di[c]	7	0.72	0.92	$<10^{-3}$	0.99	0.89
D11S1760	di	10	0.77	0.44	$<10^{-3}$	0.61	$<10^{-3}$
D11S3178	di	10	0.67	0.84	$<10^{-3}$	0.99	0.92
D11S3179	di	7	0.70	0.93	$<10^{-3}$	0.99	0.93
D3S3585	di	7	0.58	0.92	$<10^{-3}$	1.00	0.80
D3S3665	di	6	0.55	0.93	$<10^{-3}$	0.99	0.80
D4S1582	di	7	0.78	0.89	$<10^{-3}$	0.98	0.88
D5S2005	di	7	0.66	0.91	$<10^{-3}$	0.99	0.10
D6S273	di	6	0.70	0.84	$<10^{-2}$	0.99	0.95
D6S291	di	8	0.72	0.92	$<10^{-3}$	0.99	0.20
D7S2422	di	15	0.83	0.78	$<10^{-3}$	0.93	0.01
DRD5	di	13	0.79	0.51	$<10^{-3}$	0.93	0.21
RH27315	di	5	0.64	0.86	$<10^{-3}$	1.00	0.87
D19S400	tetra[d]	9	0.84	0.98	0.96	0.98	0.97
GAAT	tetra	6	0.69	0.98	0.45	0.98	0.44
TH01	tetra	7	0.77	0.99	0.85	0.99	0.85
Average[a]		**8**	**0.71**	**0.85**		**0.96**	
kk3	di	6	0.75	0.81	0.06	0.94	0.75
kk7	di	6	0.74	0.90	0.29	0.99	0.93
kk9	di	5	0.77	0.81	0.02	1.00	1.00
kk11	di	6	0.73	0.81	0.02	0.99	0.99
kk16	di	6	0.57	0.86	0.07	1.00	0.69
kk20	di	7	0.55	0.95	0.19	0.99	0.19
kk24	di	12	0.82	0.78	$<10^{-3}$	0.99	0.49
kk26	di	7	0.77	0.80	0.05	0.96	0.90
kk28	di	14	0.79	0.76	0.08	0.95	0.77
kk31	di	14	0.86	0.77	0.78	0.96	0.96
kk37	di	9	0.72	0.85	0.19	0.99	1.00
kk42	di	6	0.31	0.90	$<10^{-3}$	1.00	0.63
kk43	di	9	0.72	0.82	0.01	0.99	0.91
kk45	di	11	0.70	0.79	0.02	0.99	0.94
kk56	di	9	0.76	0.83	0.11	0.99	0.82
kk58	di	9	0.84	0.83	0.67	0.95	0.97
kk61	di	9	0.78	0.88	0.26	0.99	0.98
kk62	tetra	7	0.75	0.87	0.28	0.85	0.16
Average[b]		**8**	**0.72**	**0.83**		**0.97**	

[a] Upper half of the table: values for characterized markers, analyzed in pooled DNA from 109 individuals.

[b] Lower half of the table: values for "home-made" markers, analyzed in pooled DNA from 64 individuals.

[c] Dinucleotide repeat marker.

[d] Tetranucleotide repeat marker.

[e] Number of marker alleles, determined by individual genotyping of pool samples.

[f] Correlation coefficient of individual genotyping results and estimates from uncorrected as well as stutter corrected pools.

[g] *P* value of chi-squared tests after combining alleles with expected low values, as calculated by the CLUMP algorithm.

with the true pools (for definitions see Materials and Methods section). An example is shown in Figure 9.3. In total five genotypes from 4 different markers could not be determined reliably, and these were discarded. Correlation coefficients of uncorrected and corrected pools vs. true pools for all 34 markers are given in Table 9.1. A graphical representation of the data for 16 characterized markers is given in Figure 9.5.

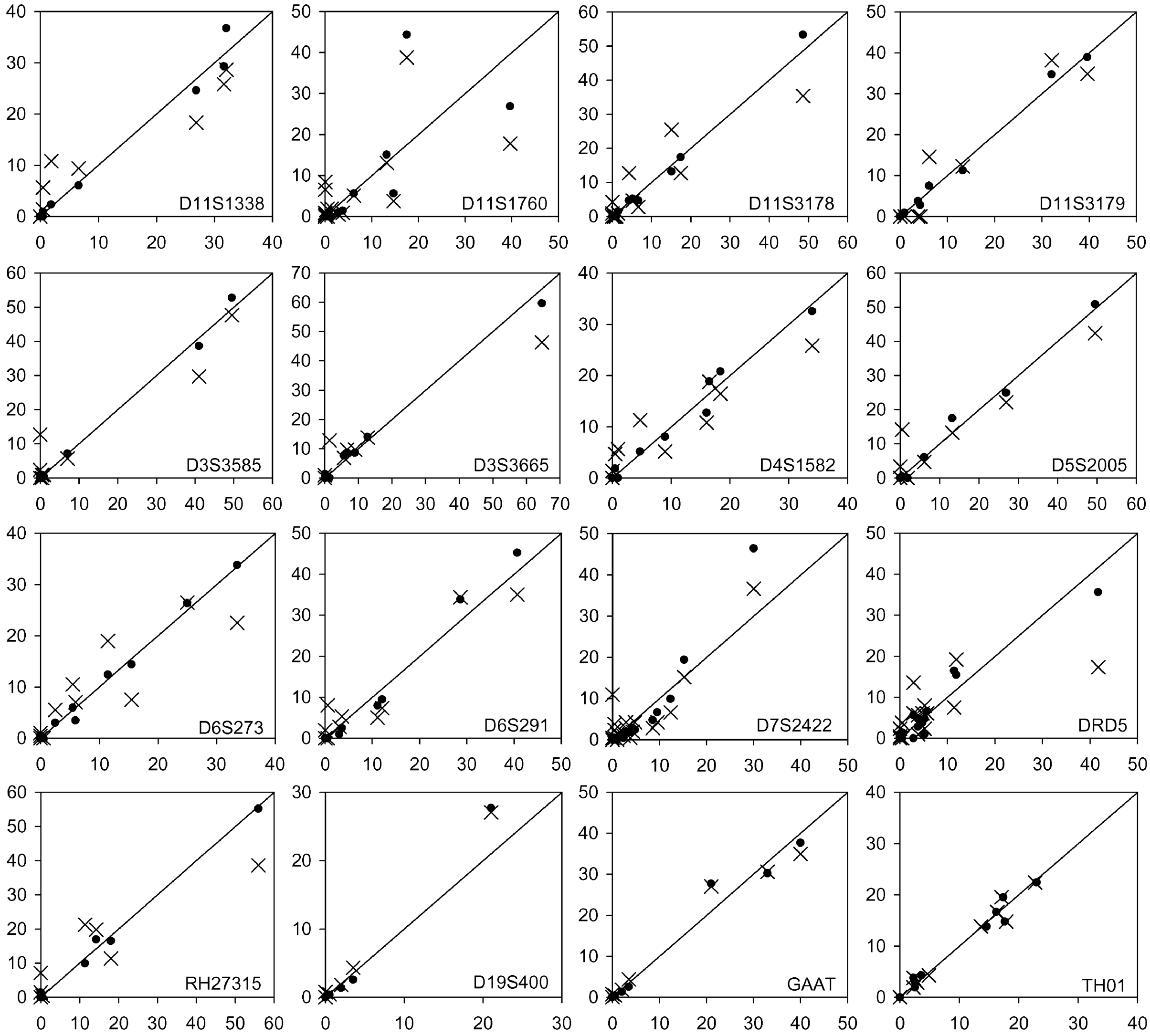


Figure 9.5. Comparison of summed individual genotyping results with allele estimates from pools. X-axis: allele frequencies (x100%) determined by summing individual genotyping results of all individuals in a pool. Y-axis: allele frequencies (x100%) estimated from pooled samples. Uncorrected estimates are indicated with crosses and corrected estimates with dots. The diagonal (y=x) indicates perfect agreement between summed individual genotyping results and estimates from pool. Marker names are shown in the lower right corner of each graph.

The only markers in which uncorrected pools approached true pools were the four tetranucleotide markers. For the dinucleotide markers uncorrected pools were generally very different from the true pools, whereas corrected and true pools did not differ significantly, with the exception of markers D11S1760, D7S2422, and kk9. For marker D11S1760 there was a large overestimation of the frequency of the shortest allele in both uncorrected and corrected pools. Analysis of all individual genotype patterns for this marker revealed that stutter did not increase with allele length in a regular fashion (see Figure 9.6), which is an underlying assumption in the correction model. Marker D7S2422 showed a systematic overestimation of the peak height of the shorter allele in heterozygotes in the PoolFitter program, which persisted after correction for stutter. This suggested preferential amplification of shorter alleles, and after applying a simple compensation in the program, the differences between corrected and true pools were no longer significant (data not shown).

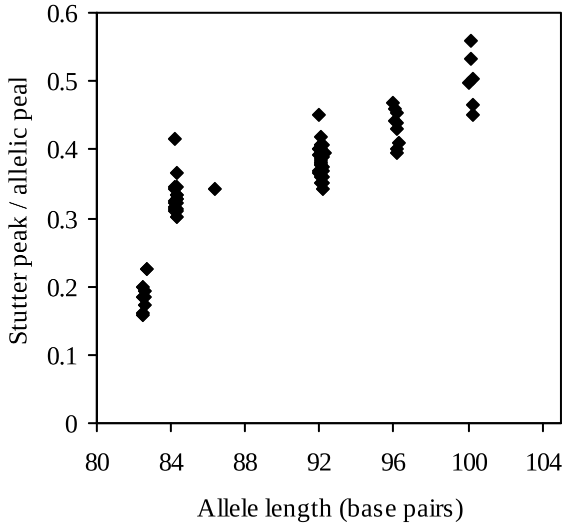


Figure 9.6. Relative stutter height for each allele of dinucleotide marker D11S1760, as determined from individual genotypes. Stutter height is depicted as the ratio of the highest stutter peak and the allelic peak.

No evidence for preferential amplification was found in the other markers (see Figure 9.7). Marker kk9 had two extra alleles (together accounting for 18% of all alleles in the true pool), with a size exactly between alleles at the regular 2-bp intervals. These aberrant alleles were discarded from the analysis, since the correction method ignores alleles at irregular intervals.

Case-control study

We investigated the application of the correction method in a case-control study in celiac disease (CD). DNA from 50 CD patients and 100 healthy controls was combined into two pools. Five microsatellite markers that had previously been used in association studies of CD patients were blinded and analyzed in CD and control pools.

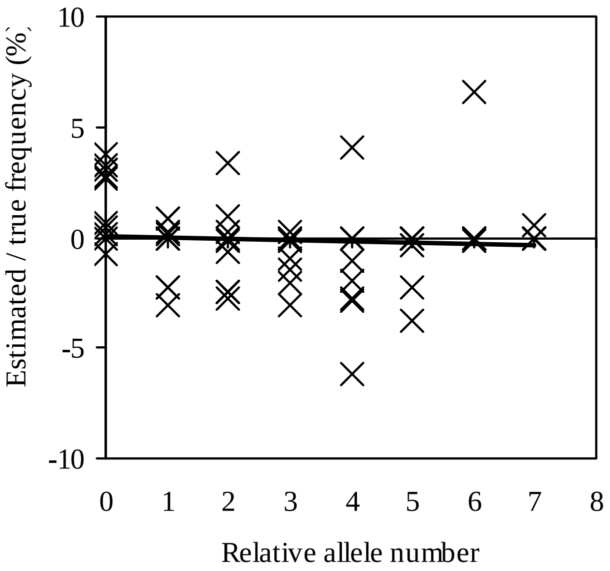


Figure 9.7. Average deviation from true frequencies of estimated frequencies after stutter correction. Data for 10 random dinucleotide markers are shown. Y-axis= (estimated/ true frequencies) * 100%. Only averages of at least three independent alleles of each size are shown. Y-axis= allele number relative to the shortest allele of each marker. The shortest allele is indicated with 0, allele 1 has one additional repeat, etc. A trend line for the combined measurements is shown.

Interpretation: in the case of preferential amplification of alleles, one would expect a systematic overestimation of shorter alleles, and an underestimation of longer alleles after correction for stutter, which is not the case. Apparently, higher stutter of longer alleles accounts for most of the observed peak height reduction for longer alleles in raw genotype data.

For three markers, allele frequencies did not differ significantly between cases and control pools, in either individual genotyping or pooled analysis. The other two markers showed significant differences between cases and controls. In each marker, one allele was very strongly associated, and already detectable in uncorrected pools, but after stutter correction, both markers also showed a much weaker but significant association with a second allele (see Figure 9.8). Both weak and strong associations were also demonstrable in the summed individual analysis.

Discussion

Although the use of pooled DNA enormously reduces the amount of genotyping in comparing cases and controls, it suffers from the inability to generate haplotype information. As a result, microsatellite markers, with their high information content, are much more suitable than SNPs for use in pooled DNA samples.

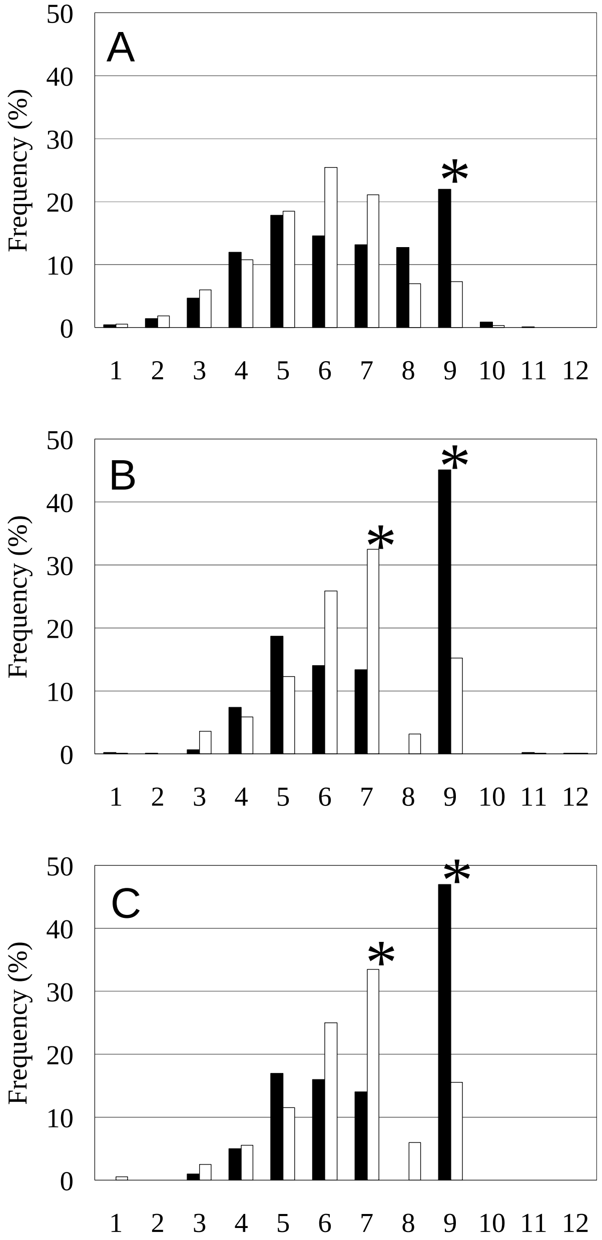


Figure 9.8. Comparison of allele frequencies in pools of celiac disease patients (black bars) and healthy controls (white bars) for marker D6D273. A: uncorrected pools, B: corrected pools, C: true pools. X-axis: allele number (increasing size). Y-axis: frequency of individual alleles (%). Significant differences for single alleles (p<0.05, not corrected for testing multiple alleles) are indicated with (*).

The number of potentially polymorphic microsatellites in the genome is much higher than the number of characterized markers in public databases. For example, in 11 schizophrenia candidate genes we have tested 19 polymorphic microsatellites, 8 of which were intragenic, while flanking markers were on average at 45 kb distance from the gene (max 130 kb). In a schizophrenia candidate region, we found nearly 250 potentially microsatellites with an average spacing of 55 kb (max 168 kb). However, the widespread application of microsatellite markers in DNA pooling may have been prevented by uncertainties induced by stutter artefacts and the consequent distortion of allele frequency estimates.

We have developed a novel method, which enables accurate extraction of allele frequencies from microsatellite pool signals. A prerequisite for the application to large studies is that the correction method does not entail much additional analysis time. Our method meets this requirement, since the same training set of only 10 independent DNA samples plus the pool samples is required to carry out an analysis for a given marker. An apparent advantage of our approach is that there is no requirement for stutter and allelic peak signals of heterozygous individuals to be clearly separated, which greatly reduces the number of individuals required. There was little gain in accuracy when more than 5-10 individual genotypes were used, and accordingly, we choose one set of the same 10 independent individuals for all analyses, to allow for occasional dropouts. Other advantages of our fit algorithm are the simultaneous fitting of all data, which decreases the sensitivity to aberrant data, and that the size distribution of alleles and stutter, or alleles with an anomalous stutter height, had little influence on the predictive accuracy of the model.

The model was tested on DNA pool patterns with 34 different microsatellite markers, 18 of which were newly defined from human sequence data, since well-characterized markers could have been selected for their accuracy in genotyping. Our results with tetranucleotide markers confirm previous reports that stutter is low in these markers (generally < 5%) and that no stutter correction is required [90, 100, 101]. Significantly, for the two dinucleotide markers in which correction remained inaccurate, the presence of an aberration was readily detected in the PoolFitter program, even though it could not correct the stutter distortion.

In a case-control study involving celiac disease, marker alleles that were weakly associated in individual genotyping were also found to be associated in the pool analysis, but only after stutter correction. These two exceptionally strongly associated markers in the HLA region would have been detected even without correction, but would have been missed if only the weakly associated alleles had been present. This clearly demonstrates the benefit of stutter correction in DNA pooling.

Taken together, stutter correction generally resulted in accurate estimates of true allele frequencies in DNA pools. Compared with methods that use uncorrected pool patterns, several important advantages are apparent. The recently proposed ΔAIP and ΔTAC methods compare overall differences in peak area or peak height between pool patterns [93, 96]. However, both methods assume a single fixed stutter profile for all markers and simulate large numbers of pool patterns to determine what proportion by chance will deviate significantly. Since the heights as well as the number of stutter peaks can differ greatly between markers, these methods raise the question whether realistic significance levels can be calculated in this way. In any case, such an approach prevents ascribing differences between pools to single alleles and summing results from different sub-pools or different experiments [90]. These drawbacks are not evident in our method.

We found that technical measures, such as reducing the number of PCR cycles, and adding pig-tail sequences to primers to eliminate plus-A artefacts, and separation on a capillary sequencer instead of a slab gel machine [116], consistently improved the accuracy of DNA pool measurements. However, the nature of DNA pooling will inevitably result in some loss of sensitivity compared to individual genotyping. Furthermore, a four-parameter model is not a perfect description of reality.

Despite these and other limitations, such as the lack of haplotype information, until cheap and rapid large-scale individual genotyping of markers for single individuals becomes technically feasible, DNA pooling methods allow efficient initial screening of candidate regions, and candidate gene systems. In pooled DNA, microsatellites are much more informative than single SNPs. In a second phase, associated microsatellites could then be followed-up by individual genotyping of high-density SNP markers, and haplotype analysis. Even if cases and controls were divided into pools of only 100 individuals each, as recently advocated [85, 97], and all amplified in triplicate, DNA pooling decreases genotyping by a factor of 30 in studies involving 500 cases and 1000 controls.

Our results confirm that the accuracy of analyzing corrected pool patterns generated from microsatellites approaches that of individual genotyping. Particularly in complex disorders, where the association of marker alleles with disease loci is likely to be only moderate or weak, a gain in sensitivity with stutter correction in pooled analyses justifies the limited amount of extra genotyping required to create a small training set.

In conclusion, we have demonstrated that accurate estimates of microsatellite allele frequencies from DNA pools are feasible with a novel stutter correction method requiring one standard training set of only ten additional individual genotypes. This method opens the way for realistic large-scale genetic association studies using microsatellite markers.

Acknowledgements

We acknowledge the special contribution of Lodewijk Sandkuijl in formulating some of the basic concepts embodied in this work; he died shortly before completion of the manuscript. Further thanks go to Martine van Belzen for providing the celiac disease samples and individual genotypes, and to Jackie Senior for critically reading the manuscript.

Eur J Hum Genet (2004) 12:925-934 (reprinted with permission).

10 SINAPSE: SINGLE-PCR SNP DETECTION USING STANDARD DNA SEQUENCING EQUIPMENT

S.C. Bakker, J.C.J.M. Hendriks, P.H.A. van Zon, P.L. Pearson and R.J. Sinke

Abstract

Single nucleotide polymorphisms (SNPs) are widely used as DNA markers in genetic studies. Current techniques for large-scale SNP typing require considerable initial investments in equipment, while techniques that use widely available equipment are generally more labour-intensive and more expensive per sample.

We present a single-PCR, efficient SNP genotyping method on standard DNA sequencers. SNP discrimination is based on allele-specific primers with a high-affinity 3' locked nucleic acid (LNA) to enhance specificity. A tail sequence allows a universal labelled oligonucleotide to be incorporated during PCR, which reduces assay cost considerably. SNP alleles are amplified using one standard PCR protocol, and separated on a DNA sequencer under the usual conditions for microsatellite genotyping. A four base pair spacer sequence allows allele identification by size. Primer design and data analysis are facilitated by two freely available spreadsheets.

The method was tested on 36 candidate SNPs. Specific primer sets could be designed for 35 SNPs, 31 of which were successfully amplified without optimization. Pooling of PCR product before separation allowed simultaneous analysis of up to four SNPs, which further increased throughput. For 13 SNPs genotyping accuracy was verified in 25-49 individuals, using the existing TaqMan technique or direct sequencing. On 298 genotypes there was one discrepancy between two genotyping methods. In amplified DNA pools, signal intensity was proportional to the amount of alleles present in the pools, which makes the technique suitable for accurate determination of allele frequencies in DNA pools.

In conclusion, this new SNP typing technique provides flexible and inexpensive medium- to high-throughput SNP genotyping on standard sequencing equipment.

Introduction

Single nucleotide polymorphisms (SNPs) are DNA variations at a single base position, which are widely used as genetic markers. In ongoing large collaborative efforts, several millions of these abundant polymorphisms have already been identified in the human genome. Since association of a genetic marker with a disease locus is expected to be detectable over small distances only, many SNPs have to be available for the screening of candidate regions or –genes. Efficient, inexpensive SNP detection methods are therefore essential for the success of studies involving many samples and markers.

The choice of genotyping method will depend on the numbers of SNPs that need to be typed, and the availability of equipment. Multiple different SNP detection methods have recently been published, each with specific advantages and drawbacks [16, 17]. There are platforms for high-throughput SNP detection at low costs per sample, such as mass-spectrometers or micro-array based methods, which require large initial investments for equipment. SNP detection on equipment that is already available in many genetic laboratories, such DNA sequencers or real-time PCR machines, have other drawbacks. These methods are generally labour-intensive, requiring several PCR- and purification steps, or optimization of conditions for each SNP to be typed. Many methods require labelled primers, probes or nucleotides, which substantially increase the costs per genotype.

We have developed a flexible, medium- to high-throughput SNP typing method on standard DNA sequencing equipment that requires only one standard PCR step with unlabelled, high-affinity allele-specific primers. The accuracy of the method was tested by comparing genotypes with those obtained using the existing TaqMan technique or direct sequencing. In addition, the usefulness of the technique for DNA pooling experiments was investigated. Two Microsoft Excel-based spreadsheets were developed to facilitate the design of allele-specific PCR primers, and to analyze the data.

Materials and methods

Samples. DNA was isolated from peripheral blood lymphocytes using established procedures. Analyses were performed on anonymous samples from individuals that were referred to our Diagnostics department, as well as anonymous controls and individuals from CEPH reference families, the DNA of which was isolated from cultured cell lines.

SNP selection. Validation experiments were based on possibly disease-related mutations, which had previously been identified at our Diagnostics department. Other SNPs were selected for candidate gene studies from dbSNP or the Celera database [2].

Principle of the SiNaPse technique. SNP discrimination is based on allele-specific amplification by two allele-specific primers with a SNP-binding 3' nucleotide, and one reverse primer. One of the allele-specific primers has a 4 base pair (bp) spacer sequence that allows the discrimination of alleles by product size. The reverse primer has a universal tail sequence, which allows the incorporation of a separate oligonucleotide with a fluorescent label (labelled tail) during the same PCR reaction. PCR will produce labelled allele-specific PCR products that can be separated by size using electrophoresis (see Figure 10.1). SiNaPse is an acronym for SNP Investigation using Non-labelled Allele-specific Primers and Separation by Electrophoresis.

Design of SiNaPse PCR primers. A Microsoft Excel Spreadsheet was developed to facilitate primer design (available on request).

