From DEPARTMENT OF CELLULAR AND MOLECULAR BIOLOGY
Progamme for Genomics and Bioinformatics
Karolinska Institute, Stockholm, Sweden

COMPARATIVE SEQUENCING OF CANDIDATE GENES IN COMPLEX DISEASE

Shane McCarthy

Stockholm 2006

All previously published papers were reproduced with permission from the publishers
Published and printed by:
Larserics Digital Print AB,
Sundbyberg, Stockholm, Sweden

ISBN 91-7140-663-8

To you who shared the path,
who held me up,
who watched over me

Abstract

Complex multi-factorial diseases such as cardiovascular, metabolic, neurological and respiratory disorders affect a great number of people across the world. In the post-Human Genome Sequencing era, genome wide association studies are increasingly viable alternatives to linkage approaches in locating disease genes. In addition, positional and hypothesis-driven candidate genes can be assessed for their role in susceptibility to sporadic common diseases. The efficiency and success of these approaches depend on knowledge of DNA variation and linkage disequilibrium. This thesis describes the comparative sequencing of regions across the candidate genes *HTR2C*, *MAOA*, *MAOB*, *IDE*, *KIF11* and *HHEX* to illustrate the importance of understanding the fine scale nucleotide and LD distribution for improvements in association study design with Obesity, Depression and Alzheimers disease.

In *HTR2C*, recombination between the commonly used nsSNP marker, *Cys*23*Ser*, and the promoter was observed (Paper I). Furthermore, nucleotide and haplotype analysis showed that gene conversion in the promoter contributed to the complexity of LD. If the functional promoter polymorphisms act in the susceptibility to serotonergic-related phenotypes, this work suggests that the unknown structure of LD across *HTR2C* could have been an issue in previous association studies using *Cys*23*Ser*. Support for *HTR2C* promoter polymorphisms in obesity was provided by the associations of promoter haplotype, *TA13GCG* (P>0.0001) with high body mass index (BMI, 30 $\geq$ kgm^{-2}) and promoter SNP -995*G*>*A*, (P = 0.01) with serum-leptin/%body fat (Paper I). The *HTR2C* promoter haplotype *GGCC* effects were suggestive, but not significant, as having a role in depression (Paper IV).

Sequence variation was scarce in the *MAO* regions studied. This contributes to the hypothesis that these genes are under selective pressure and that much of the variants in public databases for *MAOA*/*B* could be population specific (Paper II). Lack of *MAO* variation was reflected in the poor validation rate of SNPs and LD complex structure in a Swedish twin sample group. While *MAOB* SNPs were found to correlate with depressive state in elderly Swedish Twins, no association was observed with polymorphisms in this gene and trbc-activity (Paper II). Conversely, low trbc-activity was found to associate with *MAOA* SNPs and haplotypes. A potentially additive effect on the risk for depression per *MAO* haplotype was observed. In a larger sample, no significant associations were found with any of the *MAO* SNPs or haplotypes (Paper IV). However, a trend towards departures from Hardy-Weinberg Equilibrium between the genes may suggest that this region warrants further sequencing to identify potential regulatory mechanisms of MAO expression.

A wealth of polymorphisms was found in re-sequencing *IDE*, *KIF11*, *HHEX* and conserved regions within a haplotype block associated with Alzheimers Disease. However, no significant associations between *IDE* and *KIF11* SNPs, and Alzheimers disease were observed.

These works demonstrate the advantages of re-sequencing in providing a better understanding of the various genetic factors influencing studies of polymorphisms with complex diseases. With advances in technology and throughput, sequencing will become instrumental in the location of disease genes and the identification of causative polymorphisms.

List of Publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals

I **Shane McCarthy**, Salim Mottagui-Tabar, Yumi Mizuno, Bengt Sennblad, Johan Hoffstedt, Peter Arner, Claes Wahlestedt, Björn Andersson
Complex *HTR2C* linkage disequilibrium and promoter associations with body mass index and serum leptin
Hum Genet (2005) 117: 545557

II Mårten Jansson*, **Shane McCarthy***, Patrick F. Sullivan, Paul Dickman, Björn Andersson, Lars Oreland, Martin Schalling and Nancy L. Pedersen
MAOA haplotypes associated with thrombocyte-MAO activity
BMC Genetics 2005, 6:46

III Lars Feuk*, **Shane McCarthy***, Björn Andersson, Jonathan A. Prince, and Anthony J. Brookes
Mutation Screening of a Haplotype Block Around the Insulin Degrading Enzyme Gene and Association With Alzheimers Disease
Am J of Med Gen Part B (Neuropsychiatric Genetics) 136B:6971 (2005)

IV **Shane McCarthy**, Annica Dominicus, Måarten Jansson, Björn Andersson, Nancy L. Pedersen
Serotonin Receptor 2C Polymorphisms and Risk for Depression in Swedish Twins
Manuscript

* These authors contributed equally to this work

List of Abbreviations

DNA	Deoxyribonucleic acid
PCR	Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
LD	Linkage disequilibrium
HapMap	Haplotype Map
HGP	Human Genome Project
RNA	Ribonucleic Acid
dNTP	deoxytribonucleotide
ddNTP	dideoxytribonucleotide
ISM	Infinite-sites Model
EPG	Environmental Genome Project
CD/CV	Common Disease/Common Variant
CD/FV	Common Disease/Fixed Variant
CD/RV	Common Disease/Rare Variant
HTR2C	Serotonin Receptor 2C
MAO	Monoamine oxidase
IDE	Insulin-degrading enzyme
KIF11	Kinesin Family Member 11
HHEX	Hematopitically Expressed Homeobox

Contents

Preface

Two groundbreaking events in the field of human genetics mark the beginning and ending of this thesis; the 2001 release of the Human Genome draft sequence [1, 2] and the 2005 announcement of Phase I completion of the Human *HapMap* Project. The goal of the *Human Genome Project* (HGP) was to understand the structure, organization and content of the human genome. The identification of approximately 1.42 million differences (variants) between the sequences used to establish the draft [3], supplied researchers with material of anthropological, evolutionary and medical importance. For these reasons, the objective of the International HapMap Consortium was validating common variants and analyzing their distribution across the genome to determine if they are the same across different populations for a better understanding of human origins and future simplification of disease gene location [4].

Analyzing the genomic and global spread of these variants can teach us about the development of our ancestors, their evolution and migrations out of Africa into Europe, Asia and the Americas an estimated 150,000 years ago. For medical purposes, some of these differences may explain variation in a range of features such as eye color, height, and disease susceptibility to autoimmune, cardiovascular, metabolic, neurological and respiratory diseases.

The concepts behind these breakthroughs are amalgamated into the aims of this thesis. We have comparatively re-sequenced candidate genes for obesity, depression and Alzheimers disease to discover variants between individuals in the population with or without the disease. The relationship between these variants to those already known is compared as a way to understand better the genes role in complex disease. As sequencing technologies improve and their costs reduce re-sequencing of genes will be a very useful tool for discovery and detection of variants functioning in disease susceptibility, mechanisms and treatment.

Chapter 1

An Introduction to DNA, Genes and the Human Genome

Deoxyribonucleic acid (DNA) is composed of a strand of molecules known as deoxyribonucelotides, connected by phosphodiester bonds at the 5' and 3' ends (Figure 1.1). There are four types of deoxyribonucelotides; two purines: Adenosine (*A*) and Guanine (*G*); two pyrimadines: Thymine (*T*) and Cytosine (*C*). Each have the same sugar, phosphate and base arrangement but vary in their base composition. In 1953, Watson and Crick discovered that DNA did not exist purely as a single string of deoxyribonucleotides but as a double helix, two strands of DNA held together by hydrogen bonds between the bases [5]. These bases pair specifically such that *A* is opposite *T* and C pairs with *G*. This complementation (Chargaffs rules) is a fundamental feature of DNA and is at the center of DNAs ability to replicate, repair and maintain its integrity.

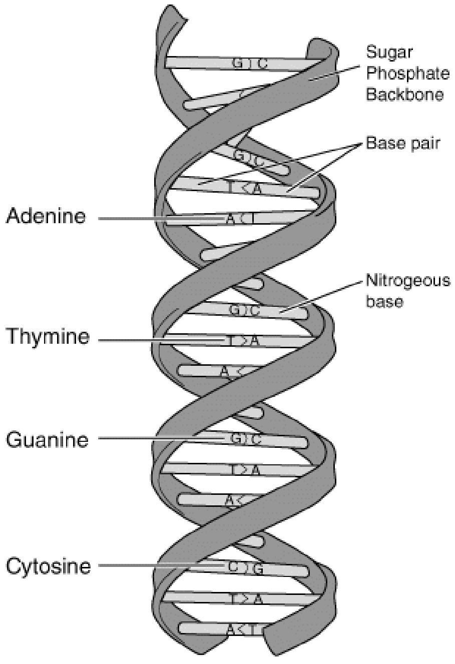


Figure 1.1: Fundemental Structure of DNA
image credit: NHGRI Talking Glossary; http://www.nhgrinig.gov

A *gene* is a regional stretch of DNA whose nucleotide composition has particular properties that are important for the production of ribonucleic acid (RNA) and proteins. Regulatory regions, known as *promoters*, are involved in gene transcription i.e. the production of mRNA. *Introns* separate *exons*, which are sections of genes that are spiced together to form the mRNA (Figure 1.2). Parts of the mRNA are translated into the amino acid sequence of a protein. Gene length is variable; with the largest known human gene, the dystrophin gene, (DMD) approximately 2.4 million base pairs long [1]. Not all genes code for proteins but also for RNA molecules that function in regulation of gene transcription and translation [6].

The *human genome* is the collection of all 30,000 human genes and the DNA that separates them (intergenic) totaling approximately 3.2 billion nucleotides. Protein coding genes may compose only 1-2% of the genome and transcribed genes comprise an additional 33%.

So the function of over 60% of the human genome is unknown. However in this vast DNA wilderness are the tools for cell maintenance, cell communication, tissue organization and organ co-operation: there is a blueprint for life.

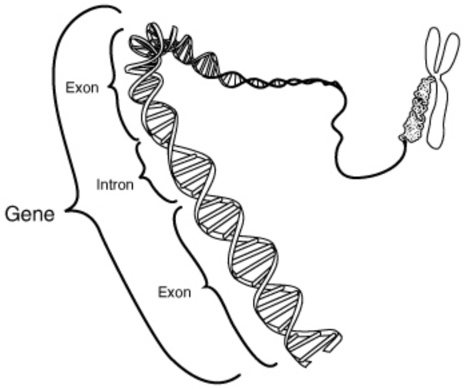


Figure 1.2: Basic organization of a gene image credit: NHGRI Talking Glossary; http://www.nhgrinig.gov

The human genome is broken up into higher order DNA and protein-rich structures known as *chromosomes* (Figure 1.3). Most cell nuclei of the human body contain 46 chromosomes (*diploids*): 23 pairs of autosomal chromosomes and one pair of sex chromosomes with the exception of the gametes, sperm and oocytes. These cells contain only 23 chromosomes (*haploids*), one random chromosome from each pair (Random Segregation). At fertilization, the pairing of these chromosomes is restored; with one member of each pair originating maternally and the other paternally. Those who inherit two X sex chromosomes (XX) are female and those with one X and one Y are males.

The inheritance of chromosomes, dense with genes, from parental lines is the pillar of human and medical genetics. The transmission of parental genes and gene variants is what provides some shared characteristics with their offspring and at the same time makes them unique to everyone else.

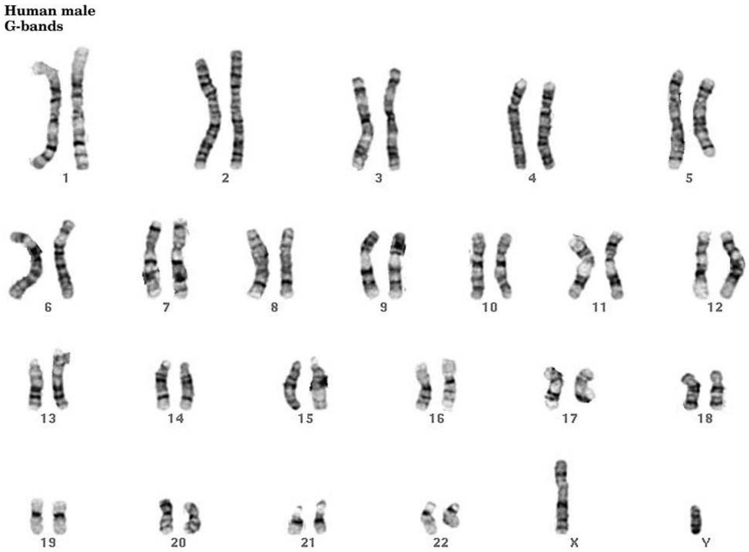


Figure 1.3: Staining of male chromosomes. Note that there is only one X and one Y chromosome

Chapter 2

Basics of Human Genetic Variation

That “The Human Genome” is the same for all humans is a common misnomer. Rather it is a global representation of a number of human genomes. From sequencing multiple human genomes and in-depth gene analysis, much is now known about the range of genetic variation between two haploid genomes in the general population. On average, two chromosomes share 99.9% similarity in their DNA sequence [3]. The remaining 0.1% signifies differences that may have no effect or variants with potential medical consequences such as disease or provide protective effects against infectious agents like viruses and bacteria. The discovery of these differences in the genome, their utility and contributions to disease is the focus of this thesis.

The most frequently observed difference is a *single nucleotide substitution* (Figure 2.1). For the same position in the DNA sequence (*Locus*), alternative base pairs are observed (*Alleles*), whose combination in one individual is a genotype. One allele (major allele) may be found more frequently than the others (minor allele) in a population of chromosomes. Where the minor allele characterizes 1% or more of these chromosomes, the locus is termed a *Single Nucleotide Polymorphism* (SNP).