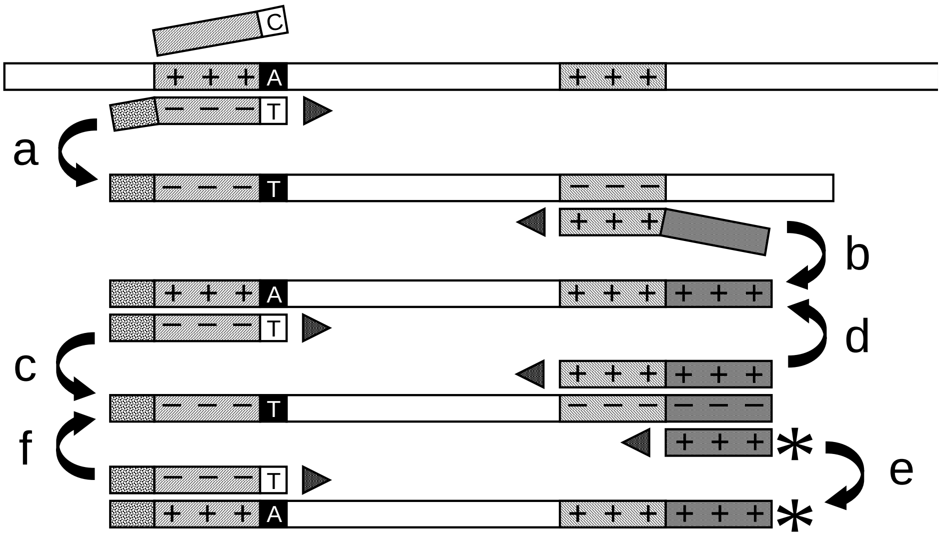


Figure 10.1. Principle of the SiNaPse PCR reaction.
In this example, a G/A SNP is located on the plus strand (+) of the genomic template DNA (top of figure). The direction of primer elongation is indicated by grey triangles, while curved arrows indicate the formation of the resulting PCR products. If an individual is homozygous for the A allele, only the allele-specific primer with a 3' T can anneal to the template; this primer has a 4-bp spacer sequence at its 5' end. In reaction (a), this primer produces a product (-), to which the reverse primer with the tail sequence can now anneal. Elongation of this primer in reaction (b) results in a product (+), to which the allele-specific primer can once more anneal. The next reaction step (c) results in a product that incorporates a sequence that is complementary to the tail sequence (-). The reverse primer and the labelled tail will now compete for this binding site, and two parallel reactions will take place. The reverse primer will form more unlabelled product through steps (d) and (c), while the labelled tail is incorporated in a fluorescently labelled PCR product through reaction (e). The labelled tail is present in a 50-fold higher concentration than the reverse primer, which will consequently be exhausted after a few cycles. This will terminate reaction (d), and the PCR will exclusively proceed through reaction (e), which forms fluorescently labelled product.
If the G-allele is also present, the other allele-specific primer, with a 3' C, will in a similar reaction produce a second labelled product that lacks the 4-bp spacer sequence. These products can then be separated through their 4-bp size difference.

Allele-specific primers. For each SNP, a pair of allele-specific primers with a melting temperature (T_m) of 60°C +/- 3°C was sought on either the plus or the minus DNA strand flanking the SNP. Allele-specific primers were ordered as TrueSNP oligos (Proligo, Paris, France), which have a 3' locked nucleic acid (LNA), complementary to the respective SNP alleles. LNA was recently reported to bind to DNA with a high affinity en to result in highly specific amplification [153, 154]. The increase in primer annealing temperature caused by the 3' LNA was determined using the Exiqon web tool. To one of the allele-specific sequences, a four base pair spacer sequence was added at the 5' side, with a sequence identical to the genomic sequence at that position, to prevent annealing. Both primers were further extended at the 5' side with a 7-bp 'pig tail' sequence (GTGTCTT) that reduces plus-A artefacts [278]. This addition is not essential, but we found that the genotyping software can more easily distinguish the resulting sharper peak patterns.

Reverse primer. Allele-specific primer sequences were entered as either right or left primer into the Primer3 program [152], which suggested reverse primers of the desired product size and T_m. Forward and reverse primers were blasted against the human genome sequence using the Ensembl database, and if no specific primers could be found, alternative primers on the opposite DNA strand were chosen. The reverse primer was extended on the 5' side with a universal tail sequence (5'_TGGTAAAACGACGCCGAC_3'), based on the phage M13-21 sequence. Several modifications were made to improve the physical characteristics, which were tested with the Netprimer web tool (PREMIER Biosoft International, Palo Alto, CA, USA) and to reduce non-specific binding (investigated by blasting candidate sequences against the human genome sequence).

Labelled tail. The fourth primer in the reaction consists of the tail sequence only, which is labelled with the fluorescent HEX dye at the 5' end.

PCR conditions. PCR reactions were performed on a Gene Amp PCR system 9700 (Applied Biosystems, further called ABI), in 384-well plates. Each 5 µl volume contained 10 ng template DNA, 200 mM of each dNTP (Amersham Biosciences), and 0.2 units Amplitaq Gold (ABI), 1 × PCR buffer II (ABI), 2.5 mM $MgCl_2$ (ABI), 400 nM of each allele-specific primer, 400 nM HEX-labelled tail and 8 nM reverse primer. DNA was initially denatured at 94°C for 7 min and was then subjected to 9 cycles of 94°C for 30 s, (66-57°C, with 1°C decrease in temperature per cycle) for 30 s and 72°C for 1 min, followed by 30 cycles of 94°C for 30 s, 57°C for 30 s, 72°C for 1 min, and a final extension step of 72°C for 10 min.

Electrophoresis and analysis. PCR products were pooled in equal amounts, or single products were diluted 4 times with 1x PCR Gold buffer. To 1 µl of product 4 µl HiDi and 0.02 µl ROX size standard were added, and this mixture was separated in POP 6 polymer on an ABI 3700 capillary sequencer (ABI). Data were analyzed with Genescan 3.5 and imported into Genotyper 3.7 (ABI). If one or both allelic peaks were not between 250 and 6000 units, both labels were removed. Peak heights were imported into an Excel spreadsheet that facilitates genotyping by graphically displaying the data and converting peak heights to genotypes (available on request).

DNA pooling experiments. DNA from one individual that was homozygous for the C-allele of SNP rs2619522, and from one individual homozygous for the T-allele was diluted to approximately 10 ng/µl and the concentration was measured in triplicate with Pico Green (Molecular Probes, Leiden, the Netherlands) on a Genios plate reader (Tecan, Männedorf). DNA from two different homozygous individuals was combined in ratios of 0:10 to 10:0. Reactions were performed in triplicate, and peak heights were averaged. Experiments were replicated with two independent homozygous individuals.

TaqMan analysis. Genotyping was performed on a 7900HT TaqMan system according to the recommendations by the manufacturer.

Comparison of genotyping results. Genotyping of the same individuals using different techniques was performed independently, and results were later compared.

Results

Incorporation of the fluorescent label will depend upon competition of the labelled tail with the unlabelled reverse primer. A relatively lower concentration of the reverse primer is therefore likely to favour the production of labelled product, but a certain amount of reverse primer is necessary to start the PCR. We therefore first tested different ratios (up to 1:800) of reverse primer and labelled tail, and found that a ratio of 1:50 for reverse primer and labelled tail, respectively, resulted in the strongest signal (data not shown). This ratio was used in all subsequent experiments.

In total 36 SNPs had previously been selected as relevant for DNA diagnostics or candidate gene studies in psychiatric disorders. Table 10.1 shows the results of SiNaPse analysis of these SNPs. Twenty-two SNPs were specifically tested for amplification efficiency and genotyping accuracy, while 14 additional SNPs were later analyzed in association studies involving up to 300 cases and 600 controls. For the latter studies, we know that the quality of DNA samples is less constant than in the validation experiments, and dropout rates were indeed somewhat higher in general (data not shown).

For all SNPs except DAX1g315c, primer sets could be designed for the standard 60°C PCR protocol, and with good predicted specificity. Thirty-one primer sets (86.5%) produced specific PCR products of the expected size, with clearly distinguishable peaks (see Figure 10.2).

Up to four different PCR products were pooled before electrophoresis, which still resulted in signals of sufficient intensity (Figure 10.3). The two alleles of a SNP will have fixed sizes, with only a 4 bp difference, and therefore primer sets can be designed to obtain PCR products with only small (10 to 15 bp) size differences between SNPs. For the SNPs that could be amplified, the dropout rate due to failure of individual samples was 4.7% on average in the validation experiments.

Table 10.1. SiNaPse genotyping results

Variation	Gene	PCR[a]	Alleles[b]	Individuals	Dropout	Identical[c]
RGS4_18	RGS4	+	2	24	12.5%	95.2%
RGS4_1_b	RGS4	+	2	24	12.5%	100.0%
hCV11558870	PIP5K2A	+	2	24	8.3%	100.0%
hCV9591220	PIP5K2A	+	2	24	8.3%	100.0%
B2a7470g	BRCA-2	+	2	24	4.2%	100.0%
B2g203a	BRCA-2	+	2	24	4.2%	100.0%
B1a4956g	BRCA-1	+	2	24	0.0%	100.0%
B1t4427c	BRCA-1	+	2	24	0.0%	100.0%
B2a1093c	BRCA-2	+	1	24	0.0%	100.0%
B2a1593g	BRCA-2	+	1	24	0.0%	100.0%
B2c1342a	BRCA-2	+	2	24	0.0%	100.0%
B2t2457c	BRCA-2	+	1	24	0.0%	100.0%
HFE	HFE	+	2	49	8.2%	100.0%
B1a1186g	BRCA-1	-	n.a.	24	n.a.	
rs1051332	ATP7b	+	2	27	0.0%	
rs1801249	ATP7b	+	2	27	0.0%	
rs732071	ATP7b	+	2	30	3.3%	
hcV3123355	ATP7b	+	2	31	0.0%	
rs2277448	ATP7b	+	2	31	0.0%	
rs754610	ATP7b	+	2	31	0.0%	
MSX1t1140c	MSX	-/+	2	49	32.7%	
HFEg6722a	HFE	-	n.a.	49	n.a.	
BDNF_270	BDNF	+	1			
BDNF1	BDNF	+	2			
DAX1g315c	DAX	-	n.a.			
hCV9012157	PIP5K2A	-	n.a.			
NRG221533	NRG1	+	2			
RGS4_1	RGS4	+	1			
RGS4_4	RGS4	+	1			
RGS4_7	RGS4	+	2			
rs1047552	PSFL	+	2			
rs1472500	PSFL	+	2			
rs2619522	dysbindin	+	2			
rs3213207	dysbindin	+	2			
rs3743319	PSFL	+	1			
rs760761	dysbindin	+	2			
TsC0123902	PIP5K2A	+	2			
Average		13.5%	18.9%	28.9	4.7%	99.6%

[a] Observed specific PCR product at the expected size.

[b] Number of different alleles observed in the tested samples.

[c] Percentage of genotypes in agreement in SiNaPse vs. TaqMan / direct sequencing. The first four SNPs were verified using direct sequencing, the others by TaqMan analysis.

NOTE: SNP RGS4_1 was tested using two sets of primers on complementary DNA strands, one of which seemed to amplify only one allele.

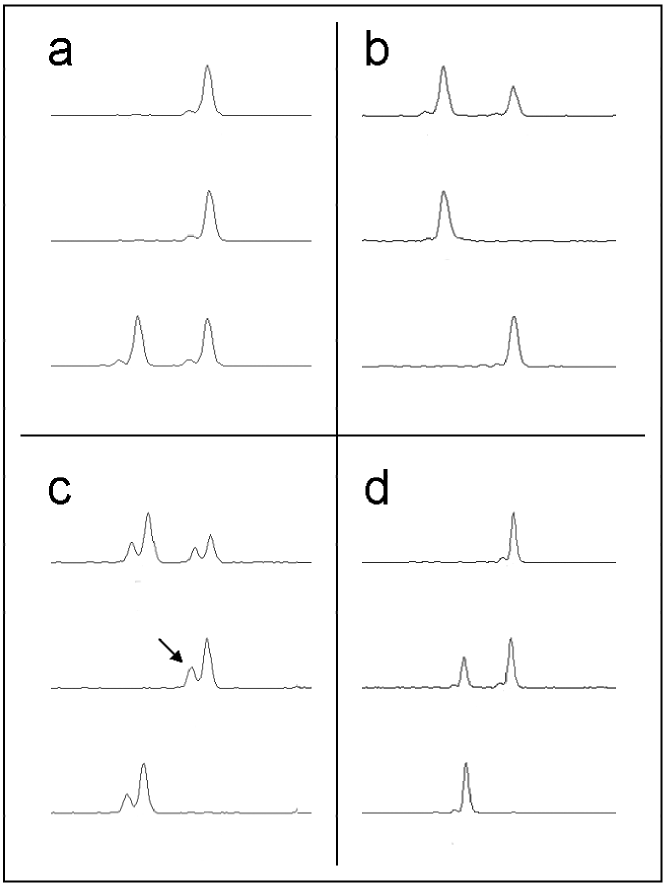


Figure 10.2. Representative examples of raw genotyping data for four different SNPs. a: BRCA2a7470g (only one type of homozygote observed); b: MSX1t1140c; c: B1a4956g; d: B2c1342a. The arrow in Figure c indicates a plus-A artefact, resulting in an additional peak at 1 bp from the allelic peak.

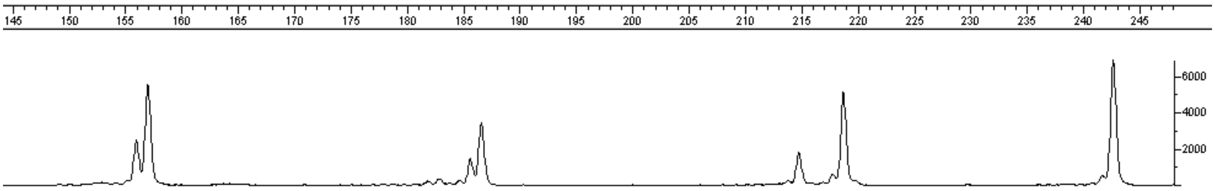


Figure 10.3. Electrophoresis pattern of pooled PCR products from four SNPs. Fragments from different SNPs have sizes of approximately 155, 185, 215 and 245 bp. NOTE: all depicted genotypes are homozygotes, except for the SNP with product sizes around 215, which shows two alleles separated by 4 bp.

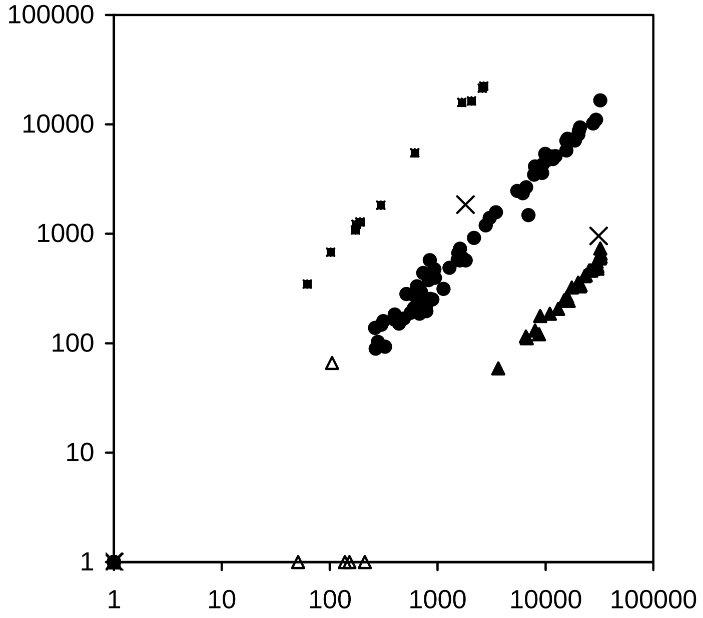


Figure 10.4. Graphical representation of genotyping results for SNP BDNF-1. x-and y-axis units indicate fluorescence intensity, measured by the sequencer. Solid circles indicate heterozygous individuals, while squares and triangles represent the two types of homozygotes. Crosses and open triangles represent measurements with aberrant peak heights or -ratios. NOTE: in order to show multiple measurements per genotype, results of an experiment with 192 individuals are shown, which were run on an ABI 3730 sequencer (units for peak heights are different from the 3700 sequencer).

Amplification by the SNP-matching primers was generally efficient, while the alternative primers formed little or no detectable product. The ratio of peak heights between the two alleles of heterozygous individuals was highly reproducible, although different for each SNP. Clusters of different genotypes could therefore be clearly separated (see Figure 10.4).

Running undiluted product was found to overload the detector of the sequencer, with flattening of the signal for the allelic peak, and a relative increase for the other PCR reaction, which diminished the allelic discrimination. Products were therefore always run using a standard 1:4 dilution with 1x PCR buffer, or they were pooled before the analysis.

Seven SNPs only seemed to display one allele. Three of these SNPs were also genotyped using a TaqMan system, which confirmed that they were not polymorphic in the sample. In SNP RGS4_1, we found evidence for failure of amplification of the second allele, since this allele was present after using a primer set on the alternative strand.

Genotyping accuracy was tested by comparing SiNaPse results for 24-49 individuals per SNP with those obtained on a TaqMan system (10 SNPs) or by direct sequencing (4 SNPs). On a total of 298 compared genotypes (596 alleles), there was one discrepancy, in which SiNaPse and TaqMan analyses identified one sample as homozygous for different alleles.

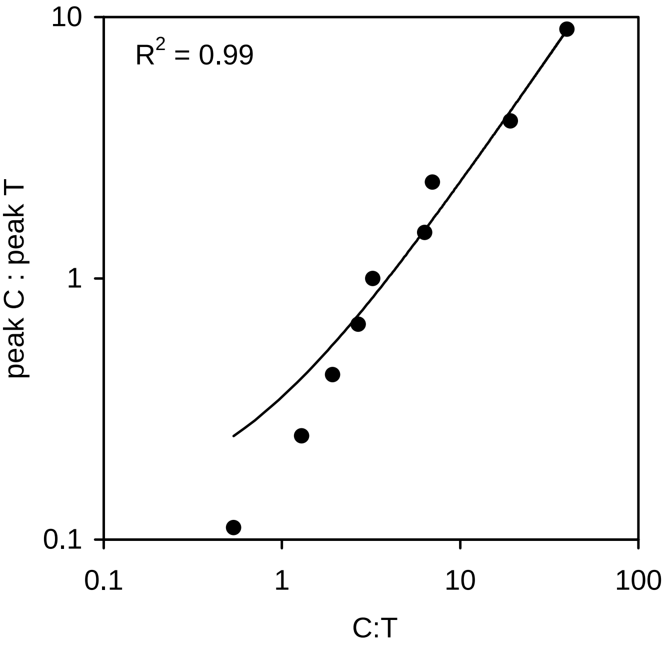


Figure 10.5. SiNaPse result obtained from pooled DNA samples. Different amounts of DNA from individuals homozygous for the C or T-allele of SNP rs2619522 were combined, and signal intensities for both alleles were measured. Averaged values for triplicate experiments are shown. X-axis: ratio of the amount of C and T allele present in the pool. Y-axis: ratio of allelic peak heights. Data are plotted on a logarithmic scale, with a linear trend line shown.

To investigate the usefulness of the technique for DNA pooling, DNA from two individuals that were homozygous for different alleles of the same SNP was added in different ratios. There was a linear increase of the relative peak height of the respective alleles with the amount of the alleles present in the pool (see Figure 10.5)

Discussion

Many different reliable techniques for SNP detection have been described. However, some methods require large initial investments in equipment or are only cost-efficient for the simultaneous analysis of hundreds of different SNPs. Other methods are labour-intensive and relatively expensive per genotype, because they require labelled nucleotides, primers or probes. The presented SiNaPse method brings together several established techniques, in order to develop a flexible and efficient SNP detection method on widely available DNA sequencing equipment. The technique is based on a single allele-specific PCR with a modified, high-affinity SNP binding nucleotide. During the PCR under standard conditions a universal fluorescent label is incorporated into the product, which allows semi-automatic detection on DNA sequencers. Although most experiments were performed on an ABI 3700 sequencer, comparable results were obtained with the newer ABI 3100 and 3730 sequencers, and the method is likely to be suitable for slab gel machines as well.

Allele-specific PCR, which is based on primers with a SNP-matching nucleotide at the 3' end, has been used for many years [280]. Several published SNP detection methods are based on allele-specific PCR primers, with different readout systems [281-283]. The specificity of the amplification will depend on both binding affinity and successful elongation by the polymerase. Unfortunately, most allele-specific PCRs require optimization of conditions for each reaction, which limits the use for high-throughput genotyping. We aimed to enhance the competitive advantage of the SNP-matching primer by using a touchdown PCR protocol during the first cycles. The reaction starts with a T_m at which none of the primers can anneal, followed by a stepwise reduction of the temperature during the first PCR cycles, which will allow the primer with the highest affinity to anneal one or several cycles before the alternative primer. In order to further increase specificity, primers were ordered with a 3' locked nucleic acid (LNA) at the position of the SNP. LNA was recently reported to bind DNA with a high affinity, and to result in highly specific amplification [153, 154]. Indeed, we found that nearly all reactions could be performed under standard conditions, while allele-specificity was high. We also tested one primer set with and without LNA, and found that alleles that could not be accurately distinguished using regular primers, gave good results with LNA primers (data not shown). Therefore, primers with a 3' LNA, which are hardly more expensive than regular primers, appear to be useful in enhancing the specificity PCR reactions.

Failure of amplification of one allele is always a theoretical possibility. In one SNP, we found evidence for failure of amplification of one allele, and it cannot be excluded for three other SNPs. However, SNPs were not selected for a high minor allele frequency, and with the relatively low number genotyped chromosomes per SNP, the minor allele may simply not have been present. This is illustrated by the observation that the three apparently monomorphic SNPs that were also genotyped using TaqMan were indeed not polymorphic in the tested sample. In case one of the alleles should fail to amplify, however, the problem will usually be readily detectable, and alternative primers could be defined.

The reverse primer tail sequence allows using one labelled universal tail primer for all SNPs, instead of labelling each SNP-specific primer separately, which reduces primer costs considerably. The method has been described for genotyping microsatellites in several independent papers, which do not seem to have received much attention [251, 284]. We have typed many different microsatellites using this method, and the results are comparable to those obtained by regular labelled primers. The recently described Amplifluor SNP detection method on plate readers uses two different tail sequences and two dyes in a comparable fashion [281]. However, even though tail sequences are defined in such a way as to reduce non-specific binding, it is possible that part of a tail sequence by chance will anneal to the template DNA. The use of two different tails in one reaction carries the risk of annealing differences between the tails that could result in different melting temperatures of both allele-specific primers. Separating alleles by size instead of colour allows the same reverse primer and tail to be used for both alleles, which circumvents this problem.

Efficient genotyping requires minimal optimization and sample handling, and reliable allelic discrimination. Therefore, we have investigated the success rate of primer design and SNP amplification under standard conditions, and compared genotyping results with those obtained by established methods.

Designing primer sets. Allele-specific primers have their 3' end at the SNP, and their position is therefore fixed. The SiNaPse method allows choosing primer sets on both DNA strands, which increases the flexibility of primer design. The Microsoft Excel spreadsheet facilitates the design of primers with the required T_m, tail sequence and spacer sequence. For only one out of 36 SNPs, no primer set with the desired properties could be designed.

Efficiency of amplification. Overall, we obtained clearly distinguishable allelic peaks with 86.5% of the primer sets using the standard PCR protocol. For single SNPs, the dropout rate of less than 5% of analyzed samples on average seems acceptable.

Allelic discrimination and accuracy of genotyping. In general, amplification was allele-specific, with an occasional sample with an aberrant allelic peak ratio, and genotype clusters could therefore usually be clearly distinguished. It is important not to overload the camera of the sequencer, since this reduces the separation of alleles. Genotyping results for 14 SNPs were verified using the established TaqMan assay or by direct sequencing. On a total of 298 compared genotypes, there was only one sample with a genotyping discrepancy. The reason was not clear, since both sequencing data and SiNaPse data seemed reliable. Quite unexpectedly, both techniques found the sample to be homozygous for a different allele, which may indicate an error in sample handling. Overall, these data suggest that the allele-specific PCR produced reliable genotyping results.

DNA pooling. Pooled analysis of DNA samples is an efficient way to screen large numbers of cases and controls for differences in allele frequencies [84, 85]. When different amounts of DNA from homozygous individuals for one allele were combined and analyzed, we found a linear increase of peak heights with the amount of alleles present in the DNA pool. These results indicate that the SiNaPse technique is suitable for DNA pooling experiments. As with most methods, when alleles are present in equal amounts, one of the alleles will usually give a stronger signal than the other. Therefore, if one is interested in exact allele frequencies instead of differences between pools, peak heights should be corrected with a factor that is derived from measured peak heights in a number of heterozygous individuals [86].

Suitability for high-throughput genotyping. The use of one standard PCR without additional purification steps greatly reduces sample handling, but for high-throughput genotyping, genotyping capacity and low costs per sample are also essential. Signal intensities of undiluted PCR products were generally high. Although we did not test this, pooling of more than the tested 4 markers is likely to be possible (note that sequencer running conditions are identical for SiNaPse and microsatellites, and we found that the two types of markers can also be run together, as long as allele sizes do not overlap- data not shown). If we assume pooling of six SNPs per run, a capillary machine such as the ABI 3700 will have a capacity of approximately 7.900 genotypes per day (11,800 genotypes for an ABI 3730 machine). For unpooled samples, the costs per genotype were € 0.41 per genotype, including all costs for primers, PCR and run on an ABI 3700. By multiplexing reactions, pooling of SNPs and using a capillary sequencer like the ABI 3730, costs per genotype can be further reduced. We have recently used the method for candidate gene studies involving multiple SNPs in up to 900 samples, with satisfying results (data not shown).

In conclusion, we have developed a technique that allows flexible medium- to high-throughput SNP genotyping on standard DNA sequencing equipment that is already present in many laboratories. This method is a useful addition to very high-throughput platforms, which are currently still affordable for large centres only.

Acknowledgements

This project was supported by the UMC Utrecht's Genvlag program, the Makaria Foundation and ABC Neurogenomics.

11 DISCUSSION

Family studies have consistently shown that many common psychiatric disorders are genetically complex, i.e. the result of multiple heritable and environmental risk factors. Variations in the heritable material probably determine the major part of the total susceptibility to psychiatric disease. Knowledge about the nature of these DNA variations should provide insight into the causes of these disorders, and, perhaps more importantly, open ways to new forms of prevention and treatment. In theory, disease-related variations should be detectable by comparing the entire DNA in large numbers of individuals with and without a disorder. Unfortunately, this is technically not feasible. Instead, current genetic linkage and association studies analyze a selection of all genetic variations, and use these as markers that could indicate the presence of nearby disease-related variations. The underlying assumption is that a disease-related variation arose in a single person some time in history, and that this variation, as well as the DNA region surrounding it, is still shared by living individuals with the disease.

This thesis describes genetic studies in schizophrenia and ADHD, carried out in large and homogeneous patient samples. In addition, it presents new techniques for the efficient analysis of microsatellite markers in pooled DNA from many individuals, and for the analysis of single nucleotide polymorphisms (SNPs) on standard genotyping equipment. The first part of this chapter will discuss the results of the studies and relate them to findings by others. The second part will be a more general discussion of factors that may influence the chances of finding disease-related variations in complex disorders, and how these factors may guide the design of future studies. Finally, the possible implications of progress in genetics for understanding and treating psychiatric disorders will be explored.