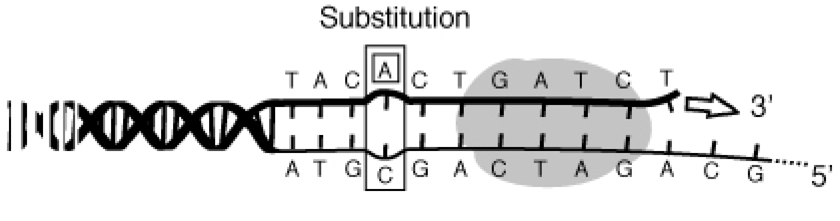


Figure 2.1: An error in DNA replication has introduced *A* instead of *G*

Deletions of nucleotides are less occuring than SNPs in the genome but are as three times as common as insertions [7]. Insertion/deletion polymorphisms are known as Indels. The average deletion and insert is 4.2bp and 4.3 bp respectively. Longer indels are less frequent whereas those of size 3 bp or less are more common.

The human genome is 35%- 50% repeat sequence [2]. There is exceptional variability in the range of these repeated sequences: polynucleotide repeats; tandem repeats (VNTRs), such as mini and microstatellites; bigger repeat elements, like transposons, LINEs, SINEs (Alu). Although the first variations in genomic structure to observed were chromosomal aberrations and aneuploidy, little has been known until recently about population variation in large-scale duplications or rearrangements of chromosomal segments in the human genome. Gains or losses of up to several hundred kilobases potentially containing genes, could be influencing gene expression levels and contributing to disease [8]. Furthermore, the number of (gene) copies appears to be variable in the population and have implications on methods of tracing human history and disease.

2.1 Single Nucleotide Polymorphisms

2.1.1 Classification

SNPs can be characterized in a hierarchy of levels: *Type*, *Location* and *Function*. There are two types of SNP, *transversions*: the substitution of a purine to a pyrimidine or visa versa; and *transitions*: replacement of a purine or pyrimidine with a purine or pyrimidine respectively. There are three forms of transversions owing to base pairing rules: (*G*-*C* & *C*-*G*), (*A*-*T* &*T*-*A*), (*A*-*C* & *T*-*G*), while only one transition: (*C*-*T* & *G*-*A*). Approximately 66% of all SNPs are *C*-*T* transitions followed by *A*-*C*, *C*-*G* and *A*-*T* transitions [9]. This higher *C*-*T* transition rate is primarily due to the increased spontaneous mutability of 5-methylcytosine (5*mC*) through deamination to thymine [10]. Regional *GC* content, nucleotide neighbouring effects [11] and if the strand is coding further influence the rate of this mutation.

SNPs are located in regions with out any protein coding potential are *ncSNPs*. Such are the preponderance of SNPs located within the introns of genes, while exonic SNPs are known as coding SNPs (*cSNPs*). At the functional level, SNPs may affect protein action or the regulatory properties of the gene. Non-synonymous cSNPs (*nsSNPs*) result in an amino acid substitution, which may have deleterious effects on protein function by altering the activity or structure. Conversely, synonymous cSNPs (*sSNPs*) modify the codon sequence but do not alter amino acid. Some result in an amino acid with the same properties while others do not and so the functional consequences of all nsSNPs remain unclear. A number of efforts have been made to develop algorithms designed to predict the impact of cSNPs on protein integrity, taking into account amino acid properties, homologies and effects on overall protein structure [12, 13].

Promoter SNPs or those in the 5' and 3' UTRs may have subtle alterations on gene expression, mRNA stability, translation or transportation to other locations in the cell (regulatory SNPs, *rSNPs*) [14]. ncSNPs adjacent to exons could potentially affect the splicing mechanisms required to construct mRNA [15]. Disruption to mRNA secondary structure stability may be influenced by sSNPs that are located in exonic splicing enhancers elements [16]. As for cSNPs programs have been developed to predict the effects of alternative SNP alleles on gene regulation. However with all prediction algorithms [17], experimentation is required to validate these effects.

2.1.2 SNP Discovery

Determining the range of SNPs in the genome is a fundamental fact if researchers are going to use them to their full potential for human history and disease gene identification. While development of the *polymerase chain reaction* (PCR) [18, 19] has helped the advancement of methods for discovery of low level variation, the requirement for high through put technologies with maximum efficiency and minimal costs has spurred the innovation of new approaches. Techniques for SNP discovery can be categorized into four classes:

1. Confirmation Based Discovery
2. Based Discovery
3. DNA Sequencing
4. In Silico Discovery

Confirmation Based Discovery

The approaches for confirmation-based discovery are based on the principles of sequence context and base pairing. A single strand of DNA is an unstable molecule and will form different structures depending on the context of the sequence. Furthermore, DNA duplex of complementary base pairs is a more stable structure than that with mismatched base pairs.

In Single Strand Confirmation Polymorphism (*SSCP*) amplified DNA fragments are denatured to produce single strands. Strands with nucleotide differences between them will form alternative structures, which will migrate at different rates through non-denaturing gels thereby allowing detection of variants [20]. Advancements have been made to develop different platforms for SSCP, such as capillary electrophoresis, however assessing various conditions is still required and it therefore remains relatively low throughput by comparison to other methods.

The migration of double-stranded DNA with mismatches forms the basis of Confirmation-Sensitive Gel Electrophoresis [21]. Amplified DNA is denatured and allowed to re-nature forming homoduplexes and/or heterduplexes, which will have different migratory patterns. Advantage of this instability has been utilized on a number of different with various detection systems. Denaturing Gradient Gel Electrophoresis (*DGGE*) monitors the early separation strands forming heteroduplexes in comparison to the homduplexes [22]. Denaturing High Pressure Liquid Chromatography (*DHPLC*) monitors the retention times of the heteroduplex strands, which should be lower than homoduplexes. This method can be automated and allows for rapid analyis but requires much optimization. Cleavage Fragment Length Polymorphism utilizes this principle of DNA repair mechanisms through chemical or enzymatic means to identify variants [23].

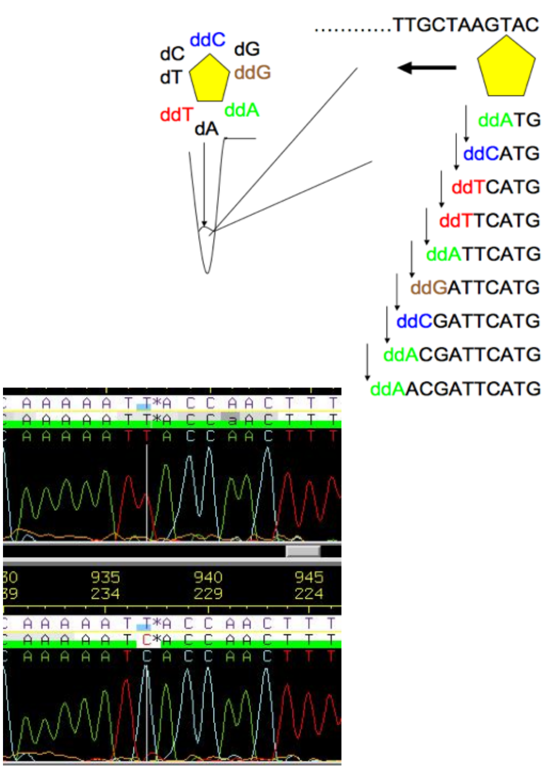
Recent methods for high throughput discovery based on confirmation principles have used the cleavage approach. Photosensitive metallointercallators have been used to cleave heteroduplex strands at the mismatch sites. These strands are be labeled with fluorescent tags and identified by capillary separation techniques [24]. As an alternative, the Mismatch Repair Detection (*MDR*) system in *E. coli* to selectively separate colonies with mismatch containing plasmids [25]. They have used this method in the identification of mismatches from a 1000-plex PCR in a 96-well format, which can be labeled with fluorescent probes and hybridized to arrays using tags. This approach appears to be very sensitive to rare SNPs and in application to Autism, they sequenced amplicons showing significant differences between cases and control confirming nsSNPs in the same exon of *CMYA3* [25].

DNA Sequencing

The principles of DNA sequencing have been around for 30 years since Sanger introduced the plus-minus method, based on DNA elongation with DNA polymerase [26]. A chemical cleavage method was developed by Maxam and Gilbert [27], however, it was Sanger who coupled

DNA elongation with dideoxynucleotides (*ddNTPs*), to revolutionize the field of genetics. Consequences of the Sanger method have been vast including the deciphering of human genome, but it is instrumental for SNP discovery in candidate genes for disease. It permits direct SNP identification and the characterization of its alleles.

The concepts of Sanger sequencing are based on the *extension* by DNA polymerase, of an oligonucelotide (*primer*) that is hybridized to a single strand of DNA, Polymerase incorporates a complementary base of the single strand to the 3' end of the primer however, if ddNTPs are used, the reaction is terminated and elongation no longer continues. This is due to the fact that ddNTPs lack the 3' hydroxyl group and the incoming nucleotides' 5' triphosphates cannot form phosphodiesterbonds bringing the reaction to a halt. Therefore, when the correct ratio between a ddNTP and its corresponding dNTP is pooled with the other three dNTPs, repeated processes of primer annealing and elongation (*cycles*) will produce different lengths of elongated products (*nested products*) due to ddNTP termination at different positions. Consequently, by resolving the length of the nested products, the order of ddNTPs and their positions can be determined.



............TTGCTAAGTAC
dC
ddC
dG
dT
ddG
ddT
ddA
dA
ddATG
ddCATG
ddTCATG
ddTTCATG
ddATTCATG
ddGATTCATG
ddCGATTCATG
ddACGATTCATG
ddAACGATTCATG
A A A A A T T*A C C A A C T T T
A A A A A T T*A C C a A C T T T
A A A A A T T A C C A A C T T T
935
234
940
229
945
224
A A A A A T T*A C C A A C T T T
A A A A A T C*A C C A A C T T T
A A A A A T C A C C A A C T T T

Figure 2.2: Basics of a DNA sequencing reaction. Flourescently labeled ddNTPs, dNTPs, a DNA template, a DNA primer and a DNA polymerase (pentagon) are combined. Cycles of primer extension produce nested products of various lengths which can be seperated to identify the ddNTP at the postion. Chromatogram shows a single nucelotide difference T>C

Initial studies performed reactions for each ddNTP separately, running on polyacrylamide gels for high-resolution separation of strands and were aided by the development of PCR and asymmetric PCR [28]. Advancements have been made in ddNTP chemistry [29], detection [30] and reaction kinetics [31]. Contemporary high throughput approaches now involve the labeling of each ddNTPs with an unique flourophore and separation is carried out using capillary-based instruments [32]. Each base is determined from the quality of the fluorescent signal [33, 34] and tools are available to visualize the finished sequence [35]. Through the assembly and alignment of multiple sequences, low-level alterations such as SNPs can be observed and extensions to algorithms for base-calling and quality checks can be implemented to aid SNP discovery [36].

Sequencing is a much more expensive method than confirmation-based strategies. However, as it provides more information, improvements are being developed to increase the high throughput capacity of current sequencing technologies, reduce running times and costs. These are either based on Sanger sequencing principles or new alternatives. Microelectrophoretic sequencing is an upgrade of Sanger sequencing technology, using microfabricated capillaries with more control over sample injection and less running time [37]. PCR amplification-free methods are being introduced for sequencing single molecules of DNA [38], including reversibly terminated nucleotides [39] and nanopore technology [40]. Alternatives to the Sanger sequencing, but require amplification, are fluorescent in situ sequencing (*FISSEQ*) [41] hybridization [42] and the Pyrosequencing [43]. With improvements to the latter there is the potential for whole genome sequencing or extensive candidate gene sequencing in under 4 days however of the three alternatives, hybridization has had the most impact on our knowledge of SNP patterns and diversity [44, 45].

In silico Discovery

The depositing of DNA sequence to large databases permits the alignment and comparison of sequence reads for SNP discovery. The sources of these sequences were initially from the Human Genome Project, which used a number of ways clone large amounts of DNA sequence from diploid individuals such as *bacterial artificial chromosomes* (BAC), *P1-based artificial chromosomes* (PAC) and *expressed sequence tags* (EST). The databases housing these reads are therefore pools of sequences from many chromosomes and by careful analysis; it is possible to identify SNPs and other variants. The same tools in sequence analysis aid the process of identifying SNPs from database sequences, however additional tools were developed for high throughput purposes such as POLYBAYES [46, 47]. These were primarily based on the quality and dept of reads but required efficient alignment to reduce false positives. By 2001 75% of the SNPs submitted to dbSNP were from alignments of BACs [48], PACs [48] and ESTs [49, 50]. Nonetheless, SNPs discovered through such alignments require confirmation by sequencing. As sequencing technologies improve and confidence in the quality of data increases, alignments of reads from genome data bases will greatly facilitate medical genetics.

Chapter 3

Nucleotide Diversity and Neutrality

Understanding nucleotide variation is an important element of re-sequencing projects. Insights into SNP diversity and patterns can tell us about the factors that have shaped them, which is important for tracing human evolution, but critical for medical genetics and the identification of disease genes.