11.1. Genetic studies in schizophrenia

The causes of schizophrenia remain basically unknown, but segregation studies in families have consistently shown that heritable factors play a major role in the development of the disorder [48].

Given this high genetic contribution, it may seem surprising that the identification of schizophrenia susceptibility genes has proved to be difficult. We hypothesized that both clinical and genetic heterogeneity may have contributed to these difficulties. Most likely, there are multiple genetic factors involved, and different genetic variations may predominate in different populations. In addition, clinically defined schizophrenia includes different classes of symptoms, some of which are also shared by other psychiatric disorders. Specific symptoms may or may not be present in individual patients, and the highly variable course of the disorder suggests the presence of disease subtypes with different causal factors. The identification of such factors in broadly defined schizophrenia samples may then be difficult.

Considering this, we attempted to collect patients with a homogeneous clinical and ethnical background, and performed genetic studies in patient subgroups with and without prominent negative symptoms, as defined by the criteria for the deficit syndrome [36].

Functional candidate genes

Genes involved in neurotransmission, such as those belonging to the dopamine, serotonin and glutamate systems, are obvious candidates genes for schizophrenia, since drugs used to either alleviate or induce symptoms of psychosis are known to influence mainly neurotransmitter receptors. Other genes involved in the synthesis, transport and breakdown of these neurotransmitters are therefore equally interesting candidates. **Chapter 2** describes a systematic screening of 12 genes with a function in dopamine neurotransmission, using microsatellite markers and a new DNA pooling technique that is described in chapter 9.

Although there were some indications for involvement of the dopamine D5 receptor in the pooled analyses, individual genotyping in a larger sample could not confirm these results. This suggests that dopaminergic genes do not play a major role in schizophrenia, at least not in the Dutch population. It is possible, however, that we have missed variations with small effects. In the first place, at the time of the study, patient inclusion was still ongoing, and a sample of slightly over 200 cases provides only modest power for detecting genes with a low relative risk. The use of a DNA pooling technique may have further reduced statistical power [98]. Second, even though our aim was to select several microsatellite markers within or close to each gene, these markers may not have been in sufficient linkage disequilibrium with a possible disease-related variant. Finally, we cannot exclude that some dopaminergic genes are only associated with specific symptoms of schizophrenia, and that we failed to detect these in an unselected sample.

Taken together, however, the results of different candidate gene studies, including ours, indicate that if dopaminergic or serotonergic candidate genes contribute to schizophrenia, their effect must be small. Meta-analyses of multiple studies have indicated a very modest contribution (with typical RRs around 1.5) of the serotonin 2a (5HT2a) receptor [49, 50], the dopamine D2 [51] and D3 receptors [52-54], and the COMT gene [55]. These results are in agreement with findings from linkage studies, which do not show much evidence for linkage in the regions containing the most obvious candidate genes. Many other negative or inconsistent studies of functionally interesting genes indicate that the obvious candidates do not explain most of the genetic susceptibility to schizophrenia.

Positional candidate genes

Genetic linkage studies do not make assumptions about gene function, since they screen the entire DNA for regions shared by family members with a disorder. The results from multiple whole-genome linkage studies in schizophrenia at first sight suggest linkage on almost every chromosome. As in other complex disorders, however, some regions have emerged more frequently than others. Recent meta-analyses of available data have confirmed this impression and mean it is highly unlikely that the observed pattern of linkage is the result of chance alone [63, 64]. In 2002, a group from Iceland gave a new impulse to the genetics of schizophrenia by screening a region of repeated linkage on chromosome 8 with a high-density marker set [66]. A certain combination of marker variants (a haplotype) in the neuregulin 1 gene was seen to be associated with schizophrenia. However, the haplotype was not unique to schizophrenia patients, and increased the risk of developing schizophrenia only 1.5 to 2 times. In **chapter 3** we have investigated the role of the neuregulin 1 gene in patients with deficit and non-deficit schizophrenia. The SNP that was most strongly associated in previous studies was also associated with schizophrenia in the Dutch population, but with a different allele. This finding provides support for neuregulin 1 as a schizophrenia susceptibility gene, but it also indicates there is genetic heterogeneity in different populations, as has been reported for other genes [132, 135]. Interestingly, we detected association in the non-deficit group, but not in deficit patients. These results provide evidence for the existence of genetically distinct schizophrenia subtypes, and they suggest that careful selection of patients based on symptoms could increase the chances of success in finding schizophrenia susceptibility genes.

Similar association studies in other linkage regions soon followed the first report on the neuregulin 1 gene. Four of these genes, dysbindin, G72/G30, RGS4 and PIP5K2A, are investigated in **chapter 4**. In the dysbindin and G72/G30 genes, no indications for association with either non-deficit or deficit schizophrenia were found. This could mean that neither gene contributes substantially to schizophrenia in our population, although it is also possible that other markers than those reported in previous studies are associated. Even between European populations, the associated dysbindin haplotypes seem to be different [132, 135]. A recent study in a Scottish sample reported no association with markers in dysbindin from the first published studies [133]. Later, however, new haplotypes that were associated in an English sample also showed strong evidence for association in the Scottish population [132]. In the presence of such allelic heterogeneity, one should perhaps be careful before concluding that a certain gene is not associated, and instead only exclude the association with specific haplotypes.

In RGS4, we confirmed the association with a haplotype that was first found to be associated with both schizophrenia and bipolar disorder [144]. Again, the association was mainly due to the group of non-deficit patients. The results of the available studies suggest that RGS4 is related to symptoms or traits that are common to schizophrenia and bipolar disorder. As a regulator of G-protein signalling, this gene modifies the function of dopamine and glutamate receptors. Recently it was suggested that glutamate neurotransmission may be central in the development of schizophrenia, since neuregulin 1, dysbindin, G72/G30 and RGS4 may all be involved in this system [70]. Interestingly, the first indications for involvement of RGS4 came from gene expression studies [69]. This underlines the importance of using biological information from related disciplines in genetic studies.

We found strong association of two SNPs in the PIP5K2A gene with schizophrenia, in both the deficit and non-deficit groups. Interestingly, the region on chromosome 10 in which this gene is located, repeatedly showed linkage to both schizophrenia and bipolar disorder [148]. Lithium, the substance most widely used to treat bipolar disorder, influences the phosphatidylinositol pathway, which includes PIP5K2A. Like RGS4, PIP5K2A may therefore be a susceptibility gene for both disorders. A recent study reported an interaction between RGS4 and phosphatidylinositol-3,4,5,-trisphosphate (PIP3), also part of the phosphatidylinositol pathway, in cardiac myocytes [285]. This suggests that the phosphatidylinositol pathway, G-protein signalling and dopamine and glutamate neurotransmission are all coupled, and possibly interact in increasing susceptibility to schizophrenia. It would be interesting to screen other genes from these pathways for association with schizophrenia and bipolar disorder as well.

A special sample

Several measures were taken to collect a homogeneous patient sample. Patients were only included if they had at least three grandparents of Dutch, Caucasian ancestry, in an attempt to reduce possible genetic heterogeneity introduced by the large-scale immigration to the Netherlands during the past decades. In order to reduce clinical heterogeneity, the same trained rater verified the diagnosis of schizophrenia in all patients, using a standardized diagnostic interview (CASH, [117]). Patients with schizoaffective disorder, who have prominent mood symptoms as well as psychotic symptoms, were excluded, since they may share biological causes with patients with mood disorders [286]. It is difficult to prove how effective this selection has been, but the relatively strong association signals, compared to other studies involving samples with comparable sizes, suggest that selecting an ethnically and clinically homogeneous sample may be worthwhile.

Patients from long-stay wards of psychiatric hospitals were specifically included, which resulted in a high proportion of patients with deficit schizophrenia. In an unselected sample of schizophrenia patients, approximately 15% of first-episode patients, and 25-30% of chronic patients will meet the criteria of the deficit syndrome, compared with over 50% in our sample. Our results may therefore not be fully representative for schizophrenia in an unselected population sample. We used the Schedule for the Deficit Syndrome (SDS) to define a subgroup with negative symptoms, because it seemed a more precisely defined measure than other proposed clinical criteria, such as years of hospitalization, or the persistent inability to take care of oneself [35]. The SDS has certain problems, however. For instance, in some patients, it may be difficult to exclude the influence of factors like depression or the use of drugs as the cause of negative symptoms. Moreover, the SDS requires the evaluation of negative symptoms during the past year, which was provided by the patient's own physician. There is not much information, however, about the inter-rater reliability of the SDS. These factors may therefore have introduced some noise into the comparison of genetic variants in deficit and non-deficit patients. If so, it may have led to an underestimation of differences in allele frequencies between the groups, but not necessarily to false-positive findings. Still, the distinct differences in allele frequencies between the groups for neuregulin 1 and RGS4 seem to confirm that the SDS delineates two groups with different characteristics. Until more basal endophenotypes become available, deficit schizophrenia may be a useful clinical subtype for genetic studies.

Our schizophrenia sample has certain limitations. In the first place, even though the total number of included patients now exceeds 300, the sample size may be relatively modest for future large-scale association studies. Therefore, possibilities for expanding the sample are currently being studied.

Second, we do not know parental genotypes, and haplotypes must therefore be estimated using allele frequency data. This is a generally accepted procedure, but genotyping errors can substantially affect haplotype frequency estimates and the power to detect association with them [287]. We decided to collect a large sample of unrelated patients, when we experienced that collecting a sib pair sample of sufficient size would probably be beyond our reach. At the time, however, population stratification was seen as a major threat to association studies, and TDT analysis using parents was regarded as the optimal study design. The case-control design has become generally accepted recently, and with hindsight, the decision to collect a large case-control sample seems to have been right, given our resources. It should be noted, however, that phase known haplotypes would be valuable for future studies, and that parental DNA should preferably be collected as well.

Third, there is little information about the background of the control sample, which consisted of anonymous blood bank donors, supplemented with control samples from the Department of Biomedical Genetics, UMC Utrecht. Consequently, population stratification cannot be ruled out as an explanation for different allele frequencies between the samples. However, the prevalence of the disorder is largely comparable in different populations worldwide, and therefore, even if different subgroups were present, stratification is perhaps less likely to have caused false-positive results. In retrospect, another argument against stratification effects is that we have now typed a large number of markers in the sample, mostly with comparable allele frequencies between cases and controls. There are formal ways to quantify stratification, however, and to correct for its presence [29, 30]. These methods, which basically involve comparing a substantial number of random markers throughout the genome in cases and controls, should be considered for future studies in this sample.

Relevance and future studies

As a whole, our findings have confirmed the involvement of several recently reported susceptibility genes in the development of schizophrenia. The available data suggest that susceptibility to schizophrenia is, at least in part, the result of common variations, as suggested by the common disease-common variant hypothesis. Interestingly, we found the first indications that some genes are involved in only non-deficit schizophrenia. This finding has important consequences. In the first place, stratification of study samples may indeed be beneficial to genetic studies. Genes involved in deficit schizophrenia are likely to go unnoticed in unselected samples, whereas this relatively large subgroup with severely disabling symptoms certainly deserves separate study. New genome scans in deficit samples, or re-analysis of combined data from existing genome scans, taking into regard negative symptoms, my be necessary to indicate new candidate regions. Ultimately, the genetic distinction of patient groups with specific symptoms may allow better prediction of the disease course in individuals, and lead to targeted treatment. Our sample is relatively unique in having a high proportion of deficit patients, and we are currently investigating other candidate gene systems for a different involvement in deficit and non-deficit schizophrenia. For a substantial number of patients (approximately 100), follow-up brain scans are available, which allows data on brain volumes to be used as an alternative endophenotype in association studies.

The first schizophrenia susceptibility genes now seem to have been identified. Support for involvement of genes like neuregulin 1, G72/G30 and dysbindin is becoming more and more robust, but the evidence is still entirely based on statistical association with genetic markers that probably do not play a causal role in the disease. Each reported marker or haplotype was only associated in a minority of patients, and it is likely that other genes, and perhaps different biological systems, are involved in other patients. Until we have identified more of the remaining genetic factors, testing for the presence of the associated haplotypes has no diagnostic or prognostic value. Moreover, until we know how genetic variations contribute to disease by changing the amount or structure of relevant proteins, it will be difficult to develop targeted therapies. These facts may be disappointing to both patients and clinicians, but should be emphasized when the recent genetic findings become public knowledge.

Still, even without knowing functional variants, the relevance of these findings cannot be overestimated. We may have gained the first insight into the biological basis of schizophrenia, and its relationship with other disorders. In the near future, we will learn which biological systems are involved, and in which groups of patients. Our basic concepts of schizophrenia are almost certainly about to change.

11.2. Genetic studies in ADHD

The estimated contribution of genetic factors to the development of attention deficit hyperactivity disorder (ADHD) is at least as high as for schizophrenia. ADHD is assumed to be a genetically complex disorder, but despite considerable research efforts, the responsible genetic variations remain largely unknown. The genetic studies described in this thesis were performed in a sample of families consisting of one or more affected individuals and their parents. This design allowed both genetic linkage and association studies, in which the parents served as controls. It has been suggested that ADHD and the so-called autism spectrum disorders share some causal factors. Therefore, we included patients in a small phenotype group, consisting of full ADHD according to DSM-IV criteria, while the broader phenotype group also included family members with autism spectrum disorders and oppositional defiant disorder (ODD). In an attempt to reduce genetic heterogeneity, patients were only included if they had four Dutch-born, Caucasian grandparents.

Functional candidate genes

Until very recently, genetic studies in ADHD had been restricted to the analysis of functional candidate genes. As in schizophrenia, genes involved in neurotransmission are plausible candidates, since the most effective medication to treat the disorder, methylphenidate, is known to block neurotransmitter transporter molecules such as the dopamine transporter (DAT1). Dopaminergic genes have been reported to be associated with ADHD, but the number of negative findings is also considerable. One of the reasons for the inconsistent findings in separate studies may have been insufficient statistical power. The reported relative risk (RR) of each gene is typically modest, less than two, and the power to detect loci with such RRs in the described samples is limited. Another factor may have been the small number of markers used to study the different genes. Following the publication of the association of the VNTR polymorphisms in DAT1 and DRD4, and the microsatellite in DRD5, subsequent studies focused on these markers. Unless these variants are truly causal, detected association must be the result of linkage disequilibrium with a disease locus. If so, the reported markers, especially the VNTRs, which are not very polymorphic for these loci, may not have been optimal. **Chapter 5** describes an association study of the DAT1, DRD4, and DRD5 genes, in which the previously described VNTRs, as well as additional microsatellites, were analyzed in one of the largest samples reported so far. We found no indications for involvement of any of these three genes in the Dutch population, but as mentioned above, small genetic effects may have been missed for various reasons.

Recent meta-analyses of available studies suggest that the effects exerted by the dopamine transporter genes, as well as the dopamine receptor 4 and 5 genes are small, but real. It remains to be determined if these genes are involved across populations, and if specific symptoms can be related to them.

Positional candidate genes

Together, dopaminergic genes only explain a fraction of the total genetic risk of developing ADHD. Other, unknown genes or gene systems are therefore likely to contribute to the disorder, and linkage studies could provide new clues to other candidate regions in the genome. **Chapter 6** describes one of the first whole-genome scans in ADHD, performed in Dutch families. Affected siblings shared several chromosomal regions more frequently than expected, the most promising of which were on chromosome regions 7 and 15. According to accepted standards, the regions with the highest LOD scores were suggestive for linkage. This indicates that they are likely to contain susceptibility genes, but false-positive findings cannot be ruled out.

The locus on chromosome 7q seemed to give the strongest signal when the analysis was restricted to ADHD only. It had a small estimated effect, increasing the risk for carriers with a factor of approximately 1.2. Such a locus would require larger samples to be fine-mapped in association studies, and given the lack of independent support for this locus, fine mapping efforts were considered premature. However, the region contains dopa decarboxylase (DDC), an a priori candidate gene, since it is involved in the synthesis of dopamine. The gene is located very close to the marker that gave the highest single point LOD score in the genome scan, and modest association with this marker in ADHD has been reported [226]. Therefore, in **chapter 7** we specifically tested DDC for association using multiple markers across the gene, but had negative results. If DDC was the cause of the linkage signal in this region, we should have been able to detect association with it, unless there are many different associated alleles or haplotypes. It is therefore possible that the causal gene is not DDC, but a different gene in the region. Before further association studies are undertaken, we should perhaps await independent confirmation of the chromosome 7 locus.

In contrast with the findings on chromosome 7, several findings now seem to support a chromosome 15 locus that contributes to childhood developmental disorders. Recently, convincing LOD scores in the same region were reported in an American sample [288]. Interestingly, this region coincides with a linkage region in reading disability [221], which is frequently diagnosed in patients with ADHD. There were no remarkable functional candidate genes in the linkage region, but with an estimated relative risk of 1.6, fine-mapping studies in the same sample would have acceptable power. In a preliminary study, we have screened a four megabase (Mb) region around the top of the linkage peak in our study. This study involved over thirty microsatellites, analyzed in pooled DNA. Two markers showed significant association, which was confirmed by individual genotyping (data not shown in this thesis). However, both markers are more than one Mb apart, and it is possible that one – or both – associations are chance results. These experiments are now being replicated in an independent sample at the UMC Nijmegen.

Relevance and future studies

The results of the Dutch genome scan were published together with those of an American group. This study reported linkage on chromosome regions 5p13, 6q12, 16p13, and 17p11 [189, 289]. Linkage to 17p11 has meanwhile been replicated in a study of extended families from an isolate, which also showed linkage to yet other regions, namely 4q13.2, 5q33.3, 8q11.23 and 11q22 [290]. The differences between the available linkage studies seem to be a first indication that ADHD is comparable to other complex disorders, with genetic findings that differ considerably between studies. It is still possible, however, that only a few loci predominate in ADHD. Complex disorders may differ in that aspect, and one may speculate that the relatively strong linkage signals in moderately sized samples reflect underlying genes with a substantial effect. The few available scans in ADHD already seem to show shared linkage regions between studies. Recent progress in other disorders, such as schizophrenia or bipolar disorder, suggests that combining the findings from different studies could be essential to identifying regions that truly contain susceptibility genes. It may be worth taking a chance and fine-map the replicated regions with the highest LOD scores (15q, 16p and 17p).

Combined analysis of the Dutch and the American linkage data highlighted regions that did not seem particularly important in the individual studies, in particular a region on chromosome 5 that did not reach impressive LOD scores in the separate studies (Ogdie *et al.*, unpublished data). Large combined samples may increase the sensitivity of analyses, and allow the specific study of disease subtypes or other phenotypic characteristics. As suggested by the different results in the American and Dutch samples, however, combining samples with different ethnic backgrounds may also reduce homogeneity.

It is intriguing that the most convincing linkage regions in the available studies coincided with regions of previously reported linkage with autism and reading disability. Although coincidence cannot be excluded, it is tempting to think that these regions contain shared susceptibility genes for these childhood developmental disorders that tend to aggregate in families. Genetic studies in samples that were selected for specific phenotypic characteristics shared by the different disorders may provide further insight into common causal mechanisms, and facilitate the identification of new genes. One could speculate that genes involved in neurodevelopment may play a primary role, and that the involvement of neurotransmitter genes is secondary, or modifying in nature.

In summary, after years of studying a limited set of a priori functional candidate genes, linkage studies have now opened new ways to identify susceptibility genes for ADHD. Recent progress in other complex disorders suggests that it is possible to track susceptibility genes using linkage and association strategies. The combined research efforts in ADHD and related disorders are therefore likely to result in the identification of susceptibility genes for childhood developmental disorders in the near future.

11.3. New methods for efficient genotyping

The search for susceptibility genes in complex disorders requires large numbers of affected individuals and genetic markers. In particular, genetic association studies, which aim to identify small genomic regions that trace back to a common ancestor, require high-density marker sets. Systematic screening of linkage regions, or even the entire human genome using association approaches would therefore be a formidable undertaking, which is barely feasible at the moment even though genotyping techniques are advancing rapidly.

Accurate determination of microsatellite allele frequencies in DNA pools

DNA pooling can increase the efficiency of genetic association studies by combining DNA of patients and controls before the analysis. Unfortunately, the PCR-induced stutter artefact frequently hinders the analysis of microsatellite markers in pooled DNA. In **chapter 8**, we have investigated if there were general rules underlying the intensity of the stutter artefact, which might then be used for stutter correction in DNA pool patterns. Stutter intensity was far more pronounced in dinucleotide than in tetranucleotide repeat markers, and correlated with both the type and the number of repeated sequences. The presence of additional repeat sequences of a different sequence increased stutter intensity, whereas repeat interruptions tended to decrease stutter intensity. A decreased intensity of the stutter artefact in genotypes was indicative for the presence of 'aberrant' alleles. In order to gain more insight into the mutation dynamics of microsatellites, it would be interesting to systematically study relative stutter height in alleles of different sizes. This may be possible using the extensive databases of genotyping centres, with data from markers that have been genotyped in different populations. Relative stutter height in different alleles could be an indicator of the mutation rate in a certain repeat tract, and suggest the presence of repeats with sequence variations. Increased knowledge of microsatellite dynamics would be valuable for the selection of microsatellites as genetic markers, and for their use in such diverse fields as the study of population history and microsatellite instability in cancer.

From a stutter correction point of view, however, our findings indicated that it would be dangerous to derive general stutter correction models from public DNA databases, since an unknown mutation of the repeat tract, with resulting lower stutter, may be present in a specific population. An alternative approach is to determine stutter height empirically for each marker by genotyping a number of individuals, and use this information for marker-specific stutter correction. Previously reported stutter correction methods required either homozygous genotypes, or heterozygous patterns, in which the signals of both alleles did not overlap. However, it may take considerable extra genotyping to find a number of useful individual genotypes. This may have been one of the reasons why such methods have not been used on a large scale, even though they were shown to result in accurate allele frequency estimates from DNA pools. **Chapter 9** describes a new stutter correction method, based on a mathematical model-fitting procedure. The principal advantage of this approach is that any individual genotype pattern can be used, even overlapping heterozygous patterns. One standard set of ten individual DNA samples was shown to provide sufficient information for the accurate correction of the stutter artefact in most markers tested. In two markers, the derived stutter correction models were shown to be inaccurate, due to aberrant stutter behaviour of single alleles, or a differential amplification of alleles. These errors, however, could be readily detected, which gives the opportunity to either replace the marker, or to compare the pool patterns without stutter correction.

Microsatellites offer specific advantages over SNPs when haplotype information is lacking, as is the case in DNA pooling studies (see below). Our method for stutter correction allows the accurate determination of microsatellite allele frequencies from DNA pools at the cost of only a limited amount of extra individual genotyping. DNA pooling using microsatellites and stutter correction could be especially efficient in the initial screening phases of large genomic regions [291]. Regions with indications for allelic association could then be followed up by high-resolution individual genotyping.

It should be noted that the use of pooled DNA is likely to result in a loss of information due to factors like measurement errors and the inability to generate haplotypes [98]. Moreover, reanalysis of DNA pooling data in interesting patient subgroups is impossible, without making and genotyping new pools. DNA pooling should therefore be regarded as a second-best but efficient screening technique, as long as individual genotyping at high marker densities is not feasible for the majority of research groups.

The role of microsatellites in genetic association studies

For association studies, the use of microsatellites may offer specific advantages, if regions are to be screened at medium density. The available data suggest that LD between microsatellites is more extensive than between pairs of SNPs. At medium marker densities, microsatellites may be more efficient than single SNPs, particularly if the disease-related variant is also relatively young. Microsatellites have a higher Polymorphism Information Content (PIC value) than SNPs, which only reach comparable PIC levels when several are used in a phase-known combination to generate haplotypes. This is a situation that rarely prevails in the majority of case control studies. On average a haplotype consisting of 3 phase-known SNPs is equivalent in mapping power to a single microsatellite marker. The number of potentially polymorphic microsatellites in the genome is much higher than the number of characterized markers in public databases [292]. The study in chapter 2 demonstrated the feasibility of finding microsatellites in or near candidate genes. Comparable results were reported by others [293]. Moreover, in a schizophrenia candidate region, we found nearly 250 potentially polymorphic di-, tri- and tetranucleotides with an average spacing of 55 kb, and a maximum 'gap' of 168 kb (data not shown). Designing primers is relatively easy with existing software like the Primer3 program [152], and we found the large majority of repeats to be polymorphic, if certain rules like minimum length of the uninterrupted repeat and repeat motif are taken into account. Specific situations in which microsatellites may remain superior to SNPs, are the genotyping of pooled DNA, as discussed above, and in case-control studies where parental information is invariably lacking to be able to generate SNP haplotypes. The possibilities of microsatellite markers therefore do not seem to be exhausted. With decreasing costs for SNP genotyping, however, SNPs may ultimately become the markers of choice in both study designs.

SiNaPse: efficient SNP typing on standard DNA sequencing equipment

In high-resolution genetic association studies, SNPs have rapidly replaced other types of markers, which are less abundant in the genome. A multitude of different SNP typing techniques have been described, ranging from inexpensive, but very labour intensive genotyping with restriction enzymes, to ultra-high throughput, fully automated SNP typing facilities [16, 17, 294]. The latter type of systems brings into reach whole-genome association studies involving hundreds of thousands of markers, although the costs for completing such studies are still enormous. Therefore, until genotyping capacity further increases and the costs per genotype drop, only large specialized centres are likely to be involved in whole-genome association studies. Efficient SNP typing techniques will be essential for other, moderately sized laboratories, in order to perform high-density SNP analyses in specific DNA regions, or candidate genes, and to replicate findings by others. Most genotyping methods on widely available equipment are either labour-intensive, or expensive per genotype. **Chapter 10** describes the development of an efficient SNP-typing technique on standard DNA sequencers, which are present in most laboratories. The method is based on a PCR with two allele-specific primers with a 3' locked nucleic acid (LNA). This modified base binds to the SNP with a very high affinity. As a result, the allele-specific PCR can usually be performed using one standard PCR protocol, without the need for optimizing conditions for each SNP. The two allele-specific products can be separated by size, because one of the primers is extended with short spacer sequence. The position of the reverse primer can vary the total size of the two products for each SNP, and multiple SNPs can thus be analyzed in a single run. A short universal sequence with a labelled dye is incorporated into the products during the PCR, which circumvents the need for marker-specific labelled primers, and reduces costs considerably. The method was tested by comparing genotyping results with those obtained on a TaqMan system and direct sequencing. In the majority of cases, it was possible to amplify the SNP using standard conditions but for some markers, amplification of one of the alleles seemed to have failed. This is not likely to be a major problem, since frequently a new set of primers can be defined on the opposite DNA strand, and otherwise the failure is easily recognized. Results from TaqMan genotyping and direct sequencing were accurately reproduced using the SiNaPse technique.

This new SNP genotyping method is especially suitable for flexible and relatively inexpensive SNP genotyping in labs with operational DNA sequencers. As such, it is a useful addition to the currently available high-throughput platforms, which are more cost-effective when very large numbers of genotypes have to be generated.

11.4. Psychiatric genetics – the road ahead

Recent findings in the genetics of psychiatric disorders are very encouraging, since they provide insight into the genes and gene systems that contribute to these diseases. It should be noted, however, that all the evidence so far is based upon statistical association of marker variations within or near candidate genes. These variations will probably have no functional relevance and simply signpost the presence of a true disease-causing DNA variation in the immediate vicinity.

Finding the functional variant(s)

Finding the causal variants in the putative susceptibility genes should now be given a high priority. Such studies are likely to require high-density SNP sets, or even direct sequencing of genes and regulatory regions [295]. The disappointing results in neuregulin 1 and other candidate genes suggest that finding the 'culprit variant' among a multitude of associated SNPs will be difficult. Final proof of involvement of certain variants will require functional experiments, in which the function of genes, *in vitro* or *in vivo*, is shown to be altered by the variant. This may even require the use of 'knock-in' mice or other animal models in situations where the behaviour of the animal can be significantly modified by the introduced genetic lesion. It may seem to be a long reach to study the effects of specific point mutations of potential importance in human psychiatric disorders in mice. The majority of the background genetic variation, however, which largely masks the specific effect of the gene variation concerned on the human disease phenotype, is largely eliminated in inbred strains of mice. As a consequence, specific phenotypic effects are more easily identified.

Sample characteristics- quantity or quality?

The putative susceptibility genes in complex disorders only marginally increase the risk, typically with relative risks of less than two, and together they explain a small proportion of the total genetic genetic susceptibility. As has been suggested, these loci may represent the 'low-hanging fruit', i.e. the common variants, which are most easily detected, since they contribute to disease in a substantial proportion of patients. If the effect of single variants were small, collecting very large samples would be an option for increasing the statistical power of studies. Collecting such large samples may require collaboration, since this is difficult for single research groups to do. If the interests of single groups can be safeguarded and phenotyping procedures are carefully compared, large multi-centre studies could provide further progress. Combining samples from different origins, however, also carries a risk of increasing genetic heterogeneity.

Genetic studies in complex disorders indicate the existence of extensive allelic heterogeneity. Even between European populations, associated alleles and haplotypes in schizophrenia seem to be different, as suggested by our own results for NRG1, and data from others on dysbindin [135]. Combined analysis of our ADHD linkage data and a US sample suggested that signals within each sample reflected true linkage and that the lack of replication between samples is best accounted for by allele frequency variability at several putative 'risk' genes in ADHD (Ogdie *et al.*, unpublished data). We collected samples that were selected for a Dutch, Caucasian background, and we were able to obtain relatively strong linkage and association signals in moderately sized samples. At this stage, it is difficult to tell how much could be gained by using strict ethnic criteria. Now that large numbers of SNPs are being typed in different populations, the effects of genetic heterogeneity can perhaps be evaluated more thoroughly in the near future. Until then, our data suggest that ethnic stratification could be worth the effort. Using family history is one way to trace ethnic origin, but different subpopulations could also be defined at the DNA level, by typing a number of random genetic markers [30, 296].

We found indications that deficit and non-deficit schizophrenia may be a useful clinical subdivision for genetic studies. Likewise, selection for other clinical characteristics, or more basal endophenotypes such as volumes or activation patterns on brain scans, may be useful [113]. Despite the attractive concept of the study of endophenotypes, however, studies that unequivocally show the benefit of this approach are still sparse. Careful phenotyping of patients may be very labour-intensive, and if this impedes the collection of samples of sufficient size, studies may lack power.

It seems too early to tell whether studying large, unselected samples or smaller, narrowly defined samples will be preferable. From a practical point of view, research groups should choose the strategy that is most convenient, e.g. depending on the facilities for extensive phenotyping, the presence of large diagnosed patient groups, from which DNA could be easily obtained, or the availability of unique large families or isolated populations.

In either approach, large, carefully phenotyped patient samples will be essential for performing the next generation of genetic studies. With the rapid progress in genotyping techniques, sample collection may become the largest bottleneck in genetic research, and the quality of patient samples is likely to determine the chances of success of future genetic studies [297]. Ongoing collection of these samples should have a high priority, and clinicians and patients should be motivated to participate in such studies.

Besides large samples of unrelated patients or multiple small families, one large family with multiple affected individuals may provide valuable clues to susceptibility genes, as evidenced by the translocation between chromosomes 1 and 11 that pointed to the involvement of the DISC1 and DISC2 genes in psychiatric disorders. Moreover, whole-genome linkage studies may be possible in large families. Likewise, studying genetically isolated populations that stem from a limited number of founders could facilitate the detection of disease-related variants. A variant that is inherited with disease may be unique to a specific family or genetic isolate, but it may point to other variants in the same gene, or to the pathways involved.

Common disease, common variant?

Despite early scepticism, the most successful approach to finding genes in complex disorders so far seems to have been the 'traditional' combination of whole-genome screens based on linkage, followed by fine-mapping efforts using association studies. Positional methods without a priori assumptions about gene function are likely to remain important. It is inherent in the techniques used, that the recently found disease-related variants are common. The characteristics of additional variants will profoundly influence the chances of success of identifying them using the same approaches. In a worst-case scenario, there is extensive genetic heterogeneity, with tens or even hundreds of genes contributing to a single disease, as well as allelic heterogeneity, with many different, rare, disease-related variants in each gene. Single variants would then contribute very little to all disease in a population, and they would therefore be almost impossible to detect by either linkage or association methods.

If only a limited number of genes is involved, but with multiple different disease-related variants in each gene, linkage studies would have good chances of detecting these genes. Linkage studies can identify chromosomal regions that are shared more often than expected by affected family members. As long as members of single families share them, the specific variants in a gene are not relevant, which makes linkage studies insensitive to allelic heterogeneity. Genetic association studies, on the other hand, compare the frequency of specific variants in patients and controls. Even if a limited number of genes were involved, the presence of extensive allelic heterogeneity in each gene could therefore severely compromise the power to detect association between the disease and each allele individually [271]. In this situation, one would expect it to be very difficult to fine map convincing linkage regions by means of association studies [7].

In the most optimistic view, the susceptibility to common disorders is largely determined by a limited number of common variants, as predicted by the common disease-common variant (CDCV) hypothesis [4]. Until we find evidence to the contrary, there is no reason to abandon the successful linkage and association approaches, as long as we keep in mind that the presence of common disease alleles is a critical assumption.

Are all psychiatric disorders determined by similar variants? It is interesting to speculate if psychiatric disorders could differ with respect to their allelic structure, in addition to the genes concerned. For example, schizophrenia patients have relatively few children on average. This implies that specific combinations of schizophrenia-causing alleles are under selection pressure and being lost. One would expect this to lead to a spectrum of alleles with relatively low frequencies and short half-lives. In other disorders, the reproductive disadvantage is not evident, which might in time result in older and relatively common alleles. In the long term, the study of allelic combinations at multiple loci, or interactions, may be more powerful than gene-by-gene analyses for both schizophrenia and other complex disorders. Although this approach is theoretically appealing, it would require very large sample sizes and the statistical tools required are not yet well developed.

A new life for linkage

In several complex disorders, including schizophrenia, linkage studies have provided candidate regions in which susceptibility genes have later been pinpointed by association approaches. It is likely that linkage studies will continue to provide important clues for locating susceptibility genes in other psychiatric disorders, as shown by the recent genome scans in ADHD, including ours. Although psychiatric disorders may differ in their genetic makeup, it is suggested that many independent studies will be needed, and that one should not be discouraged by apparent non-replications and relatively modest evidence in single studies [209].

As explained above, linkage approaches may remain irreplaceable to find loci with extensive molecular heterogeneity. In a disorder like schizophrenia, the most convincing linkage regions now seem to have been fine-mapped, and simply performing more linkage studies in similar samples may not be very fruitful, unless much larger samples and marker densities can be studied. However, the majority of SNP combinations in such regions have not yet been exhaustively studied and as the SNP map continues to develop, each previously identified linkage region should be re-examined in the light of the new information.

New loci may also be detected by studying samples with specific disease subtypes or endophenotypes. Collecting sib pair samples diagnosed with a specific subtype of a disorder, however, may be an almost impossible undertaking. It may be easier to leave the clinical disease classification behind, and focus on traits that are present in patients, but also in the general population. Examples of such endophenotypes are specific cognitive deficits, schizotypical personality, or more basal physiological measures like sensory gating deficits and eye tracking patterns [113]. Whichever clinical and biological traits are chosen, their use in genetic studies will require robust data on their heritability.

The recent advances in SNP genotyping may provide an efficient and inexpensive alternative for linkage studies based on microsatellite markers. Whole-genome scans using several thousands of SNPs instead of microsatellites have been reported to provide at least comparable accuracy, at a fraction of the costs [298]. There are indications that dense SNP maps can provide an information content that is superior to that of a standard 400 microsatellite map [299, 300].

Taken together, although the concept of whole-genome association studies appears to be more fashionable at the moment, the possibilities of linkage studies should not be overlooked [32].

The promise of whole-genome association studies

Association studies have been claimed to provide more power than linkage studies with comparable sample sizes [271]. Moreover, single patients are much easier to collect than multiply affected families. Whole-genome association studies may therefore be an attractive alternative for linkage studies, and they could provide new loci, which have been missed by linkage approaches so far. Due to the limited extent of LD in unrelated individuals, the number of markers needed for a comprehensive whole-genome association study is several orders of magnitude higher than the number needed for a linkage screen. Fortunately, rapid advances in genotyping technology are bringing such studies within reach. It remains to be seen if whole-genome association studies will resolve all the problems in finding complex disease genes, however. Testing hundreds of thousands of markers introduces a major problem of multiple testing, even though many markers in the same region will not always be independent. Numerous positive findings by chance alone can be expected, unless one applies very stringent significance levels. New methods of statistical analysis are currently being developed, which will hopefully reduce the problem of multiple testing. Even then, samples that are much larger than most currently available samples will probably be required for the detection of variants with modest effects. It is questionable whether the recently reported detection of candidate genes from linkage regions would have been picked up in the described samples, if the entire genome instead of just a specific linkage region had been screened, since the stringency required to remove potential false-positives is so much higher in a total genome approach.

Further, as outlined above, association studies are vulnerable to allelic heterogeneity. If substantial heterogeneity is present, even the most elaborate association studies are doomed to fail using current analytical approaches. Finally, however dense the markers sets, association studies based on LD remain an indirect method. There will always be a chance of disease-related variants not being in LD with a marker, and characteristics of the markers used will determine what kind of variants will most likely be detected. Association between a disease-related variant and a marker is most easily detected if their frequencies are comparable. The SNPs that are currently being collected in efforts like the HapMap project [23] have been selected for high minor allele frequencies (>1%). Moreover, for reasons of statistical power, association studies will generally use SNPs with even higher frequencies, e.g. >10%. As a consequence, such studies will favour the detection of common disease alleles, while rare variants could be missed. Haplotyping may reduce this bias, since haplotypes with low frequencies can be constructed from common SNP alleles, which in a way can then be regarded as new markers. Several authors have recently pointed out that multi-allelic markers and haplotypes may have a greater power to detect association, most notably if more than one disease-related variant is present [301, 302].

The optimal marker density for whole-genome association studies has been the subject of much debate. Recent data on LD patterns throughout the genome suggest that LD is highly variable, and that marker sets should preferably be based on empirical LD data. These data are becoming available through the HapMap project, and commercial initiatives. Even though LD patterns are being determined in samples with different ethnic backgrounds, it remains to be determined how well these data describe LD in specific samples. The same holds for the LD blocks that are being derived from these data, with the aim of eliminating the typing of fully redundant SNPs. Recent data suggest that a block-like pattern is probably less evident than it first appeared. The density of genotyped markers influences the observed pattern of haplotype blocks, and although some block boundaries seem to be present in all populations, others may be more population-specific [303]. Moreover, a substantial part of the genome seems to lack any block-like structure at all [304]. More data therefore seems to be necessary, in order to determine how well these LD maps can be translated to specific study populations, and to determine if haplotype-tagging SNPs (htSNPs) from such databases truly capture all haplotype diversity.

Despite these theoretical limitations, a study involving hundreds of thousands of markers and sufficiently large samples would probably provide an unprecedented number of new candidate variants. Such studies may soon be practically feasible. The question is whether studies of this scale are the only way to proceed. Instead of screening the entire DNA, one could start with a gene-centred study, thus reducing the amount of DNA to be screened by several orders of magnitude. Alternatively, we could investigate only those regions in which some evidence for linkage has been reported. The screening of large genomic regions, starting with sparser marker sets, either in individual genotyping or in pooled DNA, can certainly be defended [305]. These restrictions will generally increase the chances of missing an association, or reduce the power of studies. Therefore, negative findings should never be taken as evidence of exclusion of involvement. It makes sense, however, to start by looking for the 'low-hanging fruit', i.e. common variants with a substantial effect. These variants may have the greatest impact on disease at the population level, and may be the most interesting for developing targeted treatment. The reported systematic screening of linkage regions in schizophrenia and other complex disorders illustrates the possibility to trace disease-associated genes using marker sets with a modest density.

In summary, many questions about the optimal study design and methods for data analysis remain to be answered, but association studies of the entire genome or of genomic regions are likely to give a new impulse to research on the genetics of complex disorders.

Candidate genes

Part of the elegance of linkage and whole-genome association approaches is that these methods make no assumptions about the function of genes that could be involved in a disorder. A multitude of studies of functional candidate genes suggests that the genes that are intuitively the most plausible candidates are not very prominent in causing disease. There may be more objective, biological indications for the involvement of certain genes, however, which it may be worthwhile to follow.

Following the new leads. The recently detected candidate genes will provide valuable starting points for the identification of other disease-related variants. Some genes may be known to belong to specific biological pathways, such as the glutamate system in schizophrenia. Knowledge about the architecture of these systems may already be, or become available from studies in other related mammalian species or more distant organism like *Drosophila* or *C. Elegans*. If the function of new genes is unknown, protein interaction studies with the product of a known susceptibility gene may reveal additional candidate genes. In the example of the G72/G30 gene, this approach has resulted in the identification of DAAO, which was subsequently shown to be associated with schizophrenia as well [68, 139]. Alternatively, modifying identified susceptibility genes in animals and investigating the effect on the expression of other genes may reveal new pathways. Breeding knock-in animals to modify expression of a specific candidate gene is time-consuming, but the new technique of RNA interference (RNAi) may allow rapid manipulation of gene expression in animals without the need to breed knock-in animals [306-308]. In general, disciplines related to genetics will become essential to guide the search for disease-associated variants. Huge amounts of information on gene function, expression, and interaction are rapidly becoming available in public databases.

Bioinformatics will become increasingly important for organizing, searching and using all this information efficiently.

Gene expression studies have repeatedly highlighted genes and gene systems that are up- or downregulated in the brains of schizophrenic patients, such as genes involved in synapse formation and myelination. RGS4 was found to be strongly downregulated in expression studies [69], and subsequent genetic association studies, including ours, now seem to confirm the involvement of the gene in schizophrenia. Micro-arrays allow the simultaneous analysis of the expression of thousands of genes, and the results of such studies may be useful for selecting candidate genes for genetic studies. It is doubtful, however, if genes with a different expression will always be causal. In diseases like schizophrenia, patients will nearly always have received antipsychotic medication before their death, while controls most likely have not. The effects of medication, or different living conditions prior to death may therefore cause different gene expression between patients and controls. A problem that is more or less specific to psychiatric disorders is the fact that brain tissue is relatively inaccessible in living patients, and sample collection may therefore be difficult. Expression patterns in peripheral blood lymphocytes may be an attractive substitute, however [309].

Cytogenetic abnormalities may provide yet other clues for the involvement of certain genes. The DISC1 gene, for Disrupted In SChizophrenia, was found to be disrupted by a chromosomal translocation, which is inherited together with psychiatric disorders in a large Scottish family. The gene has recently been found to be associated with schizophrenia in large samples of independent patients, which indicates that chromosomal abnormalities may provide valuable candidate genes for psychiatric disorders. However, chromosomal abnormalities in one family will probably not always play a significant role at the population level. In contrast with disorders like autism, cytogenetic studies in schizophrenia and many other psychiatric disorders have received relatively little attention. CGH microarray-based techniques now allow high-resolution screening for minor cytogenetic copy number abnormalities, and systematic screening of selected patients should therefore be considered.

Studying gene and allelic interactions

So far, genetic variants and haplotypes have mostly been investigated one-by-one. Disease may also be the result of a particular combination of susceptibility variants in individual patients, and it is also possible that certain variants are only relevant in the presence of others. In the case of multiple interactions between genes, single variants are likely to be only detectable by simultaneous analysis of multiple loci. Statistics for studying gene interactions are being developed, and hopefully, these will be able to discern genes belonging to interacting systems, that may have been overlooked so far.

Replication required

The first results in genetic studies of complex disorders suggest that replication is essential to evaluate the relevance of genetic findings. Such studies may not be the most satisfying to perform, but they should be encouraged, and rewarded. Negative findings should have equal chances of being published as positive findings, provided that the quality of the studies is guaranteed. Raw data should be made available for future meta-analyses, perhaps through coordinated web sites, and this should be a pre-requisite for publication. In addition, patients' clinical data and DNA samples should also be made available. The NIH have now made this data- and sample availability to other researchers in the field a binding condition for receiving extensive funding.

Conclusion

The classical approach of linkage and association studies in unselected samples has resulted in the identification of the first susceptibility genes in complex disorders. In psychiatric disorders, as well as in other complex disorders, the relative risks of the identified variants are low, typically less than two, which suggests that finding additional genetic risk factors in unselected patient samples will be a difficult task. Rapid technical developments have brought whole-genome association studies and gene expression studies within reach, and such studies are likely to provide wealths of new information. Handling the vast amounts of data and limiting false-positive and false-negative results will be a major challenge for the near future. At present, there is too little data on the nature of disease-related variants to determine the optimal strategy for finding additional variants. Studies in large, unselected samples will probably result in the detection of different causal factors than studies in samples with specific characteristics or symptoms, and in order to find as many contributing variants as possible, both approaches may be needed. Individual research groups should carefully determine the optimal study design in their specific situation, e.g. depending upon the availability of specific patient samples, and facilities.

Whichever road will be chosen, intelligent use of biological and clinical information from disciplines outside genetics may provide valuable shortcuts. Together, genetics and related disciplines are likely to further increase our insight into the biological basis of psychiatric illness.

11.5. From susceptibility genes to susceptible persons

After many years of research, recent findings promise to provide a glimpse into the biological basis of psychiatric disease. Long-standing questions may soon be answered, for example: are there shared causes for disorders such as ADHD and autism, or schizophrenia and bipolar disorder? Is schizophrenia a developmental or a neurodegenerative disorder, or perhaps a combination of both? Can disease subtypes with a distinct aetiology be distinguished within the currently recognized disease entities? As a direct result, our concepts and classification of these disorders are likely to change in the near future. Moreover, a firmly established biological basis of psychiatric diseases may help patients and their families to accept the illness as not fundamentally different from other chronic disorders with a heritable susceptibility, such as diabetes or asthma.

The most important question, however, is: how will increased knowledge of the causes of psychiatric diseases benefit those who suffer from them? It is not likely that we will ever be able to tell with certainty which individuals will later develop a common psychiatric disorder, since disease is the result of a unique genetic constitution and environmental factors. This uncertainty may be comforting to many, because it prevents genetic determinism, but it may also be disappointing to those who hope to banish psychiatric illness completely from society. Still, if we assume a threshold model for the development of disease, it would be valuable only if we could identify a limited number of risk factors, both genetic and environmental, in order to keep individuals at high genetic risk below the environmental threshold for the development of symptoms. Once we know specific genetic risk factors, the identification of environmental risk factors may be facilitated, by studying these in groups with a similar genetic background. In addition, it may be possible to develop more specific medication, able to target the appropriate biochemical pathways and with fewer side effects than the current medication.

In conclusion, genetics and related fields are likely to provide insight into the causes of psychiatric disorders in the near future, which in turn will hopefully benefit the prevention and treatment of these debilitating diseases.

12 NEDERLANDSE SAMENVATTING

Dit proefschrift beschrijft onderzoek naar genen die betrokken zijn bij schizofrenie en attention-deficit hyperactivity disorder (ADHD), en de ontwikkeling van efficiënte methoden voor het uitvoeren van genetische studies. Erfelijke factoren spelen een belangrijke rol bij de ontwikkeling van veel voorkomende ziekten, waaronder psychiatrische stoornissen, zoals tweelingstudies en adoptiestudies hebben uitgewezen. Het patroon van overerving binnen families wijst erop dat in het algemeen meerdere erfelijke factoren en omgevingsfactoren betrokken zijn bij het ontstaan van psychiatrische ziekten. Bij dergelijke genetisch complexe aandoeningen behoeft de bijdrage van afzonderlijke genen niet groot zijn, wat hun detectie bemoeilijkt.

Principes van genetisch onderzoek. Het erfelijk materiaal bestaat uit DNA ketens, opgebouwd uit vier verschillende bouwstenen, of basen, die de genetische code vormen. Twee complete versies van het DNA zijn aanwezig in de cellen van het lichaam: een afkomstig van de vader en een van de moeder. Genen zijn vast omschreven gebieden in het DNA, met informatie voor de vorming van eiwitten, de moleculen die bijna alle belangrijke functies in het lichaam vervullen. Varianten in genen kunnen leiden tot een veranderde eiwitfunctie, en uiteindelijk tot ziekteverschijnselen.

Ondanks de snel toegenomen capaciteit van analyseapparatuur is het nog niet mogelijk om het totale DNA te vergelijken tussen grote groepen patiënten en gezonde controles. Voor genetische studies zal daarom een voorselectie gemaakt moeten worden van de te bestuderen genen, en de in die kandidaatgenen te analyseren varianten.

De onderliggende aanname bij genetische studies is dat een ziekteveroorzakende variant, of mutatie, ergens in de geschiedenis ontstaan is, en dat de thans levende zieke nakomelingen deze mutatie, en het omringende DNA, nog steeds dragen. Om een onbekende mutatie op het spoor te komen volstaat het derhalve in eerste instantie, om bij zieke personen te zoeken naar gebieden in het DNA, die wijzen op een gemeenschappelijke zieke voorouder. Bij ieder persoon wisselen de twee versies van het totale DNA willekeurig gebieden van gelijke lengte uit, voordat een helft wordt doorgegeven aan het nageslacht. Deze gedeelde gebieden rondom een mutatie zullen daarom gemiddeld met iedere volgende generatie kleiner worden. Nauw verwante familieleden delen gemiddeld grote DNA gebieden, wat globale lokalisatie van een mutatie mogelijk maakt. Hierop berust het principe van koppelingsonderzoek in families, waarbij gedeelde DNA gebieden worden opgespoord door het totale DNA op slechts een beperkt aantal punten (300 tot 400) met elkaar te vergelijken. Zieke personen met een verre gemeenschappelijke voorouder, daarentegen, zullen nog slechts zeer klein gebieden rondom de gemeenschappelijke mutatie met elkaar delen, wat nauwkeuriger lokalisatie mogelijk maakt. Voor een dergelijke genetische associatiestudie zal het DNA echter op veel meer punten vergeleken moeten worden, om geen gebieden te missen. DNA afkomstig van verschillende personen kan worden onderscheiden met behulp van genetische markers. Dit zijn individuele variaties tussen personen in het DNA, die frequent voorkomen, maar niet noodzakelijkerwijs van functioneel belang zijn. Twee veel gebruikte typen markers zijn single nucleotide polymorphisms (SNPs) en microsatellieten, ook wel short tandem repeat (STR) markers genoemd. SNPs zijn variaties van een base in het DNA op vaste plaatsen, terwijl microsatellieten repeterende eenheden in het DNA zijn (bijvoorbeeld CACACACACA), met een verschillend aantal herhalingen per persoon. Met behulp van beide typen markers kunnen gebieden worden opgespoord die teruggaan op een zieke voorouder met een specifieke combinatie van markervariaties rondom een ziektemutatie. Voordat een variatie zichtbaar gemaakt kan worden, is het in het algemeen vereist om het DNA met de marker vele malen te vermenigvuldigen. Dit gebeurt met behulp van de polymerase chain reaction (PCR).

12.1. Genetische studies bij schizofreniepatiënten

Schizofrenie is een psychiatrische stoornis die wereldwijd bijna 1% van de bevolking treft. De ziekte wordt gekenmerkt door wanen en hallucinaties ('positieve symptomen'), en daarnaast een verslechtering van het algemeen functioneren, met onder meer verlies van initiatief en interesse ('negatieve symptomen').

Op basis van onderzoek in tweelingen en geadopteerde kinderen wordt de bijdrage van erfelijke factoren aan de ontwikkeling van schizofrenie geschat op ongeveer 80 procent. Over de betrokken genen of gensystemen is echter nog zeer weinig bekend. Al geruime tijd geleden is gesuggereerd dat de boodschapperstof dopamine een rol zou kunnen spelen bij de ontwikkeling van schizofrenie, aangezien bekend is dat effectieve medicijnen met name de receptoren voor dopamine in de hersenen beïnvloeden. **Hoofdstuk 2** beschrijft een systematische analyse van 12 genen die betrokken zijn bij de aanmaak, de werking en de afbraak van dopamine. Gebruikmakend van de in hoofdstuk 9 beschreven DNA pooling techniek, werden voor ieder gen een of meerdere microsatelliet markers onderzocht op het meer frequent voorkomen van bepaalde varianten bij patiënten dan bij gezonde controles. Er werden geen significante verschillen gevonden, wat erop duidt dat het dopamine systeem bij Nederlandse schizofreniepatiënten geen belangrijke rol speelt in de ontwikkeling van de ziekte. De grootte van de onderzochte groepen, en de keuze van de markers, sluiten echter niet uit dat subtiele effecten niet detecteerbaar zijn geweest.