3.1 Estimators

There are a number of ways to measure nucleotide diversity. Most often used are the summary statistics π, the mean number of pairwise differences per site between two sequences chosen at random [51] and θ_w the number of segregating sites [52]:

$$\pi = \sum_{ij} x_i x_j \sqcap_{ij} \tag{3.1}$$

where $\sqcap_{ij}$is the proportion of nucleotide differences between the ith and jth DNA sequences, and x_i and x_j are the frequencies of these respective sequences;

$$\theta_w = \frac{S}{a_1} \qquad \text{where } a_1 = \sum_{i=1}^{n-1} \frac{1}{i} \tag{3.2}$$

and n is the size of the sample. Essentially these are estimates of a central parameter of the neutral theory of molecular evolution (Neutral Theory), the *population mutation rate*:

$$\theta = N_e \mu \tag{3.3}$$

where N_e, is the effective population size and μ is the neutral mutation rate per generation. This was derived by Motoo Kimura [53] to explain that the high variation in protein electromorphs between species observed at that time was not due to equilibrium between selection and mutation, but rather to random genetic drift. In stating this, it was implied that much of the variation was indicative of neutral, or nearly neutral polymorphisms in a balance between mutation and stochastic sampling. Selection, background or positive, could never be ruled out in shaping the patterns of variation to some degree but it was emphasized that the variation remaining had little or no effect on fitness. Potentially departures from neutrality could then be considered as signs of selection.

The *Standard Neutral Model*, is based on the conditions of population panmixia, constant size and the *infinite-site mutation model* (ISM) where recurrent mutation at the same site does not occur. Under the ISM, estimates of π and θ_w approximate the neutral mutation parameter [54]. If the Standard Neutral Model is assumed then departures from neutrality could be estimated by comparison of π and θ_w. One common measure to compare these estimators is Tajimas D [55]:

$$D = \frac{\hat{\pi} - \hat{\theta_w}}{\sqrt{Var(\hat{\pi} - \hat{\theta_w})}} \tag{3.4}$$

where $\hat{\pi}$ and $\hat{\theta_w}$ are estimated from the data, and the denominator is the the square root of the variance between the two.

Fu and Li developed statistical tests considering the distributions of mutations occuring on the external and internal branches of a genealogy tree [56]. Older mutations are in the internal branches of the tree while younger mutations are located on the external branches. In their models there are two expectations of θ based on the mutations in the external branches and internal branches. Under neutrality there should be no difference in these estimates of theta. D* tests for the number of *singletons* (polymorphisms appearing once) to overall number of polymorphisms, whereas F* tests for the number of singletons in the external branches against all pairwise differences.

It is useful to consider what factors would effect neutrality to interpret the outcomes of these measures. For example, *Background selection* (negative or purifying selection) will create an excess of deleterious mutations in external branches because of their low frequencies. In their elimination, linked sites would also be removed and so we would expect to see an increase in rare variants compared to higher frequency polymorphisms. In such cases Tajima's D and Fu and Lis D* and F* will be negative. (See figure 3.1 below)

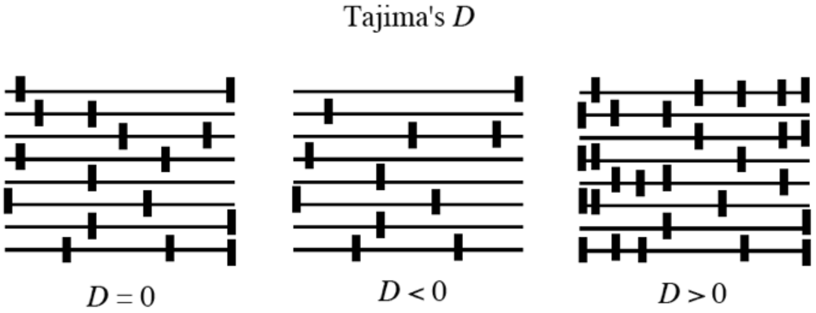


Figure 3.1: Patterns of variation for different ranges of Tajima's D

Positive and balancing selection will increase the frequency of older mutations. However positive selection will create mutations on the external branches if advantageous mutations are becoming fixed and so the majority of them are expected to by young. On the other hand balancing selection will be have a deficiency in external mutations, replacing older mutations with new alleles with intermediate frequencies. As such these forces will in turn influence π but have little affect on θ_w. Because of the excess of higher frequencies polymorphisms, Tajima's D and Fu and Lis $D*$ and $F*$ will be positive.

It is important to consider that these are measures from neutrality based on the genealogical process and are not just affected by selection. Population dynamics also has effects the genealogical process and in such influences the patterns of diversity in ways that are virtually indistinguishable from other forces.

3.2 Patterns of Human Diversity

Since the 1980s with the development of the polymerase chain reaction (PCR), restriction fragment length polymorphisms (RFLP) techniques [57] and sequencing technologies [26], a lot more information is available for interpretation of DNA nucleotide diversity within and between populations. A range of studies focused on re-sequencing particular genes involved in biological and disease systems [58, 59, 60, 61, 62] or a large number of genes in greater sample sizes to understand the distribution of SNPs in the genome [63, 44, 45, 64].

Eleven million SNPs are predicted in the contemporary population of ”human genome”, occurring on average once every 1331 base pairs (bp) (1/1331bp) between two chromosomes and increasing from 1/300bp to 1/100bp with larger sample sizes [65]. SNPs are more often bi-allelic, however instances where more than two alleles have been reported [66]. The vast majority of these are rare (MAFs $\leq$5%, greater than 4 Million approx while the average MAF freq is approx $\sim$0.11.

The distribution of SNPs based on gene location and function is consistent for all genes rather than for specific genes per say. Protein-coding polymorphisms are only a subset of all variation and functional constraints on proteins explains why cSNPs tend to be less observed and less frequent than ncSNPs, that is they are not very diverse [67]. Among cSNPs, there appears to be an even distribution in the numbers of sSNPs and nsSNPs but nsSNPs tend have lower frequencies in the population and therefore less diverse . Diversity increases as the location of the nsSNP moves from non-degenerate to two-fold degenerate to four-fold degenerate codons [68]. On average there are 4 cSNPs per gene, with an expected 120000 to 219,000 genome wide [65, 69]. Taking into account the occurrence rate and frequencies between 50,000 and 876000 (40%) are predicted to be nsSNPs. Diversity increases in introns, while 5 and 3 UTR appear to have similar diversities as the genome wide averages.

The distribution density of SNPs across the genome is non-random because of evolutionary and population dynamic forces [70]. Some regions show higher levels of diversity than others which seems to be consistent with features of the human genome, such as simple-repeat content, *GC* content, *C*pG islands [71], distance from telomeres and centromeres are also apparent predictors of diversity levels. Some regions of the genome are variation “desert”, such as on the X-chromosome [72], which has an estimated 40-80% of the autosomal variation [73].

Consistent with early protein studies are observations that most of the nucleotide variation is within populations (84.4%) rather than between [68][74]. However, based on autosomal, X-chromosomal and mitochondrial DNA studies, the African continent houses more variation than outside [75, 76, 77, 78]. Non-African variation is a subset of that within Africa. Differences in some polymorphism frequencies may possibly reflect the migration and adaptation of Homo sapiens in different environments during its evolution, which may also reflect the susceptibility of contemporary humans to common diseases today.

Chapter 4

Linkage Disequilibrium, Estimators and Haplotypes

According to the infinite sites model, nucleotide substitutions at different positions occur independently. Therefore, mutation events occur on chromosomes with the same or different backgrounds of extant nucleotide variation. As a result the mutation is correlated with the other variation that is present on the chromosome.

Consider two independent mutation events on the same chromosome. Each event creates two alleles: A and a at locus 1; B and b at locus 2. A combination of these alleles along the chromosome is known as a *haplotype* e.g. AB or ab. If a combination of alleles rises to higher frequencies in a population, then a correlation may be observed between the new and extant variation such that Allele A at locus 1 is found more often Allele B at locus 2 than predicted by random chance. This departure from independence is known as *Linkage disequilibrium* (LD) or *Gametic phase disequilibrium*.

A number of measures for LD have been proposed, and their properties have been studied extensively. The earliest known estimator was *D*, the difference between the observed haplotype frequency, P_{AB}, and what is expected under equilibrium:

$$D = P_{AB} - P_A P_B \tag{4.1}$$

D is positive or negative depending on the arbitrary labeling of alleles and is affected greatly by sample sizes and allele frequencies. Consequently, the majority of other methods for measuring LD attempt to take these factors into account. A more commonly used measures is D' [79], a scaled measure of D to control for allele frequencies:

$$D' = \frac{D}{D_{max}} \tag{4.2}$$

where D_{max} is the lesser of $P_A P_b$ or $P_a P_B$ if D is positive or $-P_A P_B$ or $-P_a$ if D' is negative. D' is a ratio of observed LD to the maximum LD possible given the polymorphism allele frequencies. Values of D' are between 1 and -1 depending on the arbitrary labeling of alleles. As D is symmetric, $\mid D' \mid$ values (absolute) are reported. $\mid D' \mid$ values equal to 1 imply complete LD whereas $\mid D' \mid$ equal to 0 indicates total independence. Recombination is a factor in reducing the associations between alleles [80]. Complete LD is then interpreted as a situation where recombination has not occurred. Intermediate values of $\mid D' \mid$ are difficult to interpret but are believed to be signatures of ancestral recombination. $\mid D' \mid$ is upwardly biased by sample sizes and rare allele frequencies [80, 81]. Although some argue that $\mid D' \mid$ is not influenced by allele frequencies, others claim that no method for measuring LD is free from the effects of allele frequencies [82]. An alternative measure of LD is the coefficient of correlation [83], r^2:

$$r^2 = \frac{D^2}{\sqrt{P_A P_B P_a P_b}} \tag{4.3}$$

Only when alleles have the same frequency r^2 equals 1 indicating that they occurred in proximity to one other on the same branch of the genealogy lineage and have not been separated by recombination [84]. This is considered perfect LD because the alleles are fully informative of each other. r^2 is typically lower than $\mid D' \mid$ for any physical distance (kb). r^2 has a number of pit falls including sensitivity to allele frequency and issues with interpretation. However, estimating r^2 provides a number of advantages in population genetics and association mapping studies using comparative re-sequencing [84].

For association mapping purposes, the inverse of r^2 may be used to estimate the increase in sample size required to obtain an association with a marker and have the same power as testing the disease allele directly [80]. For example, in order to find an association with a maker with the same power as the disease polymorphism, which has an r^2 of 0.5 with a disease polymorphism in a sample of 100 chromosomes, the sample size needs to be doubled (1/0.5). This measurement indicates of the extent of useful LD in association studies. Distances over which $\mid D' \mid$ decreases by half ("half-life" of $\mid D' \mid$) has also been suggested but since $\mid D' \mid$ overestimates the magnitude of LD, application of the half-life to fine scale mapping remains unsure [85]. To test against the null hypothesis of no LD, the significance of LD between markers can be obtained by dividing r^2 by the number of chromosomes, estimating a chi-squared value and with 2 degrees of freedom [84].

In a population genetics setting, the standard neutral model has been used to predict the extent of LD [84]. In this model r^2 is an estimator of $1/(4N_e c + 1)$, where Ne is the effective sample size and c is the per nucleotide recombination rate. Therefore, r^2 is a function of the scaled population recombination rate; high recombination rates imply lower r^2, whereas if $4N_e c$ is low then r^2 is approximately 1. An advantage of this scaling is that it allows the comparison of regions rather than pairwise measures of LD. Furthermore, other factors such as population dynamics, selection and crossover events contribute to the estimates of $4N_e c$, making it more applicable to comparisons within and across population.

4.1 Haplotype Diversity and Neutrality

The diversity of haplotypes can be estimated similarly to gene diversity ([86]:

$$H_d = 1 - \sum {p_i}^2 \tag{4.4}$$

A number of tests have been developed to study departures from neutrality based on haplotypes. These tests center on the *infinite alleles model.* According to this model, each mutation occurring in a background of pre-existing mutations creates a new haplotype. Based on this concept, Ewans described a formula using the neutral mutation parameter θ estimated from the number of segregating sites to predict the expected haplotype frequencies [87] .

Strobeck modified Ewans sampling formula by using θ estimated fom π in his S test to determine if a population was mating randomly or if substructure existed [88]. If there is substructure then the number of haplotypes is expected to be less than predicted by the modified formula. Therefore, the probability of obtaining a sample with less than or equal to the number of observed alleles which are in the sample is estimated. However, since the effects of balancing selection are proposed to be similar to that of substructure i.e. a deficit in haplotypes, then the S-test has also been suggested to be an estimate of departure from neutrality.

Fu and Li proposed the F_s test as an alternative to Strobecks S [89]. The modification S is essentially the opposite of S i.e. the probability of observing more haplotypes than that found in the sample based on π. F_s tends to be negative when there is an excess of recent mutations and appears to be more powerful for detecting background selection.

Chapter 5

Human Genome Patterns of Linkage Disequilibrium

5.1 Predictions, Empirical Data and The International *HapMap* Project

Early studies of linkage disequilibrium involved the use of micro-satellite markers and simulations to predict the extent of LD in the human genome [90, 91, 92]. However, SNPs are more abundant in the genome and with improvements in technologies for their discovery/detection, SNPs have been the main tool in research to shed light on the complexity of LD in human populations.

A number of independent groups and an international collaboration, the International *HapMap* Project, have tried to predict and study LD empirically across an extensive range of genes, genomic regions and whole chromosomes, as well as the entire human genome in order to get a better understanding of LD and haplotype structure. Typically, patterns have been examined in a number of populations from different continents to get a global perspective on LD. The data produced has provided a wealth of information for future studies of population genetics, and they will have a large impact on medical genetics for disease gene identification.

LD simulations based the expansions of a small founder population to the current population size over 5,000 generations predicted that LD would not be found beyond 3kb of DNA in the genome [90]. Some researchers argued the model used to derive this expectation was inconsistent with the LD in current populations and empirical studies have revealed regions of the human genome where LD extends far beyond 3kb [93]. Early empirical studies using microsatellites demonstrated that LD across the genome is heterogeneous, often extending over long distances with an inverse relationship with physical distance. However, over shorter distances LD was more variable and less predictable [94]. Although LD between SNPs is does not extend as far as for microsatellites, the LD pattern displays the same relationship with distance but is more consistent with a genetic (cM/Mb) rather than physical map (kb) [95].

The degree of LD correlates with features of the genome such as *G+C* content, SINE repeats, gene density and gene function for both physical and genetic maps [96, 97, 4]. There appears to be low LD within genes involved in immune response and neurophysiology whereas higher LD is observed among genes involved in DNA and RNA metabolism, DNA damage, and the cell cycle. Extensive haplotypes are observed towards autosomal centromeres and the

X-chromosome appears to have even more regions of extended LD [4].

In a number of studies spanning large regions of the genome, including entire chromosomes, short regions of low LD often interrupt areas of high LD [85, 95, 98]. These regions of high LD are characterized by low haplotype diversity, where 2-5 common haplotypes (MAF $\geq 5\%$) explain much of the variation, and low recombination rates [99]. Regions of low LD correspond to increased haplotype diversity and potentially higher recombination rates. In a number of studies, it has been demonstrated empirically that some of these regions of low LD do correspond to regions of elevated recombination and they have been termed recombination "*hotspots*" [100]. In some cases, the majority (50-80%) of all recombination occurred in a subset (10-15%) of the sequence [101]. However, regions of high LD may also have recombination hotspots that may have arisen recently [102, 103], which is consistent with the observations of different hotspots in non-human primates [104].

Regions of low haplotype diversity are termed "*haploblocks*" and can also be defined based on the extent of pairwise LD half-life $\mid D' \mid$, or by combining both haplotype diversity and LD [105]. SNP information in haploblocks is partially redundant and a subset of the SNPs can be informative of other SNPs in the region. These "*haplotype tagging*" SNPs (htSNPs) can be selected based on the methods used to define the block structure. For example, SNPs can be chosen to capture a percentage of the haplotype diversity [106], distinguish between the haplotypes [99] or exceed a threshold of LD with other SNPs. As the block definition is subjective, strategies have been developed to capture SNPs based on regression of multiple measures of LD and so are not restricted by the definitions of block boundaries but can span them [107].

The alignment of block boundaries and block patterns are similar between related populations [108]. Conversely, the degree of LD and the extent of haploblocks is much smaller in African population samples than in non-African populations [109, 110]. As for SNPs, the majority of haplotype variation is explained by differences within the populations rather than between. The sequence covered by the blocks is often dependent on which method was used to define recombination such as pairwise LD or the Four Gamete Test. Eventhough haplotypes have been found to span recombination hotspots, their length is also correlated with genetic distance [4].

Due to differences between populations, the concept of using a common subset of htSNPs to tag global variation is questionable. Difficulties in comparing patterns of LD and haplotype blocks between populations are due to polymorphism density, the measure of LD, sample sizes, large numbers of pair-wise comparisons, and inconsistent selection of SNPs with high frequencies that are not shared between all populations (*ascertainment*) [111]. For example, in studies where SNPs were selected from databases it has been shown approximately 50% of the genome is organized in this block fashion [112], while SNPs identified by re-sequencing such as that obtained from the Environmental Genome Project (EPG) showed better blocks [69]. Additionally, a number of studies have shown that factors other than recombination may result in block-like haplotypes.

Attempts to understand the fine-scale resolution of LD using genetic map are very recent. There is a consistency in recombination events recombination rates between populations measure using dense marker sets across regions. However, population specific recombination spots can be observed. This complements the pairwise measures of LD to some degree but also questions the definitions used to describe haplotype blocks and the shared patterns between populations [113].

Chapter 6

Factors Shaping Patterns of Diversity & Linkage Disequilibrium

Tests for deviations from selective neutrality are sensitive to other factors influencing the assumptions of the standard neutral model and gene history. The measures and patterns of diversity are variable possibly reflecting past events other than selection in the genealogy. However, there are correlations in levels of diversity over distances as short as 100bp to 1000kb that are consistent with patterns of LD [70]. This indicates that the same forces shaping diversity have consequences regarding the correlations between polymorphisms.

6.1 Mutation Rates and the Neutral Mutation Parameter

Error in the replication and repair of DNA is a critical source of mutation. Most of our knowledge on mutation rates is provided by comparisons of genes between species and pseudogenes. These are based on the neutral theory of molecular evolution, which assumes that the rate of neutral mutation is similar between and within species. A number of studies find that the neutral mutation rate lies between 1.3x10^{-8} and 3.4 x10^{-8} with an average of 2.5x10^{-8} [114]. Therefore, the number of new mutations per diploid genome per generation is between 115 and 175 [65, 63]. Genes show differences in their mutation rates. These are often correlated with features such as coding regions, introns, *C*pG content and repeat. The X-chromosome also appears to display lower mutation rates perhaps due to lower mutation rates in females (male-driven molecular evolution) [115], while other studies have not found significant differences between the X and autosome chromosomes.

Variation in mutation rates is predicted to affect SNP density and subsequently influence π and θ_w. Assuming regions with no function and high mutation rates, the number of potentially rare mutations will be greater, effecting θ_w more than π. A negative Tajimas D will subsequently indicate an excess of rare variants, as will Fu and Lis D* and F*. Furthermore back mutation (*homoplasies*) in these regions cannot be ruled out which violates the infinite sites model. Regions with high mutation rates will result in rare mutations on several branches of the genealogy, which creates weaker correlations between neighboring SNPs. Therefore, if the neutral mutation parameter is high, the haplotype block size is expected to be smaller than that predicted by lower rates of mutation. From the point of view of evolutionary and disease genetics, mutation will break down the size of haploblocks [116].

6.2 Random Genetic Drift

Not every mutation that arises is going to persist in the population for a long period of time. Owing to random mating, the effects of sampling from one generation to the next will either be the complete loss of new mutations and/or haplotypes or a "*drift*" (fluctuation) between frequencies. As a result, there is a potential for the new alleles to replace the ancestral allele and become fixed. This is a function of the population size $(1/2N)$. In a large population, it is expected that a new neutral mutation will survive approximately ten generations before being lost [117]. On the other hand, in smaller populations, such as geographically isolated groups, the effects of drift will be greater.

These fluctuations from generation to generation may explain some of the patterns of variation that are observed in extant populations. The majority of variants are rare because of drift. When the effects of population dynamics and selection are excluded, drift is expected to be a major driving force in molecular evolution and is the key factor in the *Neutral Theory*. The balance between rates of neutral mutation and drift will result in similar neutral parameters by pi and thetaw and Tajimas D is expected to be zero. The number of mutations from the inner and outer branches of a genealogy will reflect this, and Fu and Lis $D*$ and $F*$ will be near 0.

Linkage disequilibrium is expected to increase under random genetic drift. The effects will be more pronounced in small populations. This is due to the loss of chromosomes, which will increase the correlation and extent of association between polymorphisms that have occurred on the same chromosome or descend from a common ancestor [118].

6.3 Recombination and Gene Conversion

Recombination is the balanced crossover of genetic material during prophase I of meiosis. As a result, there is shuffling of genetic material from generation to generation without altering the density of existing polymorphisms, unless new mutations occur. Correlations have been shown between the recombination rate and nucleotide diversity [119], which may be due to selection processes or that recombination is mutatgenic [120, 121]. The average genome recombination rate is approximately 1.1cM/Mb based on genomic-physical distance maps. However, recombination is neither uniform across the genome nor similar between the genders. Moreover, for the sex chromosomes recombination is also lower than the genome average [122].

Many of the models that are used in population genetic analysis assume absent or uniform recombination across the regions under investigation. Recombination will increase allelic combinations, influencing the variances of π and the number of segregating sites without affecting their means. As a result, many of the summary tests based on θ estimated by the number of pairwise differences (e.g. F_s and Strobecks S) could potentially be either conservative or inflated.

The detection of recombination from population genetic data is difficult although a number of methods have been proposed for this purpose including the *Four Gamete Test* (FGT). Under an infinite sites model the FGT assumes that the four haplotypes observed from diallelic markers can only be created by balanced crossover and not by homoplasy [123]. Performing this test across a region will identify recombination spots that overlap giving a conservative number of recombination events R$_m$. This minimum number of recombination events can be used to estimate the population recombination rate. However, as it does not identify all potential events that have occurred in the history of the sample, it is not expected to be very precise. Hudson proposed another method by which C (also known as ρ) is estimated from the variance of the number of site differences between pairs of sequences in the sample [123]. However, a number of groups have now developed other ways of estimating recombination rates from population genetic data. These methods have been designed to be more precise by taking sample history into account through coalescent modeling and are insensitive to SNP ascertainment [124].

Recombination is the strongest force that will break down LD. A theoretical rule of thumb, LD between two sites per generation will be:

$$D_t = (1 - \theta)^t D_0 \tag{6.1}$$

where D_0 is LD at the time 0, D_t is LD at the t'th generation broken down by the recombination fraction θ [80]. Therefore, with distance and time, LD will decrease gradually, as seen with absolute $\mid D' \mid$ and r^2. However, over shorter distances LD can be very variable. A major driving force in shaping these fine scale patterns of LD is suspected to be *gene conversion*: the non-reciprocal transfer of DNA from one chromosome to another (figure). Gene conversion has been overlooked as a mechanism shaping the structure of genomic LD. If gene conversion occurs as common as predicted, and as interest in copy number variants increases, then gene conversion will have to be taken into account for evolutionary and disease mapping purposes [125, 126].

Gene conversion (Fig. 6.1) replaces the alleles on one chromosome (*Acceptor Strand* , Blue) by the alleles of the other (*Donor strand*, Red) and can be identified as tracts of two or more consecutive SNPs that could also be homoplasies [127]. Gene conversion tract lengths are on average between 350-1000bp and appear to be frequent in regions of high *GC* content, possibly to counteract the increased mutation rates at *C*pG islands [128]. As the tracts are quite small,

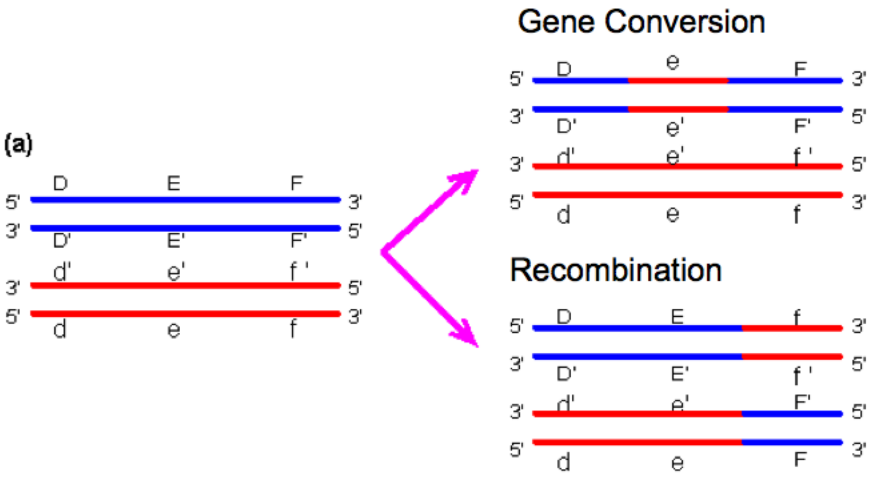


Figure 6.1:

gene conversion cannot be as easily observed between markers that are outside the tract as in low-density marker scans for LD structure determination. When higher densities of markers are used, LD varies much more and haplotype block sizes decrease. Andolfatto and Nordborg pointed out that gene conversion could contribute to the per generation recombination rate, c, if the distance between two sites is equal to or smaller than the length of the gene conversion tract [129]. If gene conversion inflates c, estimates of the population recombination rate, C, will be expected to increase. As r^2 is a function of the population recombination parameter, gene conversion will reduce allelic associations at high SNP densities.

Gene conversion can occur intra-chromasomally between duplicated sequences on either side of the same chromatid or sister chromatid and thus plays a significant role in the evolution of multi-gene families such as the globin gene, ribosomal and *MHC* [130]. Hayashida described that gene conversion involving the X-linked Zinc Finger (ZF) genes could possibly explain the unusual divergence pattern of the mammalian Y-linked ZF genes [131]. Gene conversion occurs frequently in the germ-line [132] and is important in forensics [133].

Gene conversion has been important in phenotype variation such as color vision and olfactory receptor evolution [134]. Amino acid substitutions in the red cone gene (*OPN1LW*) are caused by the gene conversion of alleles from the green cone gene (*OPNL1MW*) on the X-chromosome. In the HLA class three region, gene conversion between *CYP21B* and pseudogene *CYP21A* is responsible for 75% of the mutant alleles that cause congenital adrenal hyperplasia [135]. Of the 89% of unrelated individuals with Shwachman-Diamond syndrome, 60% have two converted alleles from the pseudogene *SBDSP* [136]. Gene conversion between two functional genes plays a significant role in chronic pancreatitis where 289bp from the anionic trypsinogen gene *PRSS2* is relocated to exon two of the cationic trypsinogen gene *PRSS1*, resulting in two exonic mutations increasing susceptibility for spontaneous activation to trypsin [137]. Interchromosomal gene conversion of alleles from homologous pseudogenes on chromosome 22 to chromosome 12 is important in von Willebrand disease [138].

6.4 Molecular Selection

Mutations that result in amino acid substitutions can potentially have a number of consequences depending on how they influence the *fitness* of the carrier. For example, deleterious amino acid substitutions may result in congenital deformations or early-onset disease, which reduces the chances of the carrier to pass on the mutation to the next generation. Therefore, the mutation is negatively selected and maintained at low frequencies. The contrary may be true for beneficial mutations. Mutations that improve the survival of the carrier to reproductive age will be positively selected and increase the likelihood of that mutation becoming polymorphic or fixed in the population. Alternatively, mutations will occur that may have no effect on fitness, be slightly deleterious, or have little effect on improving fitness. These mutations are considered neutral or slightly neutral and their frequency in the population is subject to random genetic drift [139].