Reeds vele studies zijn verricht naar genen die op basis van hun functie betrokken zouden kunnen zijn bij schizofrenie, zoals dopamine genen, en voor een aantal van deze genen hebben meta-analyses van de beschikbare studies een associatie met de ziekte aangetoond. Toch lijken dergelijke functionele kandidaat-genen slechts een fractie te verklaren van de totale aanleg voor het ontwikkelen van schizofrenie. Reeds eerder zijn echter vele koppelingsstudies voor schizofrenie verricht, waarin meerdere chromosoomgebieden zijn gevonden, die vaker dan verwacht gedeeld worden door zieke familieleden. Recent zijn in enkele van deze koppelingsgebieden genen gevonden, waarvan in verschillende bevolkingen specifieke markervarianten geassocieerd zijn met schizofrenie. Niet alle studies waren echter positief, en de effecten van de verschillende genen waren klein. Een mogelijke verklaring hiervoor is dat sommige genen overwegend een rol spelen bij patiënten met specifieke kenmerken, of in bepaalde bevolkingsgroepen. In dat geval zouden deze eerder te detecteren zijn in geselecteerde studiegroepen. Deze mogelijkheid hebben wij onderzocht door twee groepen patiënten te verzamelen, namelijk patiënten met en zonder het deficit syndroom. Deficit schizofrenie wordt gekenmerkt door aanhoudende negatieve symptomen zoals verlies van initiatief en interesse, en een slecht sociaal functioneren. Eerder is geopperd dat de ziekteoorzaken in deze groep zouden kunnen verschillen van die bij andere schizofreniepatiënten.

Het neureguline 1 gen (NRG1) was ook in onze groep geassocieerd met schizofrenie, zoals beschreven in **hoofdstuk 3**, en in dit gen was een andere variant van een eerder bestudeerde SNP geassocieerd. Opmerkelijker was echter, dat vrijwel de gehele associatie kon worden toegeschreven aan de non-deficit groep, met een relatief gunstig ziektebeloop. Op dezelfde wijze zijn in **hoofdstuk 4** de dysbindin (DTNBP1), G72/G30, RGS4 en PIP5K2A genen onderzocht. De eerste drie genen zijn eerder herhaaldelijk in verband gebracht met schizofrenie, terwijl het weinig onderzochte PIP5K2A in een eerder gepubliceerde studie in verband is gebracht met bipolaire stoornis. RGS4 was, vergelijkbaar met neuregulin 1, alleen in de non-deficit groep geassocieerd. PIP5K2A, daarentegen, was sterk geassocieerd met zowel deficit als non-deficit schizofrenie. Dit lijkt erop te wijzen dat dit gen betrokken in bij bepaalde symptomen die gemeenschappelijk zijn aan zowel schizofrenie als bipolaire stoornis. Voor dysbindin en G72/G30 werd geen associatie met schizofrenie in het algemeen gevonden, noch met de deficit of non-deficit vormen van schizofrenie. Mogelijk spelen deze genen in de Nederlandse populatie geen rol van betekenis, of zijn in de Nederlandse bevolking andere markers dan de eerder gerapporteerde geassocieerd.

Onze resultaten suggereren het bestaan van genetisch verschillende vormen van schizofrenie, en van genen die betrokken zijn bij meerdere psychiatrische stoornissen. Voor verder genetisch onderzoek lijkt het waardevol om nauwer omschreven patiëntengroepen te gebruiken, omdat hierin groepsspecifieke genen eenvoudiger gedetecteerd zouden kunnen worden. Ook vanuit klinisch oogpunt kan deze bevinding van belang blijken. Het betrouwbaar onderscheiden van ziektevormen met een verschillende oorzaak maakt het op termijn wellicht mogelijk om betere uitspraken over ziektebeloop te doen, en behandelingen gerichter te kunnen kiezen.

12.2. Genetische studies bij ADHD patiënten

ADHD is een stoornis die zich veelal op kinderleeftijd openbaart, en die wordt gekenmerkt door aandachtsstoornissen en overmatige activiteit. Het geschatte aandeel van erfelijke factoren bij het ontwikkelen van ADHD is minstens even groot als bij schizofrenie. Bij de aanvang van het in dit proefschrift gepresenteerde onderzoek waren nog geen koppelingsstudies voor ADHD verricht en beperkten genetische studies zich tot associatiestudies van functionele kandidaat-genen. Evenals bij schizofrenie is het dopamine systeem een interessant kandidaat-systeem voor ADHD, aangezien bekend is dat de meest effectieve medicatie voor de behandeling van de ziekte, methylfenidaat, de werking blokkeert van het dopamine transportmolecuul, DAT1. **Hoofdstuk 5** beschrijft een associatiestudie van het DAT1 gen en de dopamine receptor D4 (DRD4) en D5 (DRD5) genen. Er werd geen associatie gevonden tussen ADHD en deze functionele kandidaatgenen. Hiermee schaart deze studie zich onder de reeks negatieve studies die voor beide genen zijn gepubliceerd, naast de studies met positieve bevindingen. Recente meta-analyses van verrichte studies geven aan dat beide genen de kans op het krijgen van ADHD in zeer beperkte mate, doch aantoonbaar verhogen. Hoewel het in dit proefschrift beschreven studiecohort niet tot de kleinst beschreven behoort, is de kans op het niet detecteren van dergelijke genen met een laag relatief risico niet te verwaarlozen.

Om andere mogelijk bij ADHD betrokken genen op het spoor te komen, werd een in **hoofdstuk 6** beschreven koppelingsonderzoek van het totale genoom verricht, binnen 106 families met twee of meer kinderen met ADHD. Verschillende chromosoomgebieden werden vaker dan verwacht gedeeld door zieke kinderen. In chromosoom regio's 7p en 15q waren de aanwijzingen voor koppeling met ADHD suggestief, met LOD scores van respectievelijk 3,04 en 3,54. De regio op chromosoom 15 is eerder gevonden in koppelingsstudies naar dyslexie, een aandoening die frequent voorkomt binnen families met ADHD. Dit lijkt erop te duiden dat deze regio een gen bevat dat bijdraagt aan beide stoornissen. Linkage in deze regio is inmiddels door anderen gerepliceerd in een onafhankelijke groep ADHD families. De resultaten van de Nederlandse koppelingsstudies werden gezamenlijk gepubliceerd met die van een soortgelijk Amerikaans onderzoek. In deze laatste studie werden echter de sterkste aanwijzingen gevonden op chromosomen 16 en 17. Deze eerste koppelingsstudies in ADHD lijken het patroon te volgen van koppelingsstudies in andere complexe aandoeningen, zoals schizofrenie, met hun uiteenlopende bevindingen. De ervaringen die zijn opgedaan in deze meer intensief bestudeerde ziektebeelden, tonen aan dat de gecombineerde resultaten van koppelingsonderzoek daadwerkelijk kunnen leiden tot het identificeren van ziektegenen. Onafhankelijke studies zullen moeten uitwijzen welke chromosoomgebieden in verschillende populaties een belangrijke rol spelen, en samenwerking tussen verschillende onderzoeksgroepen is hierbij van groot belang.

Hoofdstuk 7 beschrijft een gedetailleerde associatiestudie van het dopa decarboxylase gen (DDC) binnen de families met ADHD. DDC is een interessant functioneel kandidaatgen voor ADHD omdat het de laatste stap verzorgt in de vorming van dopamine. Reeds eerder is het gen geassocieerd gevonden met ADHD, hoewel deze studie niet herhaald is. De resultaten van het koppelingsonderzoek maken het gen echter ook tot een interessante positionele kandidaat, aangezien het gelokaliseerd is in de koppelingsregio op chromosoom 7 in de in hoofdstuk 6 beschreven koppelingsstudie. Er kon geen associatie van DDC met ADHD worden aangetoond, waarmee betrokkenheid van het gen in de Nederlandse bevolking minder waarschijnlijk wordt.

12.3. Nieuwe methoden voor efficiënte genetische markeranalyse

Het typeren van markers bij grote aantallen patiënten en controles is tijdrovend wanneer dit voor elk persoon afzonderlijk moet gebeuren. DNA pooling is een techniek waarbij gelijke hoeveelheden DNA van grote aantallen individuen worden samengevoegd voor gelijktijdige analyse. Op deze wijze kan op efficiënte wijze bepaald worden of er verschillen zijn in de verdeling van markervarianten tussen groepen patiënten en controles. Vergeleken met SNPs hebben microsatelliet markers vaak vele varianten, in plaats van slechts twee, en dit hoge informatiegehalte maakt microsatellieten in theorie heel geschikt voor analyse in DNA pools, waarin alleen analyse van afzonderlijke markers mogelijk is. De analyse van microsatelliet markers in DNA pools wordt echter bemoeilijkt door het optreden van fouten die optreden tijdens de vermenigvuldiging van het repeterende DNA met de PCR. Hierbij kan nu en dan een repeterende eenheid worden overgeslagen, wat ertoe leidt dat naast de werkelijke lengte van het repeterende DNA ook fragmenten met een voornamelijk kortere lengte worden waargenomen. In DNA pools verstoort dit zogenaamde stutter artefact de correcte bepaling van het aantal personen met een daadwerkelijk kleiner aantal herhaalde eenheden. **Hoofdstuk 8** laat zien dat de intensiteit van het stutter artefact voornamelijk afhankelijk is van het type en het totale aantal herhaalde eenheden in het DNA. Deze kennis zou in principe gebruikt kunnen worden om het artefact te voorspellen en in DNA pools te corrigeren, ware het niet dat tevens werd gevonden, dat onderbrekingen van het herhalingsmotief, die niet altijd bekend hoeven zijn, het optreden van stutter sterk kunnen verminderen.

In **hoofdstuk 9** wordt daarom een nieuwe methode voor stuttercorrectie gepresenteerd, die is gebaseerd op meting van de stutterintensiteit in enkele individuele personen. Hiermee wordt vervolgens een model opgesteld om het signaal afkomstig van DNA pools voor het artefact te corrigeren. In tegenstelling tot eerder gepubliceerde correctie methoden maakt een wiskundige fitprocedure het mogelijk om de benodigde informatie te verkrijgen door meting van stutter in slechts 10 individuele personen. De methode werd getest met meer dan dertig verschillende microsatelliet markers, waarbij de geschatte aantallen markervarianten in DNA pools werden vergeleken met die som van de individuele metingen van alle personen die in de DNA pool vertegenwoordigd waren. In het merendeel van de gevallen bleek de correctiemethode accurate bepaling van de aantallen markervarianten in DNA pools mogelijk te maken, terwijl sporadische gevallen waarin dit niet mogelijk was tijdig konden worden opgemerkt.

Deze nieuwe methode voor stuttercorrectie maakt efficiënte analyse van microsatelliet markers in grote groepen patiënten en controles mogelijk.

Voor het analyseren van het tweede veelgebruikte type markers, SNPs, zijn vele technieken beschreven, doch deze zijn veelal duur en bewerkelijk per bepaling, terwijl technieken met een lagere prijs per typering en een grote capaciteit vaak grote initiële investeringen in de benodigde apparatuur vereisen. **Hoofdstuk 10** beschrijft een nieuwe techniek voor SNP typering op apparatuur die in veel genetische laboratoria reeds voorhanden is, namelijk DNA sequencers. De methode is gebaseerd op een allel-specifieke PCR reactie, die geen aanvullende bewerkingen vereist. Daarnaast is de typering relatief goedkoop, aangezien voor alle verschillende bepalingen een universeel fluorescerend label gebruikt kan worden, dat tijdens de PCR wordt ingebouwd. De methode werd gevalideerd door de resultaten te vergelijken met die welke verkregen waren met behulp van de bestaande TaqMan techniek en direct sequencen. De resultaten gaven aan dat de nieuwe methode geschikt is om de meeste SNPs zonder verdere optimalisatie betrouwbaar te analyseren.

12.4. Conclusie

Het in dit proefschrift beschreven onderzoek heeft aanwijzingen opgeleverd voor het bestaan van genetisch verschillende vormen van schizofrenie. Het kunnen onderscheiden van specifieke ziektevormen maakt het op termijn wellicht mogelijk om het beloop in individuele patienten beter te kunnen voorspellen, en behandelingen gerichter te kunnen kiezen. Het beschreven koppelingsonderzoek van het totale DNA in families met ADHD vormt een eerste aanzet tot het identificeren van mogelijk geheel nieuwe bij deze ziekte betrokken genen.

Na jaren van moeizame vooruitgang van het genetisch onderzoek naar de biologische grondslagen van psychiatrische aandoeningen, lijken de gezamenlijke inspanningen van vele onderzoeksgroepen nu de eerste overtuigende aanwijzingen op te leveren voor betrokkenheid van specifieke genen. Hopelijk zullen deze resultaten de komende jaren leiden tot inzicht in de oorzaken van psychiatrische aandoeningen, en tot nieuwe mogelijkheden voor preventie en behandeling

CONTRIBUTING AUTHORS AND AFFILIATIONS

S.C. Bakker, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht (present: Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht)

J.K. Buitelaar, Dept. of Psychiatry, University Medical Centre Nijmegen, Nijmegen

S. Caron, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

B.M. Groot, Institute of Information and Computing Sciences, Utrecht University, Utrecht

W.B. Gunning, Dept. of Child and Adolescent Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam

F.M.M.A. van der Heijden, Vincent van Gogh Institute for Psychiatry, Venray

J.C.J.M. Hendriks, Rudolf Magnus Institute of Neuroscience, Dept. of Pharmacology and Anatomy, University Medical Centre Utrecht, Utrecht

M.L.C. Hoogendoorn, Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht

R.S. Kahn, Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht

B.P.C. Koeleman, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

K. Kusters, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

E.M. van der Meulen, Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht

R.B. Minderaa, University Centre for Child and Adolescent Psychiatry, Groningen

A.J. Monsuur, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

N. Oteman, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

H.G. Otten, Dept. of Immunology, University Medical Centre Utrecht, Utrecht

D.L. Pauls, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

P.L. Pearson, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

L.A. Sandkuijl († 4 December 2002), Dept. of Medical Statistics, Leiden University Medical Centre, Leiden

H. Schelleman, Dept. of Biostatistics and Epidemiology, Erasmus MC, Rotterdam

H.G. Schnack, Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht

J-P.C. Selten, Rudolf Magnus Institute of Neuroscience, Dept. of Psychiatry, University Medical Centre Utrecht, Utrecht

R. J. Sinke, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

R. van 't Slot, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

W. Verduijn, Dept. of Immunohematology and Blood Transfusion, Leiden University Medical Centre, Leiden

K. Verzijlbergen, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

P.H.A. van Zon, Dept. of Biomedical Genetics, University Medical Centre Utrecht, Utrecht

NAWOORD

Gedurende het in dit proefschrift beschreven onderzoek vond een omslag plaats in het genetisch onderzoek van psychiatrische ziekten. Na jaren van verwarrende resultaten en twijfel aan de haalbaarheid van de methoden werden de eerste overtuigende aanwijzingen gevonden voor betrokkenheid van specifieke genen. Naar zich laat aanzien zullen de gevolgen voor het denken over psychiatrische stoornissen en voor het voorkomen en behandelen ervan verstrekkend zijn. Het was een belevenis om deze omwenteling mee te maken, en er zelf aan te kunnen bijdragen. De voltooiing van dit proefschrift zie ik als het begin van een ontdekkingsreis die ik graag voortzet. Zeer velen hebben geholpen om de eerste stappen te zetten. Ik dank hen allen, mij realiserend dat ik niet iedereen bij naam zal kunnen noemen. De honderden patiënten en familieleden wil ik danken voor hun bereidheid om wat van hun tijd en erfelijke materiaal af te staan. Ik dank de medepromovendi van de Complexe Genetica Groep die mij voorgingen, en degenen die spoedig volgen, voor hulp en discussies, en meer in het algemeen voor de plezierige tijd op het lab en in de hectische werkkamers. Analisten die niet direct bij het project betrokken waren, zowel van het research lab als van de DNA diagnostiek, dank ik voor hun raad, daad (zoals het tijdrovende isoleren van DNA), en de soms benodigde bemoedigende woorden. Medeauteurs dank ik voor hun hulp bij het tot stand komen van de artikelen.

Mijn promotoren dank ik voor de moed om een onbeschreven blad op het gebied van de genetica in een laboratorium los te laten. Prof.dr. R.S. Kahn begeleidde, en begeleidt, het onderzoek op directe en geïnteresseerde wijze, altijd bereikbaar voor vragen of het kritisch doornemen van schrijfwerk. Prof.dr. P.L. Pearson kwam vaak met volstrekt originele– hoewel niet altijd eenvoudige- oplossingen voor problemen, en liet tijdens soms urenlange correctiesessies van manuscripten zien hoe belangrijk het is, om elk woord te wegen. Dagelijkse begeleider en copromotor Richard Sinke dank ik voor de gelegenheid om zelfstandig te werken en eens een gok te wagen, en voor de plezierige samenwerking en begeleiding.

Mechteld Hoogendoorn en Emma van der Meulen hebben formidabel werk verricht bij het verzamelen van patiënten en families. Beiden dank ik als medeonderzoekers, met wie ik intens en goed heb samengewerkt, maar ook als sympathieke reisgenoten naar verre oorden. Uit de bijzondere uitwisseling van ideeën met Hugo Schnack, op het raakvlak van genetica en wiskunde, groeide een oplossing voor een lastig stoorsignaal bij markeranalyses.

Analisten Ruben van 't Slot, Karlijn Kusters, Judith Hendriks en Alfons Bardoel dank ik voor hun hulp bij het uitdenken en uitvoeren van de (niet altijd even routinematige) experimenten. Veel dank is ook verschuldigd aan Nicole Oteman, Sander Haijma, Hedi Schelleman, Bart Groot, Alienke Monsuur, Kitty Verzijlbergen, Mark Willis en Suzan Caron, die als student

hebben bijgedragen aan de uitvoering van de experimenten, het doen van literatuuronderzoek en het schrijven van computerprogramma's.
Jackie Senior heeft als 'native speaker' een belangrijk deel van dit proefschrift minutieus gecontroleerd op 'Double Dutch', en vele zaken rond de promotie geregeld, terwijl Harry van Someren nauwgezet alle gegevens van de DNA monsters in de database heeft ingevoerd. Bobby Koeleman dank ik voor zijn hulp bij het uitvoeren van de associatiestudies.
Prof.dr. Burbach, Prof.dr. Wijmenga, Prof.dr. Heutink en Prof.dr. Van Engeland wil ik danken voor het kritisch lezen van het manuscript, en de waardevolle opmerkingen ter verbetering. Prof.dr. Wijmenga dank ik daarbij voor de kritische discussies en adviezen gedurende het gehele project. Hoewel het persoonlijk contact beperkt is gebleven, heeft Prof.dr. J.K. Buitelaar een grote rol gespeeld in het ADHD onderzoek. Evenzo dank ik Jean-Paul Selten voor zijn begeleiding aan de 'klinische zijde' van het schizofrenieonderzoek. Mijn begeleiders in het eerste jaar van de opleiding, Marion Scholten en Wiepke Cahn, hebben momenten van verstrooidheid door de vingers gezien, en ruimte voor onderzoek gelaten. Door zijn plotse overlijden heeft Lodewijk Sandkuijl de voltooiing van dit werk, waarbij hij als statisticus onmisbaar was, helaas niet meer kunnen meemaken. Zijn heldere manier van denken en uitleggen, en tegelijk zijn vriendelijkheid en bescheidenheid zullen velen tot voorbeeld blijven.
De paranimfen dank ik bij voorbaat voor het geven van flankdekking op het moment suprème- evenzeer als bij een huwelijk bleek het organisatietalent van Rogier van Huffelen onmisbaar bij een promotie.

'O'a en O'a' Jonges dank ik namens Michiel voor de aandacht die zij hem gaven als hij op de dinsdag verwaarloosd dreigde te worden omwille van de wetenschap, en schoonpa Jonges daarnaast voor de prikkelende discussies over 'mind' en 'brain', en belangrijker zaken.
Mijn ouders dank ik voor alles waarmee zij mij in aanraking hebben gebracht, en de mogelijkheden die zij hebben gegeven om me te ontwikkelen in de richting die ik verkoos.
Michiel dank ik voor zijn geduld als papa weer eens achter de 'Toeter' zat, en voor het inzicht dat het allemaal maar spel is- maar dat spelen is, waar het in het leven werkelijk om gaat.
Liesbeth, tenslotte, dank ik voor het belangstellend aanhoren van verhalen over piekjes en snips, het kritisch lezen en de hulp bij de opmaak. Onmogelijk kan ik haar echter voldoende danken voor haar steun, en voor alles wat zij zich zonder verwijten heeft ontzegd wanneer werk en thuis weer eens werden verward.
Zij was er-

'...all the way'.

ELECTRONIC DATABASE INFORMATION

Celera database, http://www.celera.com/

CEPH/Généthon database, http://www.cephb.fr/cgi-bin/wdb/ceph/systeme/form

CLUMP (DOS version), http://www.mds.qmw.ac.uk/statgen/dcurtis/software.html

DbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/

Ensembl genome browser, http://www.ensembl.org/

Exiqon LNA melting temperature prediction, http://lna-tm.com/

Genome Database (mirror site), http://gdbwww.dkfz-heidelberg.de/

GRR, http://qtl.well.ox.ac.uk/GRR//

International HapMap Project, http://www.hapmap.org/index.html.en

Marshfield Centre for Medical Genetics, http://research.marshfieldclinic.org/genetics/

Netprimer, http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim

PoolFitter program, http://www.smri.nl/microsatellites

Primer3, http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

Tandem Repeat Finder, http://c3.biomath.mssm.edu/trf.html

TRANSMIT software, http://www-gene.cimr.cam.ac.uk/clayton/software/

CRITERIA FOR DEFICIT SCHIZOPHRENIA [36]

At least two of the following six features must be present and of clinically significant severity:

- Restricted effect
- Diminished emotional range
- Poverty of speech
- Curbing of interests
- Diminished sense of purpose
- Diminished social drive

Two or more of these features must have been present for the preceding 12 months, and always have been present during periods of clinical stability (including chronic psychotic states). These symptoms may or may not be detectable during transient episodes of acute psychotic disorganization or decompensation.

Two or more of these enduring features are also idiopathic, i.e., not secondary to factors other than the disease process.
Such factors include:

- Anxiety
- Drug effect
- Suspiciousness
- Formal thought disorder
- Hallucinations or delusions
- Mental retardation
- Depression

The patient meets DSM criteria for schizophrenia.

CURRICULUM VITAE

De auteur van dit proefschrift werd op 3 december 1970 geboren te Woerden.

In 1983 begon hij met voorbereidend wetenschappelijk onderwijs aan het Murmellius Gymnasium te Alkmaar, waar hij in 1989 het diploma gymnasium β behaalde. Aansluitend studeerde hij Geneeskunde aan de Rijksuniversiteit Leiden. Na het behalen van het artsexamen in 1998 werkte de auteur tot 1999 in Ziekenhuis Bronovo te 's-Gravenhage als arts-assistent Interne Geneeskunde, onder leiding van Dr. R. Bieger. Het in dit proefschrift beschreven promotieonderzoek nam een aanvang in 1999, onder leiding van Prof.dr. R.S. Kahn and Prof.dr. P.L. Pearson. Vanaf maart 2004 is de auteur in opleiding tot psychiater in het UMC Utrecht, met als opleider Prof.dr. R.S. Kahn. Zijn ervaring in het wetenschappelijk onderzoek hoopt hij vanaf september 2005 te verbreden met een verblijf aan het Institute of Psychiatry te Londen, onder leiding van Prof.dr. R.M. Murray.

PUBLICATIONS AND PRESENTATIONS

Publications

Bakker SC, Hoogendoorn MLC, Hendriks J, Verzijlbergen S, Caron S, Otten HG, Verduijn W, Selten JP, Pearson PL, Kahn RS, Sinke RJ. *Association of deficit and non-deficit schizophrenia with the PIP5K2A and RGS4 genes, but not with the dysbindin and G72/G30 genes.* Submitted.

Van der Meulen EM, Bakker SC, Pauls DL, Oteman N, Kruitwagen CLJJ, Pearson PL, Sinke RJ, Buitelaar JK. *High sibling correlation on methylphenidate response but no association with DAT1-10R homozygosity in Dutch sibpairs with ADHD.* Journal of Child Psychology and Psychiatry, accepted for publication.

Hoogendoorn MLC, Bakker SC, Schnack HG, Selten JP, Otten HG, Verduijn W, van der Heijden FMMA, Pearson PL, Kahn RS, Sinke RJ (2005) *No association between 12 dopaminergic genes and schizophrenia in a large Dutch sample.* Am J Med Genet B Neuropsychiatr Genet 134(1):6-9.

Bakker SC, van der Meulen EM, Oteman N, Schelleman H, Pearson PL, Buitelaar JK, Sinke RJ (2005) *DAT1, DRD4 and DRD5 polymorphisms are not associated with ADHD in Dutch families.* Am J Med Genet B Neuropsychiatr Genet.132(1):50-2.

Bakker SC, Hoogendoorn ML, Selten JP, Verduijn W, Pearson PL, Sinke RJ, Kahn RS (2004) *Neuregulin 1: genetic support for schizophrenia subtypes.* Mol Psychiatry 12:1061-63.

Schnack HG*, Bakker SC*, van ’t Slot R, Groot BM, Sinke RJ, Kahn RS, Pearson PL (2004) *Accurate Determination of Microsatellite Allele Frequencies in Pooled DNA Samples.* Eur J Hum Genet 12:925-934 (*equal contributions).

Bakker SC*, van der Meulen EM*, Buitelaar JK, Sandkuijl LA, Pauls DL, Monsuur AJ, van ‘t Slot R, Minderaa RB, Gunning WB, Pearson PL, Sinke RJ (2003) *A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q.* Am J Hum Genet 72(5):1251-60. (*equal contributions).