Selection will have local effects on the neutral polymorphisms surrounding the mutation. In the case of negative selection, diversity will be reduced, as polymorphisms surrounding the mutation will also be quickly eliminated resulting in an over-representation of rare alleles. Both a negative Tajimas D and a negative Fu and Lis D* and F* will indicate the excess of rare variants in these regions. As new alleles will occur on different chromosomes and frequencies will be low, LD between the alleles will be complex. Some alleles will show moderate LD, while others may not be in LD at all. Alternatively, a positively selected polymorphism will increase in frequency along with the neutral polymorphisms in the surrounding [120]. In the process, there will be a decrease in a number of pre-existing high frequency polymorphisms and an increase rare variants, an occurrence that is also be indicated by negative Tajimas D and Fu and Lis D* and F*. LD will be increased between the new linked sites. Because the extent of selection around the mutations will be subject to recombination, using LD as in indicator of positive selection has been of debate, as some argue it will breakdown rapidly after a selective sweep [140].

Under balancing selection, when two or more alleles at a particular locus have selective advantage, alleles will be maintained at intermediate frequencies. In contrast to positive selection, there will be a deficit in the number of rare variants. However the effects on neutral variation in the vicinity of the polymorphisms will be similar, in that existing polymorphisms hitch-hike along with the beneficial polymorphisms, leading to an excess of heterozygotes near the selected polymorphisms [141]. Subsequently, the number of pair-wise differences and subsequently Tajimas D will be positive. Fu and Li's D* and F* will also be positive also as the number of internal, older mutations will exceed that observed in the external branches of the genealogy.

Balancing selection can be further categorized into *negative frequency-dependent selection*, where the fitness of allele decreases as it becomes more common, or *generalized overdominance*, where heterozygotes have a selective advantage over homozygotes [141]. Selectively balanced polymorphisms are greatly influenced by the development and adaptation of humans to different environments, immunity and protection against infections agents. Most notable of these are the HLA loci [142], Glucose-6-phosphate dehydrogenase (*G6PD*) and protection against malaria [143], *CCR5* promoter polymorphisms and protection against HIV [144]and most recently *PRNP* [145].

Much of the polymorphisms occurring in coding regions of genes is a result of selection. Diversity in coding regions is lower especially for non-degenerate mutations. Many coding regions in human genome do not have an excess of rare alleles and indicating that balancing selection is more common than previously anticipated

6.5 Population Dynamics, History and Structure

As opposed to selection, which shapes diversity and LD locally, the dynamics of populations shape diversity across the genome. Consequently, the same forces influence associations between markers and shape LD.

Population growth will reduce genetic drift. As a result, the balance between mutation and drift is tilted towards mutation, increasing the number of rare alleles and reducing of LD. The extent of this effect depends on the growth rate, with the greatest breakdown in LD happening in a rapidly growing population [84].

A decrease in population size followed by growth (a *bottleneck*) will reduce genetic diversity and increase LD between markers that have not been eliminated. The decline in diversity is of course dependent on the severity of the reduction in population size and the rate of population growth following the bottleneck. The effects of migration are similar to that of the bottleneck effect, resulting in a reduction in diversity and an increase in LD in the new founder population. The effects on the extent of diversity loss and LD formation will be greatest in the migrating population size is small, owing to genetic drift. Within populations there may be substructure. Drift and other forces will act in each subpopulation and function in shaping the degree of within subpopulation diversity. The effects on LD within each population are as described above; however, when substructure is not detected, LD may be high because of within population allele associations [80, 116]. Migrations between the subpopulations (admixture) will also create LD but is dependent the time of admixture on the rates of gene flow stephens). The effects of subdivision will reduce in the following generations.

Chapter 7

Application of Human Genetic Variation

7.1 Molecular Anthropology & Study of Human Evolution

Over the last 50-60 years developments in molecular techniques and automation have provided more in depth knowledge of genome diversity within and between humans and other species. Antigenic protein analyses, hybridizations of non-repetitive DNA and sequence comparisons show that African Apes are closely related to humans, suggesting a most recent common ancestor (MRCA) about 5 million years ago (MYA) with chimpanzees (*Pan troglodytes*) [146] and about 8 MYA with the gorilla. *Pan troglodytes* are 98-99% similar to humans in terms of DNA sequence, but greater differences up to 5% are observed when insertions and deletions are taken into account [147]. Studies concerning variations in insertions, repeat copy number and genome structure illustrate that speciation and human development has been dynamic [148].

Early comparisons of blood proteins indicated that there was greater variation within human populations than between them, and more differences were present in the between the African compared to the non-African continents [147]. Based on such patterns, it was proposed that non-Africans were from the Africa. From studies of mitochondrial DNA (mtDNA), regions of the X and Y-chromosomes and of autosomal DNA from global samples show that diversity observed in non-African populations is a subset of that found with in the African population [76, 77, 78, 134, 149, 150, 151]. This further suggests that a population of approximately 10,000 *Homo* sapiens migrated "*Out of Africa*" an estimated 100,000 to 200,000 years ago into the middle-east replacing other species of *Homo* such as *Neanderthals*. Additionally because of the diversity between Africans and native Australians, it has also been suggested there were migrations south-west to east.

There are alternatives to the Out of Africa theory, such as the *multiregional hypothesis*, which argues that *Homo sapiens* developed globally, rather than from within Africa alone, following a migration of *Homo erectus* out of Africa. Alternatively, there may have been two migration events: first *erectus* and secondly *sapiens*. Following the sapien migration, there may have been total replacement by sapiens (*replacement hypothesis*). There may also have been interactions and gene flow between sapiens with *erectus* or *Neanderthals*. All these models are supported by substantial amounts of fossil and archeological evidence [116]. However, studies of mitochondrial DNA from *Neanderthals* show a high degree of variation with *sapiens*, suggesting that *Neanderthals* did not contribute to the variation in contemporary mtDNA. However, this cannot completely rule out gene flow to nuclear DNA. Recently, it has been hypothesized the *MAPT* gene, which shows an extreme haplotype structure and codes for Alzheimers disease protein tau, may have been derived from *Neanderthals* [152]. Therefore, nucleotide and haplotype diversity can be useful tools for tracing the potential origins of genes that are important in common diseases.

7.2 Medical Genetics

The synergistic relationship between population and medical genetics has become fundamental for locating genes influencing phenotypic traits and understanding the origins of disease alleles. By comparing the human genome to those of other species, such as *Pan troglodytes*, can provide information about the genes and genetic variation that have been important in the evolution and development of *Homo sapiens* in terms of cognitive abilities and language [153]. We can also improve our understanding of how climate and environmental exposures can influence genetic variation and gene function.

The majority of the 0.1% differences between two individuals have presumably no function and little influence on fitness. However, a subset of the variation is important for susceptibility to disease. These diseases may be caused by rare novel mutations, for example in cystic fibrosis, Huntingtons disease, Tay Sachs and muscular dystrophy. However, not all diseases have discrete inheritance patterns in families. In fact, the most common diseases, such as cardiovascular disease, respiratory diseases, and psychiatric diseases, are sporadic. Furthermore, these diseases are more frequent in the western world, which suggests that common polymorphisms and environments are contributing to disease. Therefore, to find the causes of genetic disease across populations, we need an understanding of the range, structure, effects and sharing of natural and disease causing genetic variation in the human genome both within and between populations.

Chapter 8

Mendelian Disease

Mendelian Disease Around the time Mendels work on independent segregation and assortment was rediscovered, Garrod was observing trends of "inborne errors of metabolism" (alkaptanuria) segregating in families following similar patterns of dominant and recessive inheritance. Mendelian genetic diseases are considered simple given that they have a readily identifiable pattern of inheritance, typically caused by mutations in a single gene.

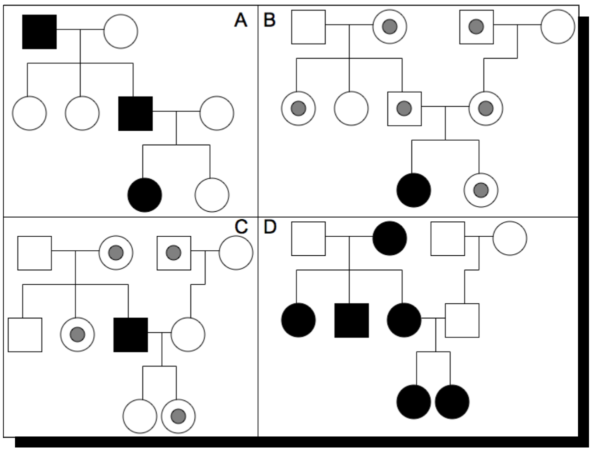


Figure 8.1: Filled Squares:- Male, Diseased; Filled Circles; Female, Diseased; Spots:- Muation carriers, not diseased. (A) Dominant Inheritance pattern (B) Recessive (C) X-linked (D) Mitochondrial

A *dominant* inheritance pattern is observed when affected individuals (probands) occur in consecutive generations (vertical), independently of gender, and are caused by the transmission of one disease allele that sufficient to cause disease (dominant). Examples of dominant diseases include Huntingtons disease, myotonic dystrophy and variegate porphyria. A *recessive* inheritance pattern is also gender independent, but two copies of the disease allele are required to cause the disease. Therefore, the disease phenotype may only be observed in one generation (horizontal) or alternative generations. Examples include cystic fibrosis, spinal muscular atrophy and phenylketonuria. The transmission patterns of mutations on the sex chromosomes are different from those on the autosomes. Affected males with an X-linked dominant mutation only inherit the disease in the maternal line while affected females can receive the disease allele from either parent, as examplified by hypophosphatemic rickets. As males carry a Y-chromosome, recessive X-linked mutations transmitted from mother to son will be dominant, e.g. color blindness, Duchene Muscular Dystrophy and heamophilia. Likewise, mitochondrial diseases are typically transmitted in the maternal line and all offspring are affected e.g. Lebers optic atrophy.

8.1 Linkage Analysis

Genetic Epidemiology and Medical Disease Genetics are born from the concept that alleles increasing the risk for disease are transmitted in pedigrees and populations. However, disease genes are not easy to identify. The allele contributing to a particular disease phenotype could be in any of the 30,000 genes in human genome. Early methods focused on markers available at that time, such as the ABO blood groups [154]. With knowledge of the genetic paradigms, improvements in technology and the identification of genetic markers, the process of locating disease genes has been built on two platforms known as forward and reverse genetics. *Forward genetics* relies on knowledge of disease mechanisms. Consistent alterations in protein activity or structure in affected individuals are taken as indications of mutations in the genetic code. The amino acid sequence could be used to determine the DNA sequence and the hybridization of probes to isolated clones can be used to locate the gene. The clones can subsequently then be sequenced to identify the mutations. However, for the majority of inherited diseases the molecular basis remains unknown and alternatives requiring no knowledge of disease mechanisms have been designed.

Reverse genetics takes advantage of deviations from Mendels law of independence: *linkage*. Linkage refers to the close physical proximity of loci on a chromosome, which are transmitted together from parent to offspring (*co-segregation*) but may be separated by recombination (*independence*). As the probability of recombination between two loci increases with distance, recombination is essentially a function of distance. This principle is the foundation of *Linkage Analysis*: by monitoring marker co-segregation through pedigrees and determining the number of recombinants (recombination fraction: θ) in order to locate the disease gene. Distinguishing which marker allele is transmitted from each parent is essential and informative, highly polymorphic markers, such as microsatellites, are most often used.

Morton devised the LOD score method (logarithm of the odds) to test the hypothesis of linkage or independence between markers [155]. The higher the odds of observing the allelic combination for a given theta against the odds of the same data under complete independence was utilized to indicate linkage between the markers. The maximum log_{10} of the odds (likelihood ratio) gives the Z-score, which can be used to test the significance of the linkage. In such studies $Z = 3$, indicates that linkage is 10^3 times more likely than no linkage with a probability of 10^{-5} which is significant at the 10^{-4} threshold.

Since recombination is a function of distance, the number of meioses is crucial for disease gene identification. As the number of generations per family is quite limited and the number of observed meioses small, the resolution of linkage studies can be quite poor. The disease gene could be anywhere in a broad region of 5-10cM, containing hundreds of other genes. The LOD method was designed to be a sequential test meaning that data from a number of different pedigrees could be combined for each given recombination fraction. The overall maximum score was used to determine the location of the markers in relation to each other. Although this may help in narrowing down the position of the disease gene, the resolution remains low at approximately 1cM. This method of linkage analysis has formed the basis for locating genes by other methods such as the multilocus method and the MOD score [156].

Chapter 9

Complex Disease: Lessons from Linkage Analysis

In 1989 the recessive mutations in the gene *CFTR* that cause cystic fibrosis were located [157]. This was a discovery heralded as the first real success for linkage analysis. Since then 1833 phenotypes have been described where the molecular basis of the disease is known(*OMIM*, January 31st, 2006). However, in the course of applying linkage analysis to a range of diseases, Mendelian and common, a number of issues have arisen concerning the genetic component and architecture of disease i.e. number of genes and mutation frequencies.

9.1 Is there a Genetic Component to Common Disease?

Most of the mutations that contribute to monogenic disease are deleterious and occur within the coding regions of structural or enzymatic proteins. These mutations tend to have a high *penetrance* i.e. a strong one-to-one relationship with the disease, and are not influenced by the environment. They may be presented congenitally or at an early age, thereby influencing fitness. Because the balance between mutation and selection is tipped, the mutations are quickly lost from the population or are kept at low frequencies (less than 1%) including disease incidence. As recessive mutations tend to have a smaller effect on fitness, they may have higher allele frequencies than dominant mutations.