Bakker SC, Zanin DE, Zweers EJ (2002) *[Treatment of hyperthyroidism caused by Graves' disease or toxic multinodular goitre by radioiodine: over 80% cure retrospectively after one calculated dose.]* Ned Tijdschr Geneeskd. 146(39):1837-41.

Toes RE, Kast WM, Blom RJ, Bakker SC, Offringa R, Melief CJ (1996) *Efficient tumor eradication by adoptively transferred cytotoxic T-cell clones in allogeneic hosts.* Int J Cancer 66(5):686-91.

Oral presentations

A new neuregulin haplotype is associated with schizophrenia in the Dutch population (2004) Twelfth Winter Workshop on Schizophrenia, Davos (abstract published in Schizophrenia Research, 68-1 supp 1, p29).

De genetica van Attention-deficit hyperactivity disorder (2004) Voorjaarscongres van de Nederlandse Vereniging voor Psychiatrie, Maastricht.

A whole-genome scan in 164 Dutch sib pairs with Attention-deficit hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q (2003) Voorjaarscongres Nederlandsche Anthropogenetische Vereniging, Veldhoven.

Hoe vindt men een schizofrenie-gen? (2003) Refereeravond psychiatrie, Utrecht

Genen en gedrag (2003) Medisch Interfacultair Congres, Leiden.

A whole-genome scan in 164 Dutch sib pairs with Attention-deficit hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q (2002) World

Congress for Psychiatric Genetics, Brussels (abstract published in Am J Med Genet 114(7), p735).

Analysis of genes from the dopamine system using a DNA pooling approach (2002) Fourth annual ADHD molecular genetics meeting, Dublin.

The genetics of attention-deficit hyperactivity disorder (2001) Marshfield Medical Research Foundation, Marshfield, USA.

DNA pooling as an efficient screening tool for ADHD candidate gene studies: correction of PCR-induced artifacts (2001) World Congress for Psychiatric Genetics, St. Louis (abstract published in Am J Med Genet 105(7), p566).

Poster presentations

Bakker SC, Hoogendoorn MLC, Selten JP, Verduijn W, Pearson PL, Kahn RS, Sinke RJ (2004) *The role of the Neuregulin 1, Dysbindin and G72/G30 genes in Dutch schizophrenia patients with and without prominent negative symptoms.* World Congress for Psychiatric Genetics, Dublin (abstract published in Am J Med Genet 130B(1), p135).

Bakker SC, Van der Meulen EM, Kusters K, Koeleman BPC, Verduijn W, Buitelaar JK, Pearson PL, Sinke RJ (2004) *Association analysis of dopa decarboxylase, a functional and positional candidate gene for attention-deficit hyperactivity disorder.* World Congress for Psychiatric Genetics, Dublin (abstract published in Am J Med Genet 130B(1), p101).

Hoogendoorn MLC, Vorstman JAS, Bakker SC, Sinke RJ, Beemer FA, Kahn RS (2004) *Prevalence of 22Q11 deletions in a Dutch population of patients with deficit schizophrenia.* World Congress for Psychiatric Genetics, Dublin (abstract published in Am J Med Genet 130B(1), p73).

Hoogendoorn MLC, Bakker SC, Van Haren NE, Selten JP, Sinke RJ, Pearson PL, Kahn RS (2004) *Association between polymorphisms in the G72 gene and lateral ventricle volume in schizophrenia.* World Congress for Psychiatric Genetics, Dublin (abstract published in Am J Med Genet 130B(1), p57).

Bakker SC, Kusters KA, Van der Meulen EM, Schnack HG, Otten HG, Buitelaar JK, Pearson PL, Sinke RJ. *Fine mapping of an ADHD linkage region on chromosome 15 using a dense set of microsatellites and DNA pools* (2003) World Congress for Psychiatric Genetics, Québec (abstract published in Am J Med Genet 122B(1): p147).

Schnack HG, Bakker SC, Sinke RJ, Kahn RS, Pearson PL (2002) *A generalized correction method for DNA pool patterns.* World Congress for Psychiatric Genetics, Brussels (abstract published in Am J Med Genet 114(7): p786).

Hoogendoorn MLC, Bakker SC, Schnack HG, Sandkuijl LA, Otten HG, Selten JP, Pearson PL, Sinke RJ, Kahn RS (2002) *A DNA pooling based association study of 10 dopaminergic genes in a large sample of Dutch schizophrenic patients.* World Congress for Psychiatric Genetics, Brussels (abstract published in Am J Med Genet 114(7): p842).

Van der Meulen EM, Bakker SC, Pauls DL, Sinke RJ and Buitelaar JK (2001) *The genetics of social skills and the association with candidate genes in a sample of 140 dutch sibpairs with ADHD.* World Congress for Psychiatric Genetics, St. Louis (abstract published in Am J Med Genet 105(7): p633).

Bakker SC, Van Belzen MJ, Sandkuijl LA, Wijmenga C and Sinke RJ (2000) *DNA pooling: towards a screening tool for genome-wide association studies in complex disorders.* Najaarscongres NAV, Noordwijkerhout.

Bakker SC, Oteman N, Van der Meulen EM, Buitelaar JK, Pearson PL and Sinke RJ (2000) *A family-based association study of a Dutch ADHD population: analysis of the DRD4 and DAT1 genes.* European Human Genetics Conference, Amsterdam (abstract published in Eur J Hum Genet 8(suppl.1): p121)

REFERENCES

1. Tandon, K. and P. McGuffin, *The genetic basis for psychiatric illness in man.* Eur J Neurosci, 2002. **16**(3): p. 403-7.
2. Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
3. McPherson, J.D., M. Marra, L. Hillier, R.H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E.R. Mardis, R.K. Wilson, et al., *A physical map of the human genome.* Nature, 2001. **409**(6822): p. 934-41.
4. Reich, D.E. and E.S. Lander, *On the allelic spectrum of human disease.* Trends Genet, 2001. **17**(9): p. 502-10.
5. Petronis, A., *The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis.* Biol Psychiatry, 2004. **55**(10): p. 965-70.
6. Burmeister, M., *Basic concepts in the study of diseases with complex genetics.* Biol Psychiatry, 1999. **45**(5): p. 522-32.
7. Pritchard, J.K., *Are rare variants responsible for susceptibility to complex diseases?* Am J Hum Genet, 2001. **69**(1): p. 124-37.
8. Pritchard, J.K. and N.J. Cox, *The allelic architecture of human disease genes: common disease-common variant...or not?* Hum Mol Genet, 2002. **11**(20): p. 2417-23.
9. North, B.V., D. Curtis, E.R. Martin, E.H. Lai, A.D. Roses, and P.C. Sham, *Further investigation of linkage disequilibrium SNPs and their ability to identify associated susceptibility loci.* Ann Hum Genet, 2004. **68**(Pt 3): p. 240-8.
10. Lohmueller, K.E., C.L. Pearce, M. Pike, E.S. Lander, and J.N. Hirschhorn, *Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.* Nat Genet, 2003. **33**(2): p. 177-82.
11. Pritchard, J.K. and M. Przeworski, *Linkage disequilibrium in humans: models and data.* Am J Hum Genet, 2001. **69**(1): p. 1-14.
12. Cardon, L.R. and G.R. Abecasis, *Using haplotype blocks to map human complex trait loci.* Trends Genet, 2003. **19**(3): p. 135-40.
13. Jorde, L.B., *Linkage disequilibrium and the search for complex disease genes [In Process Citation].* Genome Res, 2000. **10**(10): p. 1435-44.
14. Weber, J.L. and P.E. May, *Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction.* Am J Hum Genet, 1989. **44**(3): p. 388-96.
15. Michalik, A. and C. Van Broeckhoven, *Pathogenesis of polyglutamine disorders: aggregation revisited.* Hum Mol Genet, 2003. **12 Spec No 2**: p. R173-86.
16. Kwok, P.Y., *Methods for genotyping single nucleotide polymorphisms.* Annu Rev Genomics Hum Genet, 2001. **2**: p. 235-58.
17. Chen, X. and P.F. Sullivan, *Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput.* Pharmacogenomics J, 2003. **3**(2): p. 77-96.
18. Cardon, L.R. and J.I. Bell, *Association study designs for complex diseases.* Nat Rev Genet, 2001. **2**(2): p. 91-9.
19. Ardlie, K.G., L. Kruglyak, and M. Seielstad, *Patterns of linkage disequilibrium in the human genome.* Nat Rev Genet, 2002. **3**(4): p. 299-309.
20. Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, et al., *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.
21. Daly, M.J., J.D. Rioux, S.F. Schaffner, T.J. Hudson, and E.S. Lander, *High-resolution haplotype structure in the human genome.* Nat Genet, 2001. **29**(2): p. 229-32.
22. Johnson, G.C., L. Esposito, B.J. Barratt, A.N. Smith, J. Heward, G. Di Genova, H. Ueda, H.J. Cordell, I.A. Eaves, F. Dudbridge, et al., *Haplotype tagging for the identification of common disease genes.* Nat Genet, 2001. **29**(2): p. 233-7.
23. HapMap-Consortium, *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
24. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping.* Genomics, 1995. **29**(2): p. 311-22.
25. Tabor, H.K., N.J. Risch, and R.M. Myers, *OPINION: Candidate-gene approaches for studying complex genetic traits: practical considerations.* Nat Rev Genet, 2002. **3**(5): p. 391-7.
26. Mohn, A.R., R.R. Gainetdinov, M.G. Caron, and B.H. Koller, *Mice with reduced NMDA receptor expression display behaviors related to schizophrenia.* Cell, 1999. **98**(4): p. 427-36.
27. Davids, E., K. Zhang, F.I. Tarazi, and R.J. Baldessarini, *Animal models of attention-deficit hyperactivity disorder.* Brain Res Brain Res Rev, 2003. **42**(1): p. 1-21.
28. Cardon, L.R. and L.J. Palmer, *Population stratification and spurious allelic association.* Lancet, 2003. **361**(9357): p. 598-604.
29. Devlin, B. and K. Roeder, *Genomic control for association studies.* Biometrics, 1999. **55**(4): p. 997-1004.
30. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-59.

31. Risch, N. and K.R. Merikangas, *Linkage studies of psychiatric disorders.* Eur Arch Psychiatry Clin Neurosci, 1993. **243**(3-4): p. 143-9.

32. Weber, J.L. and K.W. Broman, *Genotyping for human whole-genome scans: past, present, and future.* Adv Genet, 2001. **42**: p. 77-96.

33. Mueser, K.T. and S.R. McGurk, *Schizophrenia.* Lancet, 2004. **363**(9426): p. 2063-72.

34. Roy, M.A., C. Merette, and M. Maziade, *Subtyping schizophrenia according to outcome or severity: a search for homogeneous subgroups.* Schizophr Bull, 2001. **27**(1): p. 115-38.

35. Keefe, R.S., E. Frescka, S.H. Apter, M. Davidson, J.M. Macaluso, J. Hirschowitz, and K.L. Davis, *Clinical characteristics of Kraepelinian schizophrenia: replication and extension of previous findings.* Am J Psychiatry, 1996. **153**(6): p. 806-11.

36. Kirkpatrick, B., R.W. Buchanan, P.D. McKenney, L.D. Alphs, and W.T. Carpenter, Jr., *The Schedule for the Deficit syndrome: an instrument for research in schizophrenia.* Psychiatry Res, 1989. **30**(2): p. 119-23.

37. Carpenter, W.T., Jr., D.W. Heinrichs, and A.M. Wagman, *Deficit and nondeficit forms of schizophrenia: the concept.* Am J Psychiatry, 1988. **145**(5): p. 578-83.

38. Kirkpatrick, B., R.W. Buchanan, D.E. Ross, and W.T. Carpenter, Jr., *A separate disease within the syndrome of schizophrenia.* Arch Gen Psychiatry, 2001. **58**(2): p. 165-71.

39. Thibaut, F. and M. Petit, *The deficit syndrome of schizophrenia: towards heterogeneity.* Psychopathology, 1997. **30**(5): p. 257-62.

40. Thibaut, F., J.M. Ribeyre, N. Dourmap, J.F. Menard, S. Dollfus, and M. Petit, *Plasma 3-methoxy-4-hydroxyphenylglycol and homovanillic acid measurements in deficit and nondeficit forms of schizophrenia.* Biol Psychiatry, 1998. **43**(1): p. 24-30.

41. Ross, D.E., *The deficit syndrome and eye tracking disorder may reflect a distinct subtype within the syndrome of schizophrenia.* Schizophr Bull, 2000. **26**(4): p. 855-66.

42. Fouldrin, G., M. Jay, F. Bonnet-Brilhault, J.F. Menard, M. Petit, and F. Thibaut, *Concordance of deficit and non-deficit subtypes in siblings affected with schizophrenia.* Psychiatry Res, 2001. **102**(1): p. 59-64.

43. Kirkpatrick, B., D.E. Ross, D. Walsh, L. Karkowski, and K.S. Kendler, *Family characteristics of deficit and nondeficit schizophrenia in the Roscommon Family Study.* Schizophr Res, 2000. **45**(1-2): p. 57-64.

44. Wright, I.C., S. Rabe-Hesketh, P.W. Woodruff, A.S. David, R.M. Murray, and E.T. Bullmore, *Meta-analysis of regional brain volumes in schizophrenia.* Am J Psychiatry, 2000. **157**(1): p. 16-25.

45. Haren, N.E.M.v., H.E. Hulshoff Pol, H.G. Schnack, W. Cahn, R.C.W. Mandl, L.D. Collins, A.C. Evans, and R.S. Kahn, *Focal gray and white matter changes in schizophrenia across the course of the illness: a 5-year follow-up study.* PhD. thesis, 2004.

46. Cantor-Graae, E. and J.P. Selten, *Schizophrenia and migration: a meta-analysis and review.* Am J Psychiatry, 2005. **162**(1): p. 12-24.

47. Cannon, M., P.B. Jones, and R.M. Murray, *Obstetric complications and schizophrenia: historical and meta-analytic review.* Am J Psychiatry, 2002. **159**(7): p. 1080-92.

48. Sullivan, P.F., K.S. Kendler, and M.C. Neale, *Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies.* Arch Gen Psychiatry, 2003. **60**(12): p. 1187-92.

49. Abdolmaleky, H.M., S.V. Faraone, S.J. Glatt, and M.T. Tsuang, *Meta-analysis of association between the T102C polymorphism of the 5HT2a receptor gene and schizophrenia.* Schizophr Res, 2004. **67**(1): p. 53-62.

50. Williams, J., P. McGuffin, M. Nothen, and M.J. Owen, *Meta-analysis of association between the 5-HT2a receptor T102C polymorphism and schizophrenia. EMASS Collaborative Group. European Multicentre Association Study of Schizophrenia.* Lancet, 1997. **349**(9060): p. 1221.

51. Jonsson, E.G., A. Sillen, M. Vares, B. Ekholm, L. Terenius, and G.C. Sedvall, *Dopamine D2 receptor gene Ser311Cys variant and schizophrenia: association study and meta-analysis.* Am J Med Genet, 2003. **119B**(1): p. 28-34.

52. Shaikh, S., D.A. Collier, P.C. Sham, D. Ball, K. Aitchison, H. Vallada, I. Smith, M. Gill, and R.W. Kerwin, *Allelic association between a Ser-9-Gly polymorphism in the dopamine D3 receptor gene and schizophrenia.* Hum Genet, 1996. **97**(6): p. 714-9.

53. Williams, J., G. Spurlock, P. Holmans, R. Mant, K. Murphy, L. Jones, A. Cardno, P. Asherson, D. Blackwood, W. Muir, et al., *A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia.* Mol Psychiatry, 1998. **3**(2): p. 141-9.

54. Jonsson, E.G., R. Kaiser, J. Brockmoller, V.L. Nimgaonkar, and M.A. Crocq, *Meta-analysis of the dopamine D3 receptor gene (DRD3) Ser9Gly variant and schizophrenia.* Psychiatr Genet, 2004. **14**(1): p. 9-12.

55. Glatt, S.J., S.V. Faraone, and M.T. Tsuang, *Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: meta-analysis of case-control and family-based studies.* Am J Psychiatry, 2003. **160**(3): p. 469-76.

56. Millar, J.K., J.C. Wilson-Annan, S. Anderson, S. Christie, M.S. Taylor, C.A. Semple, R.S. Devon, D.M. Clair, W.J. Muir, D.H. Blackwood, et al., *Disruption of two novel genes by a*

translocation co-segregating with schizophrenia. Hum Mol Genet, 2000. **9**(9): p. 1415-23.

57. Hodgkinson, C.A., D. Goldman, J. Jaeger, S. Persaud, J.M. Kane, R.H. Lipsky, and A.K. Malhotra, *Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder.* Am J Hum Genet, 2004. **75**(5): p. 862-72.

58. Murphy, K.C., *Schizophrenia and velo-cardio-facial syndrome.* Lancet, 2002. **359**(9304): p. 426-30.

59. Williams, N.M. and M.J. Owen, *Genetic abnormalities of chromosome 22 and the development of psychosis.* Curr Psychiatry Rep, 2004. **6**(3): p. 176-82.

60. Karayiorgou, M. and J.A. Gogos, *The molecular genetics of the 22q11-associated schizophrenia.* Brain Res Mol Brain Res, 2004. **132**(2): p. 95-104.

61. de Chaldee, M., C. Laurent, F. Thibaut, M. Martinez, D. Samolyk, M. Petit, D. Campion, and J. Mallet, *Linkage disequilibrium on the COMT gene in French schizophrenics and controls.* Am J Med Genet, 1999. **88**(5): p. 452-7.

62. Chakravarti, A., *A compelling genetic hypothesis for a complex disease: PRODH2/DGCR6 variation leads to schizophrenia susceptibility.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 4755-6.

63. Lewis, C.M., D.F. Levinson, L.H. Wise, L.E. DeLisi, R.E. Straub, I. Hovatta, N.M. Williams, S.G. Schwab, A.E. Pulver, S.V. Faraone, et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia.* Am J Hum Genet, 2003. **73**(1): p. 34-48.

64. Badner, J.A. and E.S. Gershon, *Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia.* Mol Psychiatry, 2002. **7**(4): p. 405-11.

65. Owen, M.J., N.M. Williams, and M.C. O'Donovan, *The molecular genetics of schizophrenia: new findings promise new insights.* Mol Psychiatry, 2004. **9**(1): p. 14-27.

66. Stefansson, H., E. Sigurdsson, V. Steinthorsdottir, S. Bjornsdottir, T. Sigmundsson, S. Ghosh, J. Brynjolfsson, S. Gunnarsdottir, O. Ivarsson, T.T. Chou, et al., *Neuregulin 1 and susceptibility to schizophrenia.* Am J Hum Genet, 2002. **71**(4): p. 877-92.

67. Straub, R.E., Y. Jiang, C.J. MacLean, Y. Ma, B.T. Webb, M.V. Myakishev, C. Harris-Kerr, B. Wormley, H. Sadek, B. Kadambi, et al., *Genetic Variation in the 6p22.3 Gene DTNBP1, the Human Ortholog of the Mouse Dysbindin Gene, Is Associated with Schizophrenia.* Am J Hum Genet, 2002. **71**(2).

68. Chumakov, I., M. Blumenfeld, O. Guerassimenko, L. Cavarec, M. Palicio, H. Abderrahim, L. Bougueleret, C. Barry, H. Tanaka, P. La Rosa, et al., *Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13675-80.

69. Mirnics, K., F.A. Middleton, G.D. Stanwood, D.A. Lewis, and P. Levitt, *Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia.* Mol Psychiatry, 2001. **6**(3): p. 293-301.

70. Harrison, P.J. and M.J. Owen, *Genes for schizophrenia? Recent findings and their pathophysiological implications.* Lancet, 2003. **361**(9355): p. 417-9.

71. APA, *Diagnostic and statistical manual of mental disorders.* 4th ed. 1994, Washington, DC: American Psychiatric Association.

72. Buitelaar, J.K., *Epidemiology: what have we learned over the last decade?*, in *Hyperactivity and Attention-Deficit Disorders*, S. Sandberg, Editor. 2002, Cambridge University Press.: Cambridge. p. 30-63.

73. Angold, A., E.J. Costello, and A. Erkanli, *Comorbidity.* J Child Psychol Psychiatry, 1999. **40**(1): p. 57-87.

74. Durston, S., *A review of the biological bases of ADHD: what have we learned from imaging studies?* Ment Retard Dev Disabil Res Rev, 2003. **9**(3): p. 184-95.

75. Barry, R.J., S.J. Johnstone, and A.R. Clarke, *A review of electrophysiology in attention-deficit/hyperactivity disorder: II. Event-related potentials.* Clin Neurophysiol, 2003. **114**(2): p. 184-98.

76. Barry, R.J., A.R. Clarke, and S.J. Johnstone, *A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography.* Clin Neurophysiol, 2003. **114**(2): p. 171-83.

77. Thapar, A., J. Holmes, K. Poulton, and R. Harrington, *Genetic basis of attention deficit and hyperactivity.* Br J Psychiatry, 1999. **174**: p. 105-11.

78. Smalley, S.L., *Genetic influences in childhood-onset psychiatric disorders: autism and attention-deficit/hyperactivity disorder.* Am J Hum Genet, 1997. **60**(6): p. 1276-82.

79. Bobb, A.J., F.X. Castellanos, A.M. Addington, and J.L. Rapoport, *Molecular genetic studies of ADHD: 1991 to 2004.* Am J Med Genet, 2004.

80. Biederman, J. and S.V. Faraone, *Current concepts on the neurobiology of Attention-Deficit/Hyperactivity Disorder.* J Atten Disord, 2002. **6 Suppl 1**: p. S7-16.

81. Maher, B.S., M.L. Marazita, R.E. Ferrell, and M.M. Vanyukov, *Dopamine system genes and attention deficit hyperactivity disorder: a meta-analysis.* Psychiatr Genet, 2002. **12**(4): p. 207-15.

82. Faraone, S.V., A.E. Doyle, E. Mick, and J. Biederman, *Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder.* Am J Psychiatry, 2001. **158**(7): p. 1052-7.

83. Lowe, N., A. Kirley, Z. Hawi, P. Sham, H. Wickham, C.J. Kratochvil, S.D. Smith, S.Y. Lee, F. Levy, L. Kent, et al., *Joint Analysis of the DRD5 Marker Concludes Association with Attention-Deficit/Hyperactivity Disorder Confined to the Predominantly Inattentive and Combined Subtypes.* Am J Hum Genet, 2004. **74**(2): p. 348-56.

84. Norton, N., N.M. Williams, M.C. O'Donovan, and M.J. Owen, *DNA pooling as a tool for large-scale association studies in complex traits.* Ann Med, 2004. **36**(2): p. 146-52.

85. Sham, P., J.S. Bader, I. Craig, M. O'Donovan, and M. Owen, *DNA Pooling: a tool for large-scale association studies.* Nat Rev Genet, 2002. **3**(11): p. 862-71.

86. Le Hellard, S., S.J. Ballereau, P.M. Visscher, H.S. Torrance, J. Pinson, S.W. Morris, M.L. Thomson, C.A. Semple, W.J. Muir, D.H. Blackwood, et al., *SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis.* Nucleic Acids Res, 2002. **30**(15): p. e74.

87. Ito, T., S. Chiku, E. Inoue, M. Tomita, T. Morisaki, H. Morisaki, and N. Kamatani, *Estimation of haplotype frequencies, linkage-disequilibrium measures, and combination of haplotype copies in each pool by use of pooled DNA data.* Am J Hum Genet, 2003. **72**(2): p. 384-98.

88. Bader, J.S., *The relative power of SNPs and haplotype as genetic markers for association tests.* Pharmacogenomics, 2001. **2**(1): p. 11-24.

89. Nielsen, D.M., M.G. Ehm, D.V. Zaykin, and B.S. Weir, *Effect of two- and three-locus linkage disequilibrium on the power to detect marker/phenotype associations.* Genetics, 2004. **168**(2): p. 1029-40.

90. Kirov, G., N. Williams, P. Sham, N. Craddock, and M.J. Owen, *Pooled genotyping of microsatellite markers in parent-offspring trios.* Genome Res, 2000. **10**(1): p. 105-15.

91. Taylor, B.A. and S.J. Phillips, *Detection of obesity QTLs on mouse chromosomes 1 and 7 by selective DNA pooling.* Genomics, 1996. **34**(3): p. 389-98.

92. Simonic, I., G.S. Gericke, J. Ott, and J.L. Weber, *Identification of genetic markers associated with Gilles de la Tourette syndrome in an Afrikaner population.* Am J Hum Genet, 1998. **63**(3): p. 839-46.

93. Daniels, J., P. Holmans, N. Williams, D. Turic, P. McGuffin, R. Plomin, and M.J. Owen, *A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies.* Am J Hum Genet, 1998. **62**(5): p. 1189-97.

94. Fisher, P.J., D. Turic, N.M. Williams, P. McGuffin, P. Asherson, D. Ball, I. Craig, T. Eley, L. Hill, K. Chorney, et al., *DNA pooling identifies QTLs on chromosome 4 for general cognitive ability in children.* Hum Mol Genet, 1999. **8**(5): p. 915-22.

95. Plomin, R., L. Hill, I.W. Craig, P. McGuffin, S. Purcell, P. Sham, D. Lubinski, L.A. Thompson, P.J. Fisher, D. Turic, et al., *A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling and extreme selected groups.* Behav Genet, 2001. **31**(6): p. 497-509.

96. Collins, H.E., H. Li, S.E. Inda, J. Anderson, K. Laiho, J. Tuomilehto, and M.F. Seldin, *A simple and accurate method for determination of microsatellite total allele content differences between DNA pools.* Hum Genet, 2000. **106**(2): p. 218-26.

97. Sawcer, S., M. Maranian, E. Setakis, V. Curwen, E. Akesson, A. Hensiek, F. Coraddu, R. Roxburgh, D. Sawcer, J. Gray, et al., *A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility.* Brain, 2002. **125**(Pt 6): p. 1337-1347.

98. Barratt, B.J., F. Payne, H.E. Rance, S. Nutland, J.A. Todd, and D.G. Clayton, *Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design.* Ann Hum Genet, 2002. **66**(Pt 5-6): p. 393-405.