Some genes involved in common diseases such as breast cancer (*BRCA1*, *BRCA2*) [158] and Alzheimers (APP) [159] have been identified by linkage analysis, primarily due to early onset, relatively high penetrant alleles that follow a Mendelian mode of inheritance. However these cases make up only 5-10% of all the disease incidences, while the remainder occurring spontaneously in families who may not have had a prior history of disease. Therefore, the application of the traditional linkage based method becomes more difficult as the inheritance pattern becomes less obvious due to incomplete penetrance of disease mutations [160]. Furthermore, the phenotype may manifest itself at a post reproductive age (*late onset*) and subsequently previous generations could be difficult to collect for linkage analysis. Late onset suggests weaker effects on selection and so the mutation may be subject to random genetic drift. The alleles could be lost immediately or increase in frequency, which could render descendents susceptible to the disease.

The contribution of genetics to a common disease may be ascertained from measures of “relative similarity”. *Heritability in the narrow sense* is the proportion of the phenotype variance due to additive genetic effects (V_A/V_P), which represents the extent to which the phenotype resemblance is determined from transmitted genes [161]. Additionally, λ_s measures the ratio of risk for disease in siblings of the proband as compared to the population prevalence. The increase in risk between siblings may indicate the segregation of genetic factors contributing to the risk. The risk can be observed for all possible degrees of relatedness and should decrease with distant relatives [162].

Estimates of phenotype heritability and risks indicate that common diseases are *mulifactorial*: both genetic and environmental components contribute to disease susceptibility. However, the heritability of various diseases and phenotypes is a summation of the genetic contribution and not a quantification of the number of genes the type or of effects made by each mutation. Some mutations or polymorphisms may impose a greater individual effect and risk than others and collectively increase susceptibility. It may also be the case that a large number of genes, each with a minor increase in risk could contribute to the disease (*polygeneic*) [160, 163, 164].

The existence of a genetic component to a number of late-onset, common diseases has also been questioned [165, 166]. Common diseases such as obesity and cardiovascular disease may have a large amount of influence from the environmental but have little genetic influence (*phenocopy*). Lifestyle changes such as a lack of exercise and a greater consumption of high calorie and fat foods may be contributing to higher incidence rates [167]. Environmental pathogens such as viruses may also contribute to common diseases such as Schizophrenia. For example, children born during the winter appear to have a higher risk due to *in utero* infections affecting development of the nervous system [160].

Therefore common diseases that exhibit late onset, unknown modes of inheritance, unknown genetic or environmental composition and unknown genetic architecture are labeled ***Complex Diseases***.

9.2 The Genetic Architecture of Common Complex Disease

Mendelian genetic disease architecture can be intricate due to the fact that, different mutations in the same gene (*allelic heterogeneity*; *CFTR* and cystic fibrosis) or different genes (locus *heterogeneity*; Retinitis pigmentosa and hearing loss) may cause the same phenotype. Moreover, a single gene may cause a range of different diseases (*pleiotrophy*; *ABC1*-Tangeir Disease, Familial hypoalphalipoproteinemia). Multiple genes may be needed for disease, such as mutations in the tyrosinase and *MITF* genes for digenic Warrdenburg syndrome type 2.

However, the genetic architecture that underscores the genetic contributions to complex disease remains largely unknown. Some population genetic models propose an *oligogenic* or a *polygenic* structure, which implyies that either relatively few genes with moderate increases in risk or that many genes each with a small contribution to risk are involved. The *Common Disease/Common Variant* hypothesis (*CD/CV*) proposes that high frequency predisposing polymorphisms in a limited number of genes contribute to disease susceptibility [154, 168]. Each polymorphism has a limited effect on fitness and the risk of disease. Therefore, it is expected that polymorphisms may have drifted into high frequencies either prior to or during population expansions, and could be shared among various populations because of population history. As a result, the expected allelic identity for the disease mutations is high, i.e. the diversity in disease alleles is quite old but has not changed very much [169].

It is possible that alleles with low selective constraints may have become fixed in the population. In this way, alleles would not contribute to heritability but they would be important for disease susceptibility. During human evolution alleles beneficial for metabolism and storage of energy became fixed, but in modern society food is readily available, and these alleles may not be so advantageous (The *Thrifty Gene hypothesis*) [170]. The *Common Disease/Fixed Variants* (*CD/FV*) cannot be ruled out [171], given that such alleles fixed in different populations were observed in the HapMap Project [4].

The *Common Disease/Rare Variant hypothesis* (*CD/RV*) proposes that relatively new mutations, with relatively weak effects on the fitness of the carrier, collectively increase risk for disease. These mutations may only last a few generations due to random genetic drift [172, 173]. According to such a model, the high turnover of mutations leads to a diverse allelic spectrum for the disease, causing the allelic identity to be low. The ability to find these mutations is limited unless extensive re-sequencing is carried out to discover the variation. Other models take into account the effects of selection on rare mutations [169]. The "*mutation accumulation*" model predicts that mutations accumulate in the genome and affect aging (senescence) and subsequently age-associated diseases. Alternatively, "*antagonistic pleiotrophy*" suggests that mutations may have beneficial effects early in life but deleterious effects later [171].

Understanding the genetic architecture of complex disease could be instrumental in identifying high and low risk disease groups. A study modeling the population risk of breast cancer, based on high frequency low risk alleles, demonstrated that a polygenic multiplicative model best fit the familial aggregation of non-BRCA1/2 breast cancer. By identifying the population proportion above or below a given risk that explains a percentage of all cases, shows how important it is to identify the genetic risk factors. For example, if all the factors were identified, 2% the population with a risk of 11% for breast cancer before 70 years of age would account for 50% of the cases of breast cancer [164]. To date many studies indicate the a small number of loci may contribute to most of the risk in diseases such as Diabetes (TCF7L2, CTLA-4) [174, 175]. Therefore the knowledge of the genetic architecture is important to aid the identification of genes, which could result in improved prevention, diagnosis and prognosis.

Chapter 10

Identifying Complex Disease Genes

The characteristics of complex disease reduce the power of linkage studies to identify the causative regions of the genome; therefore alternatives to the conventional parametric approach have been developed. To overcome the lack of individuals from previous generations and specification of a genetic model, *non-parametric* linkage analysis, avails of genotype sharing between two individuals. The *Affected Sibpair* (ASP) linkage method measures the relationship between sib phenotype similarity and the proportion of alleles shared between the sibs; *identical by descent* (IBD, alleles have the same common ancestor). In other words, if an allele is truly involved in the trait then it should, in theory, be shared by those affected. The probability that full sibs share 2 alleles IBD is $\frac{1}{4}$, 1 allele IBD is $\frac{1}{2}$, and none IBD is $\frac{1}{4}$. The hypothesis is that a greater proportion of alleles linked to the disease locus will be shared between the sibs, i.e IBD $> \frac{1}{2}$, and can be useful for locating disease alleles. Although sib pairs are most commonly used, methods have been developed for other degrees of relationship and extended pedigrees. However, disadvantages are that the resolution of these linkage approaches remain low and some issues with sample collection remain [176].

Alternatively, *Tests of Association* performed at the population level are proposed to be more powerful than linkage studies for finding complex disease genes [177]. The principle behind the association study is if the risk factor (disease allele) is found in or transmitted to affected individuals more than in unaffected individuals, then that allele could possibly be contributing to the disease or phenotype. Tests for association are tests of *identity by state* (IBS), where the causative allele is assumed to have arisen from a common ancestor among the cases and or controls but is too far back in time to be determined and the segregation pattern (inheritance vector) cannot be resolved [178].

The most commonly used test of association is the *case-control* (Figure 10.1). This is a relatively simple epidemiological method where *exposure* to the risk factor (allele/genotype) is compared between a group of unrelated individuals with the disease (cases) and a group without the disease (controls) from the population. Ultimately, if the exposure is contributing to the risk of disease, then it should be over-represented in the cases. If the exposure is independent of the disease then it should be found equally in both groups. As the individuals are collected based on disease status and exposure to the risk is examined, these tests are often termed *retrospective studies*. Conversely, in a *prospective study*, individuals may be collected in a cohort study and separated into groups based on the presence or absence of exposure itself (alleles/genotype). Over time, the incidence rate of disease is compared between the two groups and the difference is related to the exposure to the risk factor.

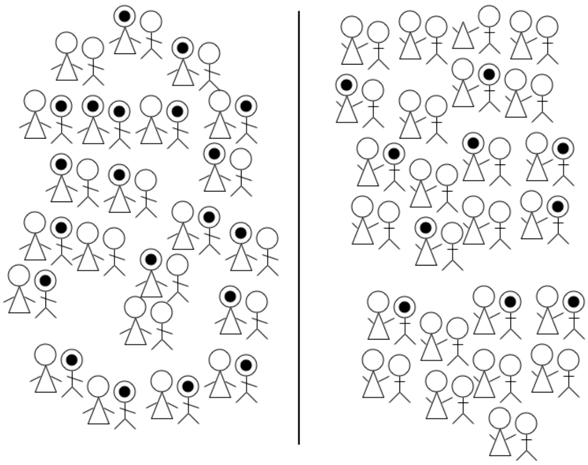


Figure 10.1: Case-Control study: To the left are Cases, to the right are controls. The polymorphism is more frequent in the cases than in controls

One of the main differences between a prospective and a retrospective study is sample size because not everyone in the exposure groups develops the phenotype. As a result, many more samples are required for analysis. However, this may not be the case if the outcome is common such as depression, or cardiovascular disease. Nonetheless, case-control studies are often the researchers choice because they are cheaper and faster.

In the application of these methods to the location of disease genes, the strongest association will be found if the disease allele is tested directly. However, finding the location of the disease allele among the estimated 11 million SNPs across the genome poses a dilemma. Testing all the SNPs in hundreds to thousands of individuals is not feasible by existing technologies and would be a nightmare statistically. To narrow down the search for the disease allele, contemporary approaches have been structured on the function of polymorphisms, genetic disease architecture and population genetics.

Contemporary approaches can be categorized into two forms: *Direct* and *Indirect Association* [179]. Direct Association is based on the *causal hypothesis*, which assumes that if all coding SNPs are discovered, catalogued and tested, the polymorphisms contributing to complex disease will be amongst them [93]. Similar to the reverse genetic design to Mendelian disease, the *Direct Whole Genome Association* (WGA) envisions genotyping all coding-polymorphisms or a prioritized subset across the entire genome to find the polymorphisms causing complex disease. However, cSNPs tend to be quite rare, not very diverse and the magnitude of their effects may be unknown. Consequently, methods are being developed to determine the contributions of cSNPs to function in order to prioritize SNPs for genotyping in collected samples [69]. Perhaps the greatest concern of the Direct Approach is the limited gene variation that is used. Without concern for polymorphisms beyond the coding region, such as promoter or splice-sites, true causative SNPs could are be missed.

The Indirect Association attempts to address the concerns of the Direct Association by assuming that the genotyped polymorphism (*marker*) is in LD with the causative SNP (*Proximity*

Hypothesis) [93]. For this reason, the indirect association is also called *LD mapping*. The disease allele is assumed to occur against a background of SNPs forming a disease haplotype. At this point the disease polymorphism is in LD with all the SNPs that occur on that background. If selection against the polymorphism is relatively negligible then the haplotype is subject to drift, potentially rising into high frequencies. In the course of this time, LD will decay due to recombination and the extent of correlations around the polymorphism will get smaller and narrower (Figure 10.2). If association is found, the assumption is that the marker is either the disease allele or in LD with it. Application of this ideology avoids the necessity to genotype all known functional polymorphisms and has the potential to locate unknown contributory polymorphisms.

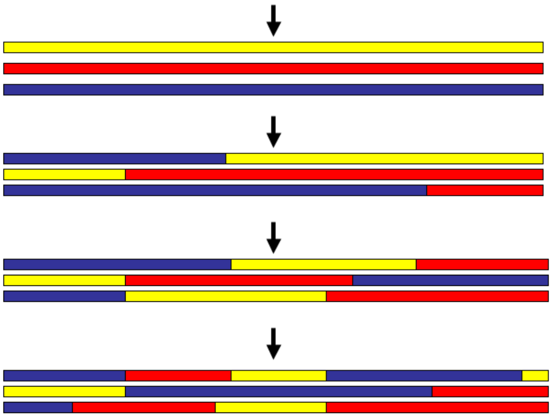


Figure 10.2: Break down of LD. Linages are in different colours. Recombination shuffles them an narrows down region around the mutation (arrow)

Both the direct and indirect association studies have the potential of whole genome application. However, a more in-depth focus can be made by selecting *Candidate Genes* based on the biological mechanisms involved in the development, actions and treatment of the disease (*Hypothesis-Driven Candidate*) Alternatively genes can be chosen based on regions linked or associated previously with the disease (*Positional Candidate*). These principles can be applied to both the direct and indirect associations with the potential to improve power for finding disease genes

10.1 Selection of Candidate Genes

To date, the *Candidate Gene* approach has been the most frequently applied method for either the direct or indirect association. Less restrictions by high throughput technologies and budgets, make the candidate gene approach appear to be an straightforward alternative to WGA. However, selecting candidate genes for study is a daunting task, particularly when the genetic architecture of the disease is unknown. The random selection of gene would not be very efficient, and such a method is best suited for WGA. Instead, there are means of prioritizing candidate genes that could improve the likelihood of finding disease alleles.

Advantage can be taken of *positional candidates* in regions previously identified in linkage or association studies [180]. Mendelian disease genes are sometimes characterized by mutations that factor in severity of the phenotype and onset. In the absence of highly penetrant variants, these genes may also contain polymorphisms with weaker effects and still contribute to complex disease [163]. Direct Associations may find polymorphisms in the particular gene significant and may select this gene for further study. Alternatively, linkage and association may identify extensive regions containing a number of genes that could possibly be involved in the disease.

The identification of plausible *hypothesis-driven candidates* with functional relevance to the disease can narrow down the selection of genes genome wide or in regions of previous linkage or association [181]. If the molecular disease or treatment mechanisms are somewhat understood, genes can be targeted for their impact on the phenotype. For example, many of the genes studied in neurological disorders have been identified by treatments targeting the gene products such as transporters, receptors, membrane channels and neurotransmitter metabolizers.

A great deal of our understanding of mechanisms behind disease has come from modeling the phenotypes in the laboratory. *In vitro* studies have been very useful for understanding the molecular mechanisms of genes and their cellular functions. Of equal importance are *in vivo* studies using model organisms such as worms, flies, mice, rats and primates. For example, a gene of interest can be "knocked out" and effects on the organism can be monitored through development into late adult age. On the other hand, genetic variation in animal models may be controlled to produce a congenic strain where the only source of variability is in the gene of interest. Phenotypes can then be monitored for the contributions by the gene and its different variants. An advantage of model organisms is the ability to produce extensive pedigrees, increasing the number of recombinations and thereby narrowing the region of interest. This is quite useful for reducing allelic and locus heterogeneity. A model state reflective of the human state may help identify the strong contributors to the disease that are shared between the species. This strategy is useful for also identifying disease genes with strong effects and is thereby well suited for follow up studies to the findings of *CD/CV* hypothesis [182].

Through *in silico* comparisons of the human genome to those of other species, such as mouse, rat and chimpanzee, chicken and cattle, not only are we gaining more insight into the evolution of the human genome but it is becoming possible to identify candidate genes for diseases such as Bardet-Biedl syndrome [183], renal failure [184] and arthrosclerosis [185]. Additionally more is being learned about the characteristics of disease genes [184, 186, 187]. For instance, the majority of human disease genes are conserved in most species, and appear to have faster evolutionary rates than non-disease genes. Neurological and malformation genes appear to be under negative selection, while immune, heamatological and pulmonary genes appear to be positively selected [186].

10.2 Genotyping Technologies

The prospects for candidate gene and genome-wide associations in hundreds to thousands of samples requires the development of technologies for SNP detection and allele discrimination with greater high-throughput capabilities (100-fold) and increased sensitivity/specificity, while decreasing the cost per genotype. SNP genotyping is built on three components: biochemical reaction principal, assay platform and detection method. A range of platforms with the same principle can be matched to different detectors in order to identify the SNP, thereby providing various approaches to the discrimination of alleles. The common reaction principles in use today are:

1. Hybridization
2. Oligonucleotide Ligation
3. Primer Extension
4. Enzymatic Cleavage

Hybridization

Besides DNA sequencing, hybridization was the earliest form of SNP detection [188], taking advantage of base-pair specificity and the instability of miss-paired nucleotides. Complementary probes that differ by one base pair to either allele of the SNP, known as *Allele Specific Oligos* (ASO), are hybridized to the target sequence forming a duplex. Mismatched probes are not as stable as matching probes and less energy is required to disrupt the duplex.

Early applications for detection commonly used Southern blot procedures [189], the dot blot [190] or reverse dot blot platforms. Improvements in technologies and detectors have allowed monitoring of duplex stability over an increasing temperature range. The signal from an intercalating dye, which only fluoresces when a duplex is formed, will be lost at a lower temperature upon melting of the mismatched probe, thereby permitting discrimination of the allele. *Dynamic Allele Specific Hybridization* (DASH) is an example of a method that relies on this principle [191]. Alternatively, probes can be labeled with flourophores that do or do not emit light when in the proximity of a quencher molecule (*Fluorescence Resonance Energy Transfer*: FRET). For example, when a probe is bound to its specific target, DNA polymerase cleaves the fluorophore from the probe, which separates it from the quencher and permits the emission of light. This can be used as both a quantitative and qualitative SNP assay and is exemplified by the TaqMan™[192] and Molecular Beacon [193] approaches.

The stability of hybridization methods can be improved by using Locked Nucleic Acid (LNA) [194] or Peptide Nucleic Acid (PNA) [195] to increase the stability of the probe. Alternatively, a minor groove binder (MGB) is bound to the TaqMan probe [196]. Multiplexing hybridization reactions is difficult but high throughput capabilities have been developed for allele specific hybridization using micro-array assay platforms such as GeneChip by Affimetrix and the semi-homogenous solution DASH-2 [197]. TaqMan and Molecular beacons have limited capabilities for multiplexing, but as they have less handling and no separation step, they are more suitable for analysis of large sample sets [198].

Oligonucleotide Ligation

The principle of OLA reaction is based on the joining of two hybridized oligonucleotides adjacent to the SNP site using the specificity of DNA ligase. Only when the correct base is incorporated, will the ligase join the oligonucleotides [199]. A number of detection systems for this method are available [200, 201]. One variation of this approach is the Padlock Probes, which works in the same way as the OLA with the exception that the probes are joined at opposite ends by a stretch of DNA. Upon ligation, circular DNA is formed, which can subsequently be amplified (rolling circle amplification, RSA) [202]. The Molecular Inversion Probe (MIP) is an adaptation of RSA where an endonuclease site is placed between the primer sites common to all probes. Probes that do not ligate are degraded by exonuclease, leaving only circular strands. Endonuclease is then used to splice the circle. A unique tag sequence is used to hybridize the probe to an array-platform, where alleles can be discriminated [203]. RSA and MIP overcome the limitation of multiplexing PCR reactions by performing the PCR after SNP detection, reducing difficulties in multiplex design, and producing enough tag sequences for both qualitative [204] and quantitative [205] high throughput genotyping.

Primer Extension

There are two separate principles of primer extension. First, there is the amplification of PCR products using allele specific primers; the second involves the incorporation of single-nucleotides and utilizing fluorescence or reaction chemistry to discriminate the SNP alleles.

One primer of an allele-specific primer pair has a complementary nucleotide to either of the SNP alleles at the 3' end and the DNA polymerase will only carryout extension when there are matches. Subsequently, fluorescent probes can be utilized for allelic discrimination. Alternatively, gel-based or capillary-based electrophoresis systems are used to separate the amplicons and visualize the outcome. However, optimization for each SNP is required and the throughput is limited to a few SNPs at a time.

Allele-specific nucleotide incorporation also utilizes base pairing but the extension reaction is limited to 1-50 bps. The concept behind "minisequencing"is the specific incorporation of fluorescently labeled dideoxyribonucleiceotides (ddNTPs) [206] detected using DNA-sequencing instruments. The ratio of fluorescence determines the allelic state of the SNP. Alternatives to fluorescence detection methods are numerous. One such approach is based on hapten labeled ddNTPs which can be detected using ELISA reactions [207]. Another approach to detection is Matrix-Associated-Laser Desorption Time-of-Flight Mass Spectrometry (MALDI-TOF), which separates the extension reactions based on the molecular mass of the products [208]. The process of Pyrosequencing uses pyrophosphate, which is released when a nucleotide is incorporated at the 3' end of a primer, as a substrate for luminescence by luciferase [209]. Each nucleotide is added sequentially and light is quantified to determine the allelic state of the SNP and surrounding sequences up to approximately 50-100bps. This technique is now a basis for modern DNA sequencing technologies.

The multiplexing of mini-sequencing reactions has been successful on a microarray format with either one or two detection primers [210, 211]. Some approaches employ primers with specific tags at the 5' ends that can be captured on arrays [212] or microbeads [213] with complementary tags. Further advancements in multiplexing have been developed that address the issue of PCR limitations in multiplexing for primer extension methods. The Golden Gate assay attaches biotinylated fragments of the genome to avidin microparticles. Specific hybridizing primers bind and DNA polymerase extends the primer. These primers contain sites for further hybridization in their 5' end. One of these primers contains a tag site compatible with GeneChip arrays or Bead arrays. The products are amplified by PCR with complementary primers, one of which is labeled, and then tagged to the arrays permitting signal detection. Modifications such as enzymatic allele-specifc primers using fluourescent nucleotides [214] or hapten ddNTPs [215] can be used to increase the multiplexing capability of this assay.

Enzymatic Cleavage

The cutting specificity of restriction nucleases was first utilized in the development of cloning techniques and was later recognized as a method for SNP detection in the form of *Restriction Fragment Length Polymorphim* (*RFLP*). Restriction enzymes will only cleave double stranded DNA if the sequence at the cleavage site is correct. In doing so, PCR amplified DNA will be cut into 1+N pieces (N= number of specific cleavage sites).

A more advanced cleavage technique is the invasive approach, which uses a probe that is complementary to the allele but the 5' end is not matched to the upstream target sequence. When the probe matches the allele, in the presence of an invader probe, the 5' end forms a structure that is recognized by a FLAP nuclease and cleaved. This will only function when the alleles match, without which, a different 5 structure is formed and that goes unrecognized by the endonuclease [216]. The cleaved product can be detected in a number of ways including FRET; however, for the purpose of large scale multiplexing, it has been adapted to solid-phase arrays [217].

Chapter 11

Issues in Association Mapping

The ability to locate alleles that contribute to disease (*power*) depends on the magnitude of the contribution of each allele and the degree of LD between markers and disease alleles. For this reason, the finding of disease genes using direct or indirect approaches may be problematic, which could limit the use of these strategies and lead to misinterpretation of results. Improvements in the design of and the approach to an association study have common goals: *Maximization* and *efficiency*. The objective is to reduce costs while increasing power or limiting the loss of power thereby minimizing the chance for obtaining false positives. The genetic architecture of complex disease has a significant role in the power of association studies. The idea of studying common variation and LD in the population suits the *CD/CV* hypothesis very well and many successes have contributed to the optimism of this approach such as *APOE* (Alzheimers disease) [218]), *PPARG* (Type 2 Diabetes) [219], ADAM33 (asthma) [220], NOD2 (Crohns disease) [221], LTA (myocardial infarction) [222] Factor *V*(deep vein thrombosis) [223]. However, the actual utility of common variation needs to be addressed once the low hanging fruit have been picked. Important alleles with weaker effects need to found. This is especially relevant to the *CV/RV* hypothesis.

11.1 Linkgage disequilibrium

The strongest associations will be found if the disease polymorphism is tested directly. For an indirect association, the strength of LD between the marker/haplotype and the disease polymorphism will determine whether an association is found. If the polymorphisms are not in perfect LD ($r^2 = 1$), then the association signal will be weakened or "diluted" This may be due to the choice of markers, or htSNPs, that do not represent the branch of the genealogy on which the disease mutation occurred or maybe due to recombination. It has been shown that studying common polymorphisms and inadequate selections of htSNPs may not be very powerful for locating disease alleles with low frequencies and risks [224]. On a finer scale ($<$ 2kb), gene conversion could be influencing LD more than anticipated. From the HapMap Project, approximately 20% of the ENCODE SNPs genotyped in Europeans and Asians are not in perfect LD with SNPs adjacent to them. Three in five SNPs have 5-19 SNPs in perfect LD, while one in five have over 20. These numbers are even lower for the Yorubas as seen by the average r^2 with the best SNP on the map: 0.85 *vs.* 0.67. Therefore the number of samples will need to increase in order to compensate for the loss of signal, or additional coverage of the region may be warranted. Having the HapMap actually will be useful for such purposes. The HapMap may help researchers identify SNPs in the region of their study that may be in LD with their markers, which could aid to strengthen signals, and improve efficiency [225]. Phase II of the HapMap project will result in a denser map of 4.2 million common SNPs, which will be even more useful for finding disease alleles and understand extent of crossovers.

11.2 Sample homogeneity

The issues with LD become even more problematic when the magnitude of the effect by the disease polymorphism is not maximized in the sample. The ideal situation would be *sample homogeneity*: an enriched sample where a minimal amount of variants render the cases liable to disease. Again this may suit the *CD/CV* hypothesis, but will be critical for issues such as allelic and locus heterogeneity, phenocopies, epistasis, gene-environment interactions and low allele frequencies (*CD/RV*), which could all be factoring in the variation of the phenotype [226]. For this reason, the definition and classification of the phenotype is very important in the choice of sample to be collected. The power of an association study could be improved by selecting cases based on sub-phenotypes (*endophenotypes*), extreme states of the disease, or highly discordant/concordant sib-pairs. Isolated or inbred populations have been useful for locating rare disease genes and can be important for reducing the heterogeneity in complex disease allele architecture. Twins offer a unique resource for the controlling of genetic and environmental factors. Due to sharing of genetics and/or environments, the genetic contributions to disease can be estimated independently of the hypothesized allelic architecture [161, 227]. Furthermore, correlations between twins with the same alleles contributing to the disease can give a strong indication of the magnitude of the disease allele. Phenotype maximization will also increase the efficiency of an association study by reducing the sample size and thereby lowered costs and time.

11.3 Genotyping and Multiple Testing

When power is compromised, the probability of spurious findings increases with the number of tests performed. This is a central issue in whole genome studies and of dense candidate gene studies. For example, if one was to extrapolate the SNPs required to genotype the ENCODE regions to the entire genome, approximately 600,000 individual assays would need to be performed. Assuming there is 90% power to detect the 150 SNPs causing the phenotypes at a significance threshold of α at 0.05, then 99% of the SNPs that test positive for association will be false positives. For this reason, stringent thresholds are needed to reduce the *Type I error rate*. Assuming the same conditions but a more stringent $\alpha = 0.0001$, the percentage of false positives will drop to 31% and the positive predictive value of the test increases. An initial *p*-value of 5x10^{-8} has been proposed to control for such false results [177].