99. Perlin, M.W., G. Lancia, and S.K. Ng, *Toward fully automated genotyping: genotyping microsatellite markers by deconvolution.* Am J Hum Genet, 1995. **57**(5): p. 1199-210.

100. Barcellos, L.F., W. Klitz, L.L. Field, R. Tobias, A.M. Bowcock, R. Wilson, M.P. Nelson, J. Nagatomi, and G. Thomson, *Association mapping of disease loci, by use of a pooled DNA genomic screen.* Am J Hum Genet, 1997. **61**(3): p. 734-47.

101. Shaw, S.H., M.M. Carrasquillo, C. Kashuk, E.G. Puffenberger, and A. Chakravarti, *Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes.* Genome Res, 1998. **8**(2): p. 111-23.

102. Heutink, P. and B.A. Oostra, *Gene finding in genetically isolated populations.* Hum Mol Genet, 2002. **11**(20): p. 2507-15.

103. Arcos-Burgos, M. and M. Muenke, *Genetics of population isolates.* Clin Genet, 2002. **61**(4): p. 233-47.

104. Varilo, T. and L. Peltonen, *Isolates and their potential use in complex gene mapping efforts.* Curr Opin Genet Dev, 2004. **14**(3): p. 316-23.

105. Wright, A.F., A.D. Carothers, and M. Pirastu, *Population choice in mapping genes for complex diseases.* Nat Genet, 1999. **23**(4): p. 397-404.

106. Ophoff, R.A., M.A. Escamilla, S.K. Service, M. Spesny, D.B. Meshi, W. Poon, J. Molina, E. Fournier, A. Gallegos, C. Mathews, et al., *Genomewide linkage disequilibrium mapping of severe bipolar disorder in a population isolate.* Am J Hum Genet, 2002. **71**(3): p. 565-74.

107. Varilo, T., T. Paunio, A. Parker, M. Perola, J. Meyer, J.D. Terwilliger, and L. Peltonen, *The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories.* Hum Mol Genet, 2003. **12**(1): p. 51-9.

108. Latini, V., G. Sole, S. Doratiotto, D. Poddie, M. Memmi, L. Varesi, G. Vona, A. Cao, and M.S. Ristaldi, *Genetic isolates in Corsica (France): linkage disequilibrium extension analysis on the Xq13 region.* Eur J Hum Genet, 2004. **12**(8): p. 613-9.

109. Shifman, S., J. Kuypers, M. Kokoris, B. Yakir, and A. Darvasi, *Linkage disequilibrium patterns of the human genome across populations.* Hum Mol Genet, 2003. **12**(7): p. 771-6.

110. Ioannidis, J.P., E.E. Ntzani, and T.A. Trikalinos, *'Racial' differences in genetic effects for complex diseases.* Nat Genet, 2004. **36**(12): p. 1312-8.

111. Helgason, A., B. Yngvadottir, B. Hrafnkelsson, J. Gulcher, and K. Stefansson, *An Icelandic example of the impact of population structure on association studies.* Nat Genet, 2005. **37**(1): p. 90-5.

112. Clark, T., C. Feehan, C. Tinline, and P. Vostanis, *Autistic symptoms in children with attention deficit-hyperactivity disorder.* Eur Child Adolesc Psychiatry, 1999. **8**(1): p. 50-5.

113. Gottesman, II and T.D. Gould, *The endophenotype concept in psychiatry: etymology and strategic intentions.* Am J Psychiatry, 2003. **160**(4): p. 636-45.

114. Almasy, L. and J. Blangero, *Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design.* Am J Med Genet, 2001. **105**(1): p. 42-4.

115. Glatt, S.J., S.V. Faraone, and M.T. Tsuang, *Meta-analysis identifies an association between the dopamine D2 receptor gene and schizophrenia.* Mol Psychiatry, 2003. **8**(11): p. 911-5.

116. Sham, P.C. and D. Curtis, *Monte Carlo tests for associations between disease and alleles at highly polymorphic loci.* Ann Hum Genet, 1995. **59 (Pt 1)**: p. 97-105.

117. Andreasen, N.C., M. Flaum, and S. Arndt, *The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology.* Arch Gen Psychiatry, 1992. **49**(8): p. 615-23.

118. Schnack, H.G., S.C. Bakker, R. van 't Slot, B.M. Groot, R.J. Sinke, R.S. Kahn, and P.L. Pearson, *Accurate determination of microsatellite allele frequencies in pooled DNA samples.* Eur J Hum Genet, 2004. **12**(11): p. 925-34.

119. Raymond, M. and F. Rousset, *GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism.* J Heredity, 1995. **86**: p. 248-9.

120. Kendler, K.S., C.J. MacLean, Y. Ma, F.A. O'Neill, D. Walsh, and R.E. Straub, *Marker-to-marker linkage disequilibrium on chromosomes 5q, 6p, and 8p in Irish high-density schizophrenia pedigrees.* Am J Med Genet, 1999. **88**(1): p. 29-33.

121. Schork, N.J., *Power calculations for genetic association studies using estimated probability distributions.* Am J Hum Genet, 2002. **70**(6): p. 1480-9.

122. Purcell, S., S.S. Cherny, and P.C. Sham, *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.* Bioinformatics, 2003. **19**(1): p. 149-50.

123. Muir, W.J., M.L. Thomson, P. McKeon, L. Mynett-Johnson, C. Whitton, K.L. Evans, D.J. Porteous, and D.H. Blackwood, *Markers close to the dopamine D5 receptor gene (DRD5) show significant association with schizophrenia but not bipolar disorder.* Am J Med Genet, 2001. **105**(2): p. 152-8.

124. Williams, N.M., A. Preece, G. Spurlock, N. Norton, H.J. Williams, S. Zammit, M.C. O'Donovan, and M.J. Owen, *Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia.* Mol Psychiatry, 2003. **8**(5): p. 485-7.

125. Stefansson, H., J. Sarginson, A. Kong, P. Yates, V. Steinthorsdottir, E. Gudfinnsson, S. Gunnarsdottir, N. Walker, H. Petursson, C. Crombie, et al., *Association of neuregulin 1 with schizophrenia confirmed in a Scottish population.* Am J Hum Genet, 2003. **72**(1): p. 83-7.

126. Yang, J.Z., T.M. Si, Y. Ruan, Y.S. Ling, Y.H. Han, X.L. Wang, M. Zhou, H.Y. Zhang, Q.M. Kong, C. Liu, et al., *Association study of neuregulin 1 gene with schizophrenia.* Mol Psychiatry, 2003. **8**(7): p. 706-9.

127. Iwata, N., T. Suzuki, M. Ikeda, T. Kitajima, Y. Yamanouchi, T. Inada, and N. Ozaki, *No association with the neuregulin 1 haplotype to Japanese schizophrenia.* Mol Psychiatry, 2004. **9**(2): p. 126-7.

128. Tang, J.X., W.Y. Chen, G. He, J. Zhou, N.F. Gu, G.Y. Feng, and L. He, *Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population.* Mol Psychiatry, 2004. **9**(1): p. 11-2.

129. Corvin, A.P., D.W. Morris, K. McGhee, S. Schwaiger, P. Scully, J. Quinn, D. Meagher, D.S. Clair, J.L. Waddington, and M. Gill, *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, 2004. **9**(2): p. 208-13.

130. Dudbridge, F., *Pedigree disequilibrium tests for multilocus haplotypes.* Genet Epidemiol, 2003. **25**(2): p. 115-21.

131. Stefansson, H., T.E. Thorgeirsson, J.R. Gulcher, and K. Stefansson, *Neuregulin 1 in schizophrenia: out of Iceland.* Mol Psychiatry, 2003. **8**(7): p. 639-40.

132. Williams, N.M., A. Preece, D.W. Morris, G. Spurlock, N.J. Bray, M. Stephens, N. Norton, H. Williams, M. Clement, S. Dwyer, et al., *Identification in 2 Independent Samples of a Novel Schizo-*

phrenia Risk Haplotype of the Dystrobrevin Binding Protein Gene (DTNBP1). Arch Gen Psychiatry, 2004. **61**(4): p. 336-44.

133. Morris, D.W., K.A. McGhee, S. Schwaiger, P. Scully, J. Quinn, D. Meagher, J.L. Waddington, M. Gill, and A.P. Corvin, *No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study.* Schizophr Res, 2003. **60**(2-3): p. 167-72.

134. van den Oord, E.J., P.F. Sullivan, Y. Jiang, D. Walsh, F.A. O'Neill, K.S. Kendler, and B.P. Riley, *Identification of a high-risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families.* Mol Psychiatry, 2003. **8**(5): p. 499-510.

135. Schwab, S.G., M. Knapp, S. Mondabon, J. Hallmayer, M. Borrmann-Hassenbach, M. Albus, B. Lerer, M. Rietschel, M. Trixler, W. Maier, et al., *Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families.* Am J Hum Genet, 2003. **72**(1): p. 185-90.

136. Van Den Bogaert, A., J. Schumacher, T.G. Schulze, A.C. Otte, S. Ohlraun, S. Kovalenko, T. Becker, J. Freudenberg, E.G. Jonsson, M. Mattila-Evenden, et al., *The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease.* Am J Hum Genet, 2003. **73**(6): p. 1438-43.

137. Kirov, G., D. Ivanov, N.M. Williams, A. Preece, I. Nikolov, R. Milev, S. Koleva, A. Dimitrova, D. Toncheva, M.C. O'Donovan, et al., *Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria.* Biol Psychiatry, 2004. **55**(10): p. 971-5.

138. Tang, J.X., J. Zhou, J.B. Fan, X.W. Li, Y.Y. Shi, N.F. Gu, G.Y. Feng, Y.L. Xing, J.G. Shi, and L. He, *Family-based association study of DTNBP1 in 6p22.3 and schizophrenia.* Mol Psychiatry, 2003. **8**(8): p. 717-8.

139. Schumacher, J., R.A. Jamra, J. Freudenberg, T. Becker, S. Ohlraun, A.C. Otte, M. Tullius, S. Kovalenko, A.V. Bogaert, W. Maier, et al., *Examination of G72 and D-amino-acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder.* Mol Psychiatry, 2004. **9**(2): p. 203-7.

140. Wang, X., G. He, N. Gu, J. Yang, J. Tang, Q. Chen, X. Liu, Y. Shen, X. Qian, W. Lin, et al., *Association of G72/G30 with schizophrenia in the Chinese population.* Biochem Biophys Res Commun, 2004. **319**(4): p. 1281-6.

141. Addington, A.M., M. Gornick, A.L. Sporn, N. Gogtay, D. Greenstein, M. Lenane, P. Gochman, N. Baker, R. Balkissoon, R.K. Vakkalanka, et al., *Polymorphisms in the 13q33.2 gene G72/G30 are associated with childhood-onset schizophrenia and psychosis not otherwise specified.* Biol Psychiatry, 2004. **55**(10): p. 976-80.

142. Hattori, E., C. Liu, J.A. Badner, T.I. Bonner, S.L. Christian, M. Maheshwari, S.D. Detera-Wadleigh, R.A. Gibbs, and E.S. Gershon, *Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series.* Am J Hum Genet, 2003. **72**(5): p. 1131-40.

143. Chen, Y.S., N. Akula, S.D. Detera-Wadleigh, T.G. Schulze, J. Thomas, J.B. Potash, J.R. DePaulo, M.G. McInnis, N.J. Cox, and F.J. McMahon, *Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33.* Mol Psychiatry, 2004. **9**(8): p. 811.

144. Chowdari, K.V., K. Mirnics, P. Semwal, J. Wood, E. Lawrence, T. Bhatia, S.N. Deshpande, B.k. T, R.E. Ferrell, F.A. Middleton, et al., *Association and linkage analyses of RGS4 polymorphisms in schizophrenia.* Hum Mol Genet, 2002. **11**(12): p. 1373-1380.

145. Chen, X., C. Dunham, S. Kendler, X. Wang, F.A. O'Neill, D. Walsh, and K.S. Kendler, *Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families.* Am J Med Genet, 2004. **129B**(1): p. 23-6.

146. Morris, D.W., A. Rodgers, K.A. McGhee, S. Schwaiger, P. Scully, J. Quinn, D. Meagher, J.L. Waddington, M. Gill, and A.P. Corvin, *Confirming RGS4 as a susceptibility gene for schizophrenia.* Am J Med Genet, 2004. **125B**(1): p. 50-3.

147. Williams, N.M., A. Preece, G. Spurlock, N. Norton, H.J. Williams, R.G. McCreadie, P. Buckland, V. Sharkey, K.V. Chowdari, S. Zammit, et al., *Support for RGS4 as a susceptibility gene for schizophrenia.* Biol Psychiatry, 2004. **55**(2): p. 192-5.

148. Stopkova, P., T. Saito, C.S. Fann, D.F. Papolos, J. Vevera, I. Paclt, I. Zukov, R. Stryjer, R.D. Strous, and H.M. Lachman, *Polymorphism Screening of PIP5K2A: A Candidate Gene for Chromosome 10p-Linked Psychiatric Disorders.* Am J Med Genet, 2003. **123B**(1): p. 50-8.

149. Sewekow, C.A., S.G. Schwab, M. Knapp, J. Hallmayer, G.N. Eckstein, S. Gabel, M. Albus, M. Borrmann-Hassenbach, B. Lerer, W. Maier, et al., *Association of SNPs with schizophrenia on chromosome 10p, a region with previously detected linkage (published abstract).* American Journal of Medical Genetics, 2003. **122B**(1): p. P244.

150. Bakker, S.C., M.L. Hoogendoorn, J.P. Selten, W. Verduijn, P.L. Pearson, R.J. Sinke, and R.S. Kahn, *Neuregulin 1: genetic support for schizophrenia subtypes.* Mol Psychiatry, 2004. **9**(12): p. 1061-3.

151. Benson, G., *Tandem repeats finder: a program to analyze DNA sequences.* Nucleic Acids Res, 1999. **27**(2): p. 573-80.

152. Rozen, S. and H.J. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers*. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, 2000: p. pp 365-386.

153. Mouritzen, P., A.T. Nielsen, H.M. Pfundheller, Y. Choleva, L. Kongsbak, and S. Moller, *Single nucleotide polymorphism genotyping using locked nucleic acid (LNA)*. Expert Rev Mol Diagn, 2003. **3**(1): p. 27-38.

154. Latorra, D., K. Campbell, A. Wolter, and J.M. Hurley, *Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers*. Hum Mutat, 2003. **22**(1): p. 79-85.

155. Abecasis, G.R. and W.O. Cookson, *GOLD--graphical overview of linkage disequilibrium*. Bioinformatics, 2000. **16**(2): p. 182-3.

156. Barrett, J.C., B. Fry, J. Maller, and M.J. Daly, *Haploview: analysis and visualization of LD and haplotype maps*. Bioinformatics, 2004. **21**(2): p. 263-265.

157. Beraki, S., F. Aronsson, H. Karlsson, S.O. Ogren, and K. Kristensson, *Influenza A virus infection causes alterations in expression of synaptic regulatory genes combined with changes in cognitive and emotional behaviors in mice*. Mol Psychiatry, 2004. **10**(3): p. 299-308.

158. Levy, F., *Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study*. Journal of the American Academy of Child & Adolescent Psychiatry, 1997. **36**(6): p. 737-44.

159. Krause, K.H., S.H. Dresel, J. Krause, H.F. Kung, and K. Tatsch, *Increased striatal dopamine transporter in adult patients with attention deficit hyperactivity disorder: effects of methylphenidate as measured by single photon emission computed tomography*. Neurosci Lett, 2000. **285**(2): p. 107-10.

160. Michelhaugh, S.K., C. Fiskerstrand, E. Lovejoy, M.J. Bannon, and J.P. Quinn, *The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons*. J Neurochem, 2001. **79**(5): p. 1033-8.

161. Mill, J., P. Asherson, C. Browes, U. D'Souza, and I. Craig, *Expression of the dopamine transporter gene is regulated by the 3' UTR VNTR: Evidence from brain and lymphocytes using quantitative RT-PCR*. Am J Med Genet, 2002. **114**(8): p. 975-9.

162. Asghari, V., S. Sanyal, S. Buchwaldt, A. Paterson, V. Jovanovic, and H.H. Van Tol, *Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants*. J Neurochem, 1995. **65**(3): p. 1157-65.

163. Schoots, O. and H.H. Van Tol, *The human dopamine D4 receptor repeat sequences modulate expression*. Pharmacogenomics J, 2003. **3**(6): p. 343-8.

164. Langley, K., L. Marshall, M. Van Den Bree, H. Thomas, M. Owen, M. O'Donovan, and A. Thapar, *Association of the dopamine d(4) receptor gene 7-repeat allele with neuropsychological test performance of children with ADHD*. Am J Psychiatry, 2004. **161**(1): p. 133-8.

165. Bakker, S.C., E.M. van der Meulen, J.K. Buitelaar, L.A. Sandkuijl, D.L. Pauls, A.J. Monsuur, R. van 't Slot, R.B. Minderaa, W.B. Gunning, P.L. Pearson, et al., *A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q*. Am J Hum Genet, 2003. **72**(5): p. 1251-60.

166. Cook, E.H., Jr., M.A. Stein, M.D. Krasowski, N.J. Cox, D.M. Olkon, J.E. Kieffer, and B.L. Leventhal, *Association of attention-deficit disorder and the dopamine transporter gene*. Am J Hum Genet, 1995. **56**(4): p. 993-8.

167. Van Tol, H.H., C.M. Wu, H.C. Guan, K. Ohara, J.R. Bunzow, O. Civelli, J. Kennedy, P. Seeman, H.B. Niznik, and V. Jovanovic, *Multiple dopamine D4 receptor variants in the human population [see comments]*. Nature, 1992. **358**(6382): p. 149-52.

168. Petronis, A., K. O'Hara, C.L. Barr, J.L. Kennedy, and H.H. Van Tol, *(G)n-mononucleotide polymorphism in the human D4 dopamine receptor (DRD4) gene*. Hum Genet, 1994. **93**(6): p. 719.

169. Daly, G., Z. Hawi, M. Fitzgerald, and M. Gill, *Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children*. Mol Psychiatry, 1999. **4**(2): p. 192-6.

170. O'Connell, J.R. and D.E. Weeks, *PedCheck: a program for identification of genotype incompatibilities in linkage analysis*. Am J Hum Genet, 1998. **63**(1): p. 259-66.

171. Hawi, Z., N. Lowe, A. Kirley, F. Gruenhage, M. Nothen, T. Greenwood, J. Kelsoe, M. Fitzgerald, and M. Gill, *Linkage disequilibrium mapping at DAT1, DRD5 and DBH narrows the search for ADHD susceptibility alleles at these loci*. Mol Psychiatry, 2003. **8**(3): p. 299-308.

172. Kustanovich, V., J. Ishii, L. Crawford, M. Yang, J.J. McGough, J.T. McCracken, S.L. Smalley, and S.F. Nelson, *Transmission disequilibrium testing of dopamine-related candidate gene polymorphisms in ADHD: confirmation of association of ADHD with DRD4 and DRD5*. Mol Psychiatry, 2003.

173. Spencer, T., J. Biederman, T.E. Wilens, and S.V. Faraone, *Adults with attention-deficit/hyperactivity disorder: a controversial diagnosis*. J Clin Psychiatry, 1998. **59 Suppl 7**: p. 59-68.

174. Gaub, M. and C.L. Carlson, *Gender differences in ADHD: a meta-analysis and critical review*. J Am Acad Child Adolesc Psychiatry, 1997. **36**(8): p. 1036-45.

175. Arnold, L.E., *Sex differences in ADHD: conference summary.* J Abnorm Child Psychol, 1996. **24**(5): p. 555-69.

176. Biederman, J., *Attention-deficit/hyperactivity disorder: a life-span perspective.* J Clin Psychiatry, 1998. **59 Suppl 7**: p. 4-16.

177. Biederman, J., S.V. Faraone, K. Keenan, J. Benjamin, B. Krifcher, C. Moore, S. Sprich-Buckminster, K. Ugaglia, M.S. Jellinek, R. Steingard, et al., *Further evidence for family-genetic risk factors in attention deficit hyperactivity disorder. Patterns of comorbidity in probands and relatives psychiatrically and pediatrically referred samples.* Arch Gen Psychiatry, 1992. **49**(9): p. 728-38.

178. Faraone, S.V., J. Biederman, E. Mick, A.E. Doyle, T. Wilens, T. Spencer, E. Frazier, and K. Mullen, *A family study of psychiatric comorbidity in girls and boys with attention-deficit/hyperactivity disorder.* Biol Psychiatry, 2001. **50**(8): p. 586-92.

179. Barr, C.L., C. Xu, J. Kroft, Y. Feng, K. Wigg, G. Zai, R. Tannock, R. Schachar, M. Malone, W. Roberts, et al., *Haplotype study of three polymorphisms at the dopamine transporter locus confirm linkage to attention-deficit/hyperactivity disorder.* Biological Psychiatry, 2001. **49**(4): p. 333-9.

180. Gill, M., G. Daly, S. Heron, Z. Hawi, and M. Fitzgerald, *Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism.* Molecular Psychiatry, 1997. **2**(4): p. 311-3.

181. Waldman, I.D., D.C. Rowe, A. Abramowitz, S.T. Kozel, J.H. Mohr, S.L. Sherman, H.H. Cleveland, M.L. Sanders, J.M. Gard, and C. Stever, *Association and linkage of the dopamine transporter gene and attention- deficit hyperactivity disorder in children: heterogeneity owing to diagnostic subtype and severity.* Am J Hum Genet, 1998. **63**(6): p. 1767-76.

182. Palmer, C.G., J.N. Bailey, C. Ramsey, D. Cantwell, J.S. Sinsheimer, M. Del'Homme, J. McGough, J.A. Woodward, R. Asarnow, J. Asarnow, et al., *No evidence of linkage or linkage disequilibrium between DAT1 and attention deficit hyperactivity disorder in a large sample.* Psychiatr Genet, 1999. **9**(3): p. 157-60.

183. Jorm, A.F., M. Prior, A. Sanson, D. Smart, Y. Zhang, and S. Easteal, *Association of a polymorphism of the dopamine transporter gene with externalizing behavior problems and associated temperament traits: a longitudinal study from infancy to the mid-teens.* Am J Med Genet, 2001. **105**(4): p. 346-50.

184. Holmes, J., A. Payton, J.H. Barrett, T. Hever, H. Fitzpatrick, A.L. Trumper, R. Harrington, P. McGuffin, M. Owen, W. Ollier, et al., *A family-based and case-control association study of the dopamine D4 receptor gene and dopamine transporter gene in attention deficit hyperactivity disorder.* Molecular Psychiatry, 2000. **5**(5): p. 523-30.

185. Kotler, M., I. Manor, Y. Sever, J. Eisenberg, H. Cohen, R.P. Ebstein, and S. Tyano, *Failure to replicate an excess of the long dopamine D4 exon III repeat polymorphism in ADHD in a family-based study.* American Journal of Medical Genetics, 2000. **96**(3): p. 278-81.

186. Roman, T., M. Schmitz, G. Polanczyk, M. Eizirik, L.A. Rohde, and M.H. Hutz, *Attention-deficit hyperactivity disorder: a study of association with both the dopamine transporter gene and the dopamine D4 receptor gene.* American Journal of Medical Genetics, 2001. **105**(5): p. 471-8.

187. Schmidt, L.A., N.A. Fox, K. Perez-Edgar, S. Hu, and D.H. Hamer, *Association of DRD4 with attention problems in normal childhood development.* Psychiatr Genet, 2001. **11**(1): p. 25-9.

188. Gainetdinov, R.R., W.C. Wetsel, S.R. Jones, E.D. Levin, M. Jaber, and M.G. Caron, *Role of serotonin in the paradoxical calming effect of psychostimulants on hyperactivity. [see comments].* Science, 1999. **283**(5400): p. 397-401.

189. Fisher, S.E., C. Francks, J.T. McCracken, J.J. McGough, A.J. Marlow, I.L. MacPhie, D.F. Newbury, L.R. Crawford, C.G. Palmer, J.A. Woodward, et al., *A genomewide scan for loci involved in attention-deficit/hyperactivity disorder.* Am J Hum Genet, 2002. **70**(5): p. 1183-96.

190. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results [see comments].* Nat Genet, 1995. **11**(3): p. 241-7.

191. Smalley, S.L., V. Kustanovich, S.L. Minassian, J.L. Stone, M.N. Ogdie, J.J. McGough, J.T. McCracken, I.L. MacPhie, C. Francks, S.E. Fisher, et al., *Genetic linkage of attention-deficit/hyperactivity disorder on chromosome 16p13, in a region implicated in autism.* Am J Hum Genet, 2002. **71**(4): p. 959-63.

192. Bailey, A., A. Le Couteur, I. Gottesman, P. Bolton, E. Simonoff, E. Yuzda, and M. Rutter, *Autism as a strongly genetic disorder: evidence from a British twin study.* Psychol Med, 1995. **25**(1): p. 63-77.

193. Luteijn, E.F., M. Serra, S. Jackson, M.P. Steenhuis, M. Althaus, F. Volkmar, and R. Minderaa, *How unspecified are disorders of children with a pervasive developmental disorder not otherwise specified? A study of social problems in children with PDD-NOS and ADHD.* Eur Child Adolesc Psychiatry, 2000. **9**(3): p. 168-79.

194. Aman, M.G. and K.S. Langworthy, *Pharmacotherapy for hyperactivity in children with autism and other pervasive developmental disorders.* J Autism Dev Disord, 2000. **30**(5): p. 451-9.

195. Jaselskis, C.A., E.H. Cook, Jr., K.E. Fletcher, and B.L. Leventhal, *Clonidine treatment of hyperactive and impulsive children with autistic disorder.* J Clin Psychopharmacol, 1992. **12**(5): p. 322-7.

196. Noterdaeme, M., H. Amorosa, K. Mildenberger, S. Sitter, and F. Minow, *Evaluation of attention problems in children with autism and children with a specific language disorder.* Eur Child Adolesc Psychiatry, 2001. **10**(1): p. 58-66.

197. Wechsler, D., *WISC-R manual. Wechsler Intelligence Scale for Children-Revised Manual.* 1974, New York: Psychological Corporation.