A number of tests have been developed to determine whether or not significant results are true positives or not. One of the most frequently used methods is the *Bonferroni Correction*, which divides the significance of each positive by the number of tests performed. One drawback of this is that it assumes all tests are independent and in association mapping this is not the case due to LD between markers and therefore Bonferroni correction is very conservative. For example, the *p*-value of 5x10^{-8} presented above is equivalent to a *p*-value of 0.05 after correction for 1 million independent tests [177]. Although other correction methods have been developed to deal with correlations between SNPs [228], their use has not been extensively investigated [229, 230]. Permutation testing of the case-control samples in order to produce a distribution of significance values is one alternative against which the *p*-values of the original findings can be compared [231]. The *False Discovery Rate* (FDR) takes into account the proportion of expected false positives in a study in comparison to the chance of finding any false positive. This approach to multiple corrections is becoming more widely used in association studies [232].

False positives may also be introduced due to technical errors or mishandling of samples, leading to the wrong assignment of individual genotypes. The assay conditions could also be variable and lead to a bias in allele scoring. The conditions of the *Hardy-Weinberg Equilibrium* (HWE) test make the assumption that alleles and genotypes are assorting independently of each other. Genotyping errors may be observed by deviations in the observed frequencies of genotypes from that expected with the allele frequencies. Error is often measured by repeating the analysis in a subset of the samples and/or by comparison to an alternative genotyping technology.

11.4 Population Stratification

Although population choice may be very useful for minimizing disease heterogeneity, population stratification may also lead to false results in association studies. Stratification arises from substructure within populations or the recent "admixture" of different populations. Within these subpopulations, there may be differences in the prevalence of the disease or phenotype. If cases and controls are selected without knowledge about the existence of substructure and allele frequency differences between the subpopulations due to drift, then these polymorphisms will show spurious associations with the phenotype although there is no real effect by the polymorphism. Furthermore, LD between polymorphisms may misdirect the location of disease alleles. Someways to detect and control for stratification have been proposed. These include genotyping polymorphisms not linked to the sites or looking for population specific alleles in the groups [233, 234]. Cases and controls should be carefully matched from the same population. However this may not always be appropriate [235]. Family-based associations can be used to control for the problems of stratification. The *Transmission Disequilibrium Test* (TDT) examines the segregation of chromosomes from parents to offspring. This eliminates the issue of parent origin. However, this test can suffer from issues with genotyping, as tests may often be biased towards conditions that suit the genotyping of particular alleles. Additionally, questions have arisen concerning transmission distortion, which might also affect these associations [236].

11.5 Study Replication and Functional Validation

Confidence in an association is a major concern for complex disease genetics. Given the high probability for false positives, replicating significant findings is of importance and can be approached in a number of ways. It is of primary interest is to repeat the analysis using an independent sample from the same population, thereby strengthening the hypothesis that the polymorphism has a potential role in the disease phenotype in that population (239). This may be difficult if the disease or sub-phenotype is not so common or the population is small. Attempts at replicating significant findings are most often performed in different populations and if successful will further suggest that the polymorphism is involved in the disease [237]. Complications arise when the results are not repeated and confidence in the trustworthiness of the original findings is undermined.

Literature reviews illustrate a bias towards positive findings and many follow up studies fail to repeat the significant associations of the initial publications [238]. This could indicate that many of the first studies contain type I errors. In addition, inconsistencies in study design, sample size and homogeneity, phenotype classification and population choice all factor [239] in follow-up studies, which may often turn out to be under powered. However, there are clear examples where published results have been verified such as Factor *V* in deep vein thrombosis (Arg507Gln), *CTLA*-4 in Graves Disease (Thr17Ala), *APOE* in Alzheimers Disease (ϵ4) and *PRNP* in CJD (Met29Val) [238]. These consistencies would indicate that polymorphisms in these genes contribute to modifications on disease risk in a number of different populations in support of *CD*/CV.

As a number of studies are underpowered, there are probably missed associations (*False Negatives*/Type II errors), especially for alleles with low frequencies and/or low penetrance. Meta-analysis combines data from a number of studies as a way to increase the sample size, and thereby potentially increasing power.

The analysis of gene function for significant associations is important for the validation of findings. Understanding the action of particular alleles is key to endorsing the biological role of the polymorphism in disease. The approach to functional validation depends on the location of the polymorphism. Examining the affinity of the corresponding protein to ligands, substrates, inhibitors, agonists and interactions with other proteins can assess the effects caused by polymorphisms in the coding region of the gene. Comparing the expression of cloned polymorphic regulatory elements via reporter vectors is a common way to examine promoter variants. Effects on RNA isoforms, levels and structure are explored for splicing variants or variants that could potentially influence RNA stability. Other methods for analyzing regulatory elements are under development such as *Haplo*ChIP, which examines the levels of phosphorylated RNA polymerase II bound for different alleles using chromosome immuno-precipitation (ChIP) and mass-spectrometry [240].

Functional studies play an important role in instances where association is found for multiple SNPs in LD. For example, if 20% of the common SNPs are in perfect LD, the ability to distinguish the real disease polymorphism will be important [241]. Functional assays provide the possibility to assess which polymorphisms are contributing to the phenotype. For example, the Thr17Ala polymorphism at *CTLA*-4 shows consistent associations, but is in LD with a regulatory polymorphism in the 3'-UTR that correlates more strongly to the phenotype [242]. Other situations have arisen where promoter SNPs are in LD but have opposite individual functional effects, which demonstrates the complexity and importance of studying haplotypes and combined SNP effects [243, 244].

11.6 Benefits of Re-sequencing Candidate Genes

Primarily due the high cost, sequencing has been one of the last steps in identifying alleles contributing to disease. However, as sequencing is both de novo discovery and an *a priori* detection tool, it is instrumental for the classification and tracking down of the genetic contributions to disease. Initiatives have been taken by the National Human Genome Research Institute (NHGRI) to entice improvements in methods for minimal cost Whole Genome Sequencing. As sequencing platforms and technologies improve towards high throughput capabilities at lower cost, the re-sequencing of regions of the genome, including candidate genes, acts in two ways: a detection method of known polymorphisms and a discovery tool for unknown genetic variants. As the genetic architecture of many common diseases is relatively unknown, re-sequencing is a way of simultaneously tackling both *CD*/*CV* and *CD*/*RV* hypotheses and a number of the issues with association studies.

Earlier examination of candidate gene sequence has provided insights into the patterns of candidate gene variation for cardiovascular disease [245, 58], HIV [246], and origins of Alzheimers disease alleles [247, 60]. By sampling a large number of individuals, we can gain insight into the distribution of genetic variants across different regions of the gene, genome, populations and species, which can tell us about the population history and forces such as mutation rates and crossovers that affect the disease gene. With *CD*/*CV* in mind, this level of detail increases our knowledge of polymorphism correlations. Recombination and gene conversion can be separated using approaches based on diversity, and together with a better understanding of the true polymorphism density, an improved selection of htSNPs for LD mapping approaches can be made.

Comparing the levels of diversity between disease genes and non-disease genes can benefit the *CD*/*RV* hypothesis. Departures from the neutral model might indicate the type of effects mutations have on the survival of the carrier and reasons for low frequencies. Perhaps the comparison of candidate gene diversity between cases and controls, or parents and offspring, will give insights into the germline mutation rates of these genes, which may be variable. This could be particularly important for the examination of allelic heterogeneity in proposed *CD*/*CV* diseases and even of greater importance to the *CD*/*RV*.

The identification of new alleles will be of great benefit to the *CD*/*CV* and *CD*/*RV* hypotheses. Knowledge of the population diversity and polymorphism correlations based on re-sequencing will be critical for the interpretation and validation of association studies where the alleles may differ or be at different frequencies in different populations. Additionally, re-sequencing of both DNA strands can help in the confirmation of rare variants such as singletons or doubletons and reduce genotyping errors. Ultimately, re-sequencing of genes will play a significant role in the maximization and efficiency of genetic association studies.

Chapter 12

Positional and Hypothesis-driven Candidates Genes

12.1 Hypothesis Driven Candidates of the Serotonergic System for Depression and Obesity

The monoamine neurotransmitter serotonin (5-hydroxytryptamine) was first purified from serum in the late 1940s and is so named for its effect on the tone of blood vessels [248]. Serotonin was later found to be expressed in the brain [249]. In the synthesis of 5-hydroxytryptamine (5-HT), dietary L-tryptophan is hydroxylated to a 5-hydoxytryptophan intermediate (5-HTP) by tryptophan hydroxylase, which is subsequently then decarboxylated to 5-HT by amino acid decarboxylase. Serotonin is catabolized by monoamine oxidase and aldehyde dehydrogenase to 5-hydroxyindoleacetic acid (5-HIAA).

The diffuse serotonergic modulatory system has one of the smallest numbers of neurons in the central nervous system (CNS) (1 in 1 million) [250] and mediates its actions thought the vast number of projections (5x10^{5}mm^{-1}). There are two main groups of nuclei that express 5-HT: the inferior and superior raphe nuclei restricted to the basal plate of the pons and the medulla [251]. The inferior nuclei innervate the spinal chord whereas the superior nuclei project into the midbrain and the forebrain [251, 252].The superior nuclei are composed of the dorsal raphe nuclei (DRN) and the median raphe nucleus, while the inferior group is composed of nucleus raphe obscruus and the nucleus raphe pallidus. Median raphe projections are more abundant in the hippocampus. The DRN also projects to the hippocampus via the cingulated cortex, but DRN projections are more abundant in the cortex and striatum [253].

DRN innervate the paraventricular nucleus of the hypothalamus and the intermediate lobe of the pituitary. The serotonergic system is also connected to other neurotransmitter systems, such as dopaminergic neurons in the substantia nigra, noradrenergic neurons in the locus coeruleus and GABA neurons in the hippocampus and midbrain. Through innervations to the *hypothalamus-pituitary-adrenal axis* (HPA) and the limbic system, serotonin plays a central role in the homeostasis of cardiovascular regulation, respiration, emotion, gastrointestinal system, circadian rhythm, aggression, cognition, learning, memory, appetite and mood. Perturbations in the components of the serotonergic system, such as synthesis, response, reuptake and degradation have effects on the system, either early or late into neuronal development. Such perturbations influence 5-HT synaptic plasticity and increase vulnerability to anxiety, depression, aggression, sleep disorders and appetite. In behavioral, mood and psychiatric disorders, alterations in the baseline levels of 5-HT are known to be very important. For example, in depression, reduced 5-HT synthesis or reduced levels in the synaptic cleft are noted. Inefficient turnover of 5-HT in the synaptic cleft can lead to antisocial behavior and aggression. Therefore, knowledge of these components, their functions and which factors influence them and when, is important in detecting the causes of a number of neurological disorders [254].

This is complicated by the scope of 5-HT innervations and by a number of system components including the large number of different 5-HT receptors. The magnitude of effects on neuronal plasticity during stages of development and neurological disease pathogenesis can be monitored by targeting the components in animal models. Most of what we know about 5-HT stems from the treatment of disorders and the knock out of components in mice. Therefore a number of serotonergic candidate genes have been identified that potentially influence an individuals susceptibility to behavioral and neurological disease.

12.1.1 Serotonin receptor 2C

The serotonin receptor 2C is one of 14 receptor subtypes which mediate the actions of 5-HT [255]. It is expressed in many areas in the CNS such as the hippocampus, nucleus accumbens, amygdala dorsal striatum and the substantia nigra but particularly abundent in the choroids plexus [256]. It is a seven transmembrane G-coupled receptor [257] and it is expressed postsynaptically [256]. The secondary signaling mechanism employs phospholipase C (PLC) to initiate a phosphoinosital second messenger cascade producing protein kinase C (PKC) and phospholipase C (PLC) which stimulates the release of calcium and the activation of protein kinase C (PKC).

The receptor is expressed in the paraventricular and ventromedial nuclei of the hypothalamus [258] and has been considered a candidate for mediating the role of serotonin in the HPA and may ultimately influence appetite and response to stress. To date, the best example of the role of the receptor in modulating feeding behavior comes from a *htr2c*-knockout mouse model, in which a stop codon was introduced in exon 5 [259]. The mutant mice exhibited elevated weight compared to wild-type mice due to abnormal eating behavior. After 5-6 months, these hyperphagic mice displayed a substantial increase in body weight coinciding with increases in white adipose tissue, hyperinsulinemia, hyperglycemia and leptin resistance [260]. Weight gain and pre-diabetic conditions were more rapid in high-fat diet fed mutant mice. As these results were observed in the late stages, hyperphagia was considered as the main effect and the increase in weight was due to a lower intake of oxygen and energy expenditure in older mutants [261].

The induction of hyperphagia in *htr2c*-knockout mice complements the observations that serotonin receptor 2c drug antagonists also induce hyperphagia [262, 263]. Conversely, the *HTR2C* agonists *m*-chlorophenylpiperazine (*m*CPP) [264], 1-(2,5-dimethoxy-4-iodophenyl)2-aminopropane (DOI) and lysergic acid diethylamide (LSD) [265, 266] reduce hyperphagia. Fenfluramine is a drug that both inhibits the reuptake and stimulates the release of serotonin. A reduction in meal size, eating rate, and an increase in intervals between meals are observed in normal mice treated with fenfluramine [267]. When *d*-Fenfluramine was marketed as a treatment for obesity (Redux®), food intake was decreased in obese subjects [268, 269]. However due to cardio-toxicity, the product was withdrawn (272).

D-Fenfluramine and *m*CPP have bee shown to stimulate the increase of 5-HT in the hypothalamus, which induces an anorexic effect, is similar to the effect of direct microinjection of 5-HT into this region. Therefore this serotonin receptor is considered to induce effects on appetite by increasing ACTH. It has also been found to activate accurate neurons expressing the melanocortin precursor pro-opiomelanocortin which gives rise to ACTH [270]. Therefore HTR2C is a strong candidate for involvement in hyperphagia.

A consequence of treatment with anorexogenic drugs that increase the circulation of ACTH and vasopressin [271] is the elevation of the levels