198. Wechsler, D., *Wechsler Preschool and Primary Scale of Intelligence.* 1967, New York: Psychological Corporation.

199. Abecasis, G.R., S.S. Cherny, W.O. Cookson, and L.R. Cardon, *GRR: graphical representation of relationship errors.* Bioinformatics, 2001. **17**(8): p. 742-3.

200. Shaffer, D., P. Fisher, C.P. Lucas, M.K. Dulcan, and M.E. Schwab-Stone, *NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses.* J Am Acad Child Adolesc Psychiatry, 2000. **39**(1): p. 28-38.

201. Achenbach, T.M. and T.M. Ruffle, *The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies.* Pediatr Rev, 2000. **21**(8): p. 265-71.

202. Goyette, C.H., C.K. Conners, and R.F. Ulrich, *Normative data on revised Conners Parent and Teacher Rating Scales.* J Abnorm Child Psychol, 1978. **6**(2): p. 221-36.

203. Leckman, J.F., D. Sholomskas, W.D. Thompson, A. Belanger, and M.M. Weissman, *Best estimate of lifetime psychiatric diagnosis: a methodological study.* Arch Gen Psychiatry, 1982. **39**(8): p. 879-83.

204. Kong, A., D.F. Gudbjartsson, J. Sainz, G.M. Jonsdottir, S.A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, et al., *A high-resolution recombination map of the human genome.* Nat Genet, 2002. **31**(3): p. 241-7.

205. Kruglyak, L. and E.S. Lander, *Complete multipoint sib-pair analysis of qualitative and quantitative traits.* Am J Hum Genet, 1995. **57**(2): p. 439-54.

206. Holmans, P., *Asymptotic properties of affected-sib-pair linkage analysis.* Am J Hum Genet, 1993. **52**(2): p. 362-74.

207. Cordell, H.J., Y. Kawaguchi, J.A. Todd, and M. Farrall, *An extension of the Maximum Lod Score method to X-linked loci.* Ann Hum Genet, 1995. **59 (Pt 4)**: p. 435-49.

208. Daly, M.J. and E.S. Lander, *The importance of being independent: sib pair analysis in diabetes.* Nat Genet, 1996. **14**(2): p. 131-2.

209. Altmuller, J., L.J. Palmer, G. Fischer, H. Scherb, and M. Wjst, *Genomewide scans of complex human diseases: true linkage is hard to find.* Am J Hum Genet, 2001. **69**(5): p. 936-50.

210. Cohen, N.J., D.D. Vallance, M. Barwick, N. Im, R. Menna, N.B. Horodezky, and L. Isaacson, *The interface between ADHD and language impairment: an examination of language, achievement, and cognitive processing.* J Child Psychol Psychiatry, 2000. **41**(3): p. 353-62.

211. Kovac, I., B. Garabedian, C. Du Souich, and R.M. Palmour, *Attention deficity/hyperactivity in SLI children increases risk of speech/language disorders in first-degree relatives: a preliminary report.* J Commun Disord, 2001. **34**(4): p. 339-54.

212. Willcutt, E.G., B.F. Pennington, S.D. Smith, L.R. Cardon, J. Gayan, V.S. Knopik, R.K. Olson, and J.C. DeFries, *Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder.* Am J Med Genet, 2002. **114**(3): p. 260-8.

213. Cook, E.H., Jr., V. Lindgren, B.L. Leventhal, R. Courchesne, A. Lincoln, C. Shulman, C. Lord, and E. Courchesne, *Autism or atypical autism in maternally but not paternally derived proximal 15q duplication.* Am J Hum Genet, 1997. **60**(4): p. 928-34.

214. Schroer, R.J., M.C. Phelan, R.C. Michaelis, E.C. Crawford, S.A. Skinner, M. Cuccaro, R.J. Simensen, J. Bishop, C. Skinner, D. Fender, et al., *Autism and maternally derived aberrations of chromosome 15q.* Am J Med Genet, 1998. **76**(4): p. 327-36.

215. Wolpert, C.M., M.M. Menold, M.P. Bass, M.B. Qumsiyeh, S.L. Donnelly, S.A. Ravan, J.M. Vance, J.R. Gilbert, R.K. Abramson, H.H. Wright, et al., *Three probands with autistic disorder and isodicentric chromosome 15.* Am J Med Genet, 2000. **96**(3): p. 365-72.

216. Collaborative Linkage Study of Autism, *An autosomal genomic screen for autism.* Am J Med Genet, 2001. **105**(8): p. 609-15.

217. Gutknecht, L., *Full-genome scans with autistic disorder: a review.* Behav Genet, 2001. **31**(1): p. 113-23.

218. Philippe, A., M. Martinez, M. Guilloud-Bataille, C. Gillberg, M. Rastam, E. Sponheim, M. Coleman, M. Zappella, H. Aschauer, L. Van Maldergem, et al., *Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study.* Hum Mol Genet, 1999. **8**(5): p. 805-12.

219. Risch, N., D. Spiker, L. Lotspeich, N. Nouri, D. Hinds, J. Hallmayer, L. Kalaydjieva, P. McCague, S. Dimiceli, T. Pitts, et al., *A genomic screen of autism: evidence for a multilocus etiology.* Am J Hum Genet, 1999. **65**(2): p. 493-507.

220. Shao, Y., M.L. Cuccaro, E.R. Hauser, K.L. Raiford, M.M. Menold, C.M. Wolpert, S.A. Ravan, L. Elston, K. Decena, S.L. Donnelly, et al., *Fine Mapping of Autistic Disorder to Chromosome 15q11-q13 by Use of Phenotypic Subtypes.* Am J Hum Genet, 2003. **72**(3): p. 539-548.

221. Grigorenko, E.L., F.B. Wood, M.S. Meyer, L.A. Hart, W.C. Speed, A. Shuster, and D.L. Pauls, *Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15.* Am J Hum Genet, 1997. **60**(1): p. 27-39.

222. Nothen, M.M., G. Schulte-Korne, T. Grimm, S. Cichon, I.R. Vogt, B. Muller-Myhsok, P. Propping, and H. Remschmidt, *Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15.* Eur Child Adolesc Psychiatry, 1999. **8 Suppl 3**: p. 56-9.

223. Morris, D.W., L. Robinson, D. Turic, M. Duke, V. Webb, C. Milham, E. Hopkin, K. Pound, S. Fernando, M. Easton, et al., *Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q.* Hum Mol Genet, 2000. **9**(5): p. 843-8.

224. Barr, C.L., Y. Feng, B. Anderson, J. Crosbie, W. Roberts, M. Malone, A. Ickowicz, R. Schachar, R. Tannock, M. Lovett, et al., *Significant evidence for linkage of attention-deficit hyperactivity disorder to the chromosome 15q region.* American Journal of Medical Genetics, 2002. **114**: p. published abstract P154.

225. Ernst, M., A.J. Zametkin, J.A. Matochik, D. Pascualvaca, P.H. Jons, and R.M. Cohen, *High midbrain [18F]DOPA accumulation in children with attention deficit hyperactivity disorder.* Am J Psychiatry, 1999. **156**(8): p. 1209-15.

226. Hawi, Z., D. Foley, A. Kirley, M. McCarron, M. Fitzgerald, and M. Gill, *Dopa decarboxylase gene polymorphisms and attention deficit hyperactivity disorder (ADHD): no evidence for association in the Irish population.* Mol Psychiatry, 2001. **6**(4): p. 420-4.

227. Roman, T., M. Schmitz, G.V. Polanczyk, M. Eizirik, L.A. Rohde, and M.H. Hutz, *Further evidence for the association between attention-deficit/hyperactivity disorder and the dopamine-beta-hydroxylase gene.* Am J Med Genet, 2002. **114**(2): p. 154-8.

228. Wigg, K., G. Zai, R. Schachar, R. Tannock, W. Roberts, M. Malone, J.L. Kennedy, and C.L. Barr, *Attention deficit hyperactivity disorder and the gene for dopamine Beta-hydroxylase.* Am J Psychiatry, 2002. **159**(6): p. 1046-8.

229. Rogeness, G.A., J.W. Maas, M.A. Javors, C.A. Macedo, C. Fischer, and W.R. Harris, *Attention deficit disorder symptoms and urine catecholamines.* Psychiatry Res, 1989. **27**(3): p. 241-51.

230. O'Malley, K.L., S. Harmon, M. Moffat, A. Uhland-Smith, and S. Wong, *The human aromatic L-amino acid decarboxylase gene can be alternatively spliced to generate unique protein isoforms.* J Neurochem, 1995. **65**(6): p. 2409-16.

231. Ichinose, H., C. Sumi-Ichinose, T. Ohye, Y. Hagino, K. Fujita, and T. Nagatsu, *Tissue-specific alternative splicing of the first exon generates two types of mRNAs in human aromatic L-amino acid decarboxylase.* Biochemistry, 1992. **31**(46): p. 11546-50.

232. Sumi-Ichinose, C., S. Hasegawa, H. Ichinose, H. Sawada, K. Kobayashi, M. Sakai, T. Fujii, H. Nomura, T. Nomura, I. Nagatsu, et al., *Analysis of the alternative promoters that regulate tissue-specific expression of human aromatic L-amino acid decarboxylase.* J Neurochem, 1995. **64**(2): p. 514-24.

233. Aguanno, A., R. Afar, and V.R. Albert, *Tissue-specific expression of the nonneuronal promoter of the aromatic L-amino acid decarboxylase gene is regulated by hepatocyte nuclear factor 1.* J Biol Chem, 1996. **271**(8): p. 4528-38.

234. Chireux, M., J.F. Raynal, A. Le Van Thai, H. Cadas, C. Bernard, I. Martinou, J.C. Martinou, and M.J. Weber, *Multiple promoters of human choline acetyltransferase and aromatic L-amino acid decarboxylase genes.* J Physiol Paris, 1994. **88**(4): p. 215-27.

235. Le Van Thai, A., E. Coste, J.M. Allen, R.D. Palmiter, and M.J. Weber, *Identification of a neuron-specific promoter of human aromatic L-amino acid decarboxylase gene.* Brain Res Mol Brain Res, 1993. **17**(3-4): p. 227-38.

236. Hitchins, M.P., L. Bentley, D. Monk, C. Beechey, J. Peters, G. Kelsey, F. Ishino, M.A. Preece, P. Stanier, and G.E. Moore, *DDC and COBL, flanking the imprinted GRB10 gene on 7p12, are biallelically expressed.* Mamm Genome, 2002. **13**(12): p. 686-91.

237. Miller, M.J. and B.Z. Yuan, *Semiautomated resolution of overlapping stutter patterns in genomic microsatellite analysis.* Anal Biochem, 1997. **251**(1): p. 50-6.

238. Murray, V., C. Monchawin, and P.R. England, *The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR.* Nucleic Acids Res, 1993. **21**(10): p. 2395-8.

239. Hauge, X.Y. and M. Litt, *A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR.* Hum Mol Genet, 1993. **2**(4): p. 411-5.

240. Clarke, L.A., C.S. Rebelo, J. Goncalves, M.G. Boavida, and P. Jordan, *PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences.* Mol Pathol, 2001. **54**(5): p. 351-3.

241. Perlin, M.W., M.B. Burks, R.C. Hoop, and E.P. Hoffman, *Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy.* Am J Hum Genet, 1994. **55**(4): p. 777-87.

242. Gill, P., J. Whitaker, C. Flaxman, N. Brown, and J. Buckleton, *An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA.* Forensic Sci Int, 2000. **112**(1): p. 17-40.

243. Bovo, D., M. Rugge, and Y.H. Shiao, *Origin of spurious multiple bands in the amplification of microsatellite sequences.* Mol Pathol, 1999. **52**(1): p. 50-1.

244. Lipkin, E., M.O. Mosig, A. Darvasi, E. Ezra, A. Shalom, A. Friedmann, and M. Soller, *Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage.* Genetics, 1998. **149**(3): p. 1557-67.

245. Lazaruk, K., J. Wallin, C. Holt, T. Nguyen, and P.S. Walsh, *Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing.* Forensic Sci Int, 2001. **119**(1): p. 1-10.

246. Walsh, P.S., N.J. Fildes, and R. Reynolds, *Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA.* Nucleic Acids Res, 1996. **24**(14): p. 2807-12.

247. Gill, P., C.P. Kimpton, A. Urquhart, N. Oldroyd, E.S. Millican, S.K. Watson, and T.J. Downes, *Automated short tandem repeat (STR) analysis in forensic casework--a strategy for the future.* Electrophoresis, 1995. **16**(9): p. 1543-52.

248. Shinde, D., Y. Lai, F. Sun, and N. Arnheim, *Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites.* Nucleic Acids Res, 2003. **31**(3): p. 974-80.

249. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu Rev Biochem, 2000. **69**: p. 497-529.

250. van Tilburg, J.H., L.A. Sandkuijl, E. Strengman, H. van Someren, C.A. Rigters-Aris, P.L. Pearson, T.W. van Haeften, and C. Wijmenga, *A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel Loci on 2q12 and 19q13.* J Clin Endocrinol Metab, 2003. **88**(5): p. 2223-30.

251. Neilan, B.A., A.N. Wilton, and D. Jacobs, *A universal procedure for primer labelling of amplicons.* Nucleic Acids Res, 1997. **25**(14): p. 2938-9.

252. Breen, G., P. Sham, T. Li, D. Shaw, D.A. Collier, and D. St Clair, *Accuracy and sensitivity of DNA pooling with microsatellite repeats using capillary electrophoresis.* Mol Cell Probes, 1999. **13**(5): p. 359-65.

253. Lederer, T., S. Seidl, B. Graham, and P. Betz, *A new pentaplex PCR system for forensic casework analysis.* Int J Legal Med, 2000. **114**(1-2): p. 87-92.

254. Chakraborty, R., M. Kimmel, D.N. Stivers, L.J. Davison, and R. Deka, *Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci.* Proc Natl Acad Sci U S A, 1997. **94**(3): p. 1041-6.

255. Perinchery, G., D. Nojima, R. Goharderakhshan, Y. Tanaka, J. Alonzo, and R. Dahiya, *Microsatellite instability of dinucleotide tandem repeat sequences is higher than trinucleotide, tetranucleotide and pentanucleotide repeat sequences in prostate cancer.* Int J Oncol, 2000. **16**(6): p. 1203-9.

256. Lee, J.S., M.G. Hanford, J.L. Genova, and R.A. Farber, *Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells.* Hum Mol Genet, 1999. **8**(13): p. 2567-72.

257. Weber, J.L. and C. Wong, *Mutation of human short tandem repeats.* Hum Mol Genet, 1993. **2**(8): p. 1123-8.

258. Eckert, K.A. and G. Yan, *Mutational analyses of dinucleotide and tetranucleotide microsatellites in Escherichia coli: influence of sequence on expansion mutagenesis.* Nucleic Acids Res, 2000. **28**(14): p. 2831-8.

259. Wu, M.J., L.W. Chow, and M. Hsieh, *Amplification of GAA/TTC triplet repeat in vitro: preferential expansion of (TTC)n strand.* Biochim Biophys Acta, 1998. **1407**(2): p. 155-62.

260. Yamada, N.A., G.A. Smith, A. Castro, C.N. Roques, J.C. Boyer, and R.A. Farber, *Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells.* Mutat Res, 2002. **499**(2): p. 213-25.

261. Schug, M.D., C.M. Hutter, K.A. Wetterstrand, M.S. Gaudette, T.F. Mackay, and C.F. Aquadro, *The mutation rates of di-, tri- and tetranucleotide repeats in Drosophila melanogaster.* Mol Biol Evol, 1998. **15**(12): p. 1751-60.

262. Ellegren, H., *Heterogeneous mutation processes in human microsatellite DNA sequences.* Nat Genet, 2000. **24**(4): p. 400-2.

263. Sibly, R.M., J.C. Whittaker, and M. Talbot, *A maximum-likelihood approach to fitting equilibrium models of microsatellite evolution.* Mol Biol Evol, 2001. **18**(3): p. 413-7.

264. Kroutil, L.C., K. Register, K. Bebenek, and T.A. Kunkel, *Exonucleolytic proofreading during replication of repetitive DNA.* Biochemistry, 1996. **35**(3): p. 1046-53.

265. Rose, O. and D. Falush, *A threshold size for microsatellite expansion.* Mol Biol Evol, 1998. **15**(5): p. 613-5.

266. Choudhry, S., M. Mukerji, A.K. Srivastava, S. Jain, and S.K. Brahmachari, *CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms.* Hum Mol Genet, 2001. **10**(21): p. 2437-46.

267. Bacon, A.L., S.M. Farrington, and M.G. Dunlop, *Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells.* Hum Mol Genet, 2000. **9**(18): p. 2707-13.

268. Jin, L., C. Macaubas, J. Hallmayer, A. Kimura, and E. Mignot, *Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence.* Proc Natl Acad Sci U S A, 1996. **93**(26): p. 15285-8.

269. Bachtrog, D., M. Agis, M. Imhof, and C. Schlotterer, *Microsatellite variability differs between dinucleotide repeat motifs-evidence from Drosophila melanogaster.* Mol Biol Evol, 2000. **17**(9): p. 1277-85.

270. Hile, S.E., G. Yan, and K.A. Eckert, *Somatic mutation rates and specificities at TC/AG and GT/CA microsatellite sequences in nontumorigenic human lymphoblastoid cells.* Cancer Res, 2000. **60**(6): p. 1698-703.

271. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases [see comments].* Science, 1996. **273**(5281): p. 1516-7.

272. Dunning, A.M., F. Durocher, C.S. Healey, M.D. Teare, S.E. McBride, F. Carlomagno, C.F. Xu, E. Dawson, S. Rhodes, S. Ueda, et al., *The extent of linkage disequilibrium in four populations with distinct demographic histories [In Process Citation].* Am J Hum Genet, 2000. **67**(6): p. 1544-54.

273. Abecasis, G.R., E. Noguchi, A. Heinzmann, J.A. Traherne, S. Bhattacharyya, N.I. Leaves, G.G. Anderson, Y. Zhang, N.J. Lench, A. Carey, et al., *Extent and Distribution of Linkage Disequilibrium in Three Genomic Regions.* Am J Hum Genet, 2001. **68**(1): p. 191-197.

274. Innan, H., B. Padhukasahasram, and M. Nordborg, *The pattern of polymorphism on human chromosome 21.* Genome Res, 2003. **13**(6): p. 1158-68.

275. Salisbury, B.A., M. Pungliya, J.Y. Choi, R. Jiang, X.J. Sun, and J.C. Stephens, *SNP and haplotype variation in the human genome.* Mutat Res, 2003. **526**(1-2): p. 53-61.

276. LeDuc, C., P. Miller, J. Lichter, and P. Parry, *Batched analysis of genotypes.* PCR Methods Appl, 1995. **4**(6): p. 331-6.

277. Sham, P.C., J.H. Zhao, and D. Curtis, *The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations.* Ann Hum Genet, 2000. **64**(Pt 2): p. 161-9.

278. Brownstein, M.J., J.D. Carpten, and S. J.R., *Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping.* BioTechniques, 1996. **20**: p. 1004-1010.

279. Press, W.H., S.A. Teukolsky, W.T. Vettering, and B.H. Flannery, *Numerical recipes in C- the art of scientific computing.* 2nd ed. 1992, Cambridge: Cambridge University Press.

280. Gibbs, R.A., P.N. Nguyen, and C.T. Caskey, *Detection of single DNA base differences by competitive oligonucleotide priming.* Nucleic Acids Res, 1989. **17**(7): p. 2437-48.

281. Myakishev, M.V., Y. Khripin, S. Hu, and D.H. Hamer, *High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers.* Genome Res, 2001. **11**(1): p. 163-9.

282. McClay, J.L., K. Sugden, H.G. Koch, S. Higuchi, and I.W. Craig, *High-throughput single-nucleotide polymorphism genotyping by fluorescent competitive allele-specific polymerase chain reaction (SNiP-Tag).* Anal Biochem, 2002. **301**(2): p. 200-6.

283. Ye, S., S. Dhillon, X. Ke, A.R. Collins, and I.N. Day, *An efficient procedure for genotyping single nucleotide polymorphisms.* Nucleic Acids Res, 2001. **29**(17): p. E88-8.

284. Shimizu, M., N. Kosaka, T. Shimada, T. Nagahata, H. Iwasaki, H. Nagai, T. Shiba, and M. Emi, *Universal fluorescent labeling (UFL) method for automated microsatellite analysis.* DNA Res, 2002. **9**(5): p. 173-8.

285. Ishii, M., A. Inanobe, and Y. Kurachi, *PIP3 inhibition of RGS protein and its reversal by Ca2+/calmodulin mediate voltage-dependent control of the G protein cycle in a cardiac K+ channel.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4325-30.

286. Evans, J.D., R.K. Heaton, J.S. Paulsen, L.A. McAdams, S.C. Heaton, and D.V. Jeste, *Schizoaffective disorder: a form of schizophrenia or affective disorder?* J Clin Psychiatry, 1999. **60**(12): p. 874-82.

287. Kirk, K.M. and L.R. Cardon, *The impact of genotyping error on haplotype reconstruction and frequency estimation.* Eur J Hum Genet, 2002. **10**(10): p. 616-22.

288. Kim, J.W., J. Fagerness, L. Arbeitman, A. Doyle, C. Petty, J. Biederman, S.V. Faraone, and P. Sklar, *Fine mapping of chromosomal regions implicated in ADHD (published abstract).* American Journal of Medical Genetics, 2004. **130B(1)**: p. 95.

289. Ogdie, M.N., S.E. Fisher, M. Yang, J. Ishii, C. Francks, S.K. Loo, R.M. Cantor, J.T. McCracken, J.J. McGough, S.L. Smalley, et al., *Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11.* Am J Hum Genet, 2004. **75**(4): p. 661-8.

290. Arcos-Burgos, M., F.X. Castellanos, D. Pineda, F. Lopera, J. David Palacio, L. Guillermo Palacio, J.L. Rapoport, K. Berg, J.E. Bailey-Wilson, and M. Muenke, *Attention-deficit/hyperactivity disorder in a population isolate: linkage to Loci at 4q13.2, 5q33.3, 11q22, and 17p11.* Am J Hum Genet, 2004. **75**(6): p. 998-1014.

291. Ohashi, J. and K. Tokunaga, *Power of genome-wide linkage disequilibrium testing by using microsatellite markers.* J Hum Genet, 2003. **48**(9): p. 487-91.

292. Subramanian, S., R.K. Mishra, and L. Singh, *Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions.* Genome Biol, 2003. **4**(2): p. R13.

293. Tanaka, G., I. Matsushita, J. Ohashi, N. Tsuchiya, S. Ikushima, M. Oritsu, M. Hijikata, T. Nagata, K. Yamamoto, K. Tokunaga, et al., *Evaluation of microsatellite markers in association*

studies: a search for an immune-related susceptibility gene in sarcoidosis. Immunogenetics, 2005.

294. Kwok, P.Y., *High-throughput genotyping assay approaches.* Pharmacogenomics, 2000. **1**(1): p. 95-100.

295. Dean, M., *Approaches to identify genes for complex human diseases: lessons from Mendelian disorders.* Hum Mutat, 2003. **22**(4): p. 261-74.

296. Hinds, D.A., R.P. Stokowski, N. Patil, K. Konvicka, D. Kershenobich, D.R. Cox, and D.G. Ballinger, *Matching strategies for genetic association studies in structured populations.* Am J Hum Genet, 2004. **74**(2): p. 317-25.

297. Terwilliger, J.D., F. Haghighi, T.S. Hiekkalinna, and H.H. Goring, *A bias-ed assessment of the use of SNPs in human complex traits.* Curr Opin Genet Dev, 2002. **12**(6): p. 726-34.

298. John, S., N. Shephard, G. Liu, E. Zeggini, M. Cao, W. Chen, N. Vasavda, T. Mills, A. Barton, A. Hinks, et al., *Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites.* Am J Hum Genet, 2004. **75**(1): p. 54-64.

299. Schaid, D.J., J.C. Guenther, G.B. Christensen, S. Hebbring, C. Rosenow, C.A. Hilker, S.K. McDonnell, J.M. Cunningham, S.L. Slager, M.L. Blute, et al., *Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility Loci.* Am J Hum Genet, 2004. **75**(6): p. 948-65.

300. Middleton, F.A., M.T. Pato, K.L. Gentile, C.P. Morley, X. Zhao, A.F. Eisener, A. Brown, T.L. Petryshen, A.N. Kirby, H. Medeiros, et al., *Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22.* Am J Hum Genet, 2004. **74**(5): p. 886-97.

301. Garner, C. and M. Slatkin, *On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci.* Genet Epidemiol, 2003. **24**(1): p. 57-67.

302. Zhang, K., P. Calabrese, M. Nordborg, and F. Sun, *Haplotype block structure and its applications to association studies: power and study designs.* Am J Hum Genet, 2002. **71**(6): p. 1386-94.

303. Wang, N., J.M. Akey, K. Zhang, R. Chakraborty, and L. Jin, *Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation.* Am J Hum Genet, 2002. **71**(5): p. 1227-34.

304. Wall, J.D. and J.K. Pritchard, *Assessing the performance of the haplotype block model of linkage disequilibrium.* Am J Hum Genet, 2003. **73**(3): p. 502-15.

305. Service, S.K., L.A. Sandkuijl, and N.B. Freimer, *Cost-effective designs for linkage disequilibrium mapping of complex traits.* Am J Hum Genet, 2003. **72**(5): p. 1213-20.

306. Wadhwa, R., S.C. Kaul, M. Miyagishi, and K. Taira, *Know-how of RNA interference and its applications in research and therapy.* Mutat Res, 2004. **567**(1): p. 71-84.

307. Mello, C.C. and D. Conte, Jr., *Revealing the world of RNA interference.* Nature, 2004. **431**(7006): p. 338-42.

308. Hannon, G.J. and J.J. Rossi, *Unlocking the potential of the human genome with RNA interference.* Nature, 2004. **431**(7006): p. 371-8.

309. Gladkevich, A., H.F. Kauffman, and J. Korf, *Lymphocytes as a neural probe: potential for studying psychiatric disorders.* Prog Neuropsychopharmacol Biol Psychiatry, 2004. **28**(3): p. 559-76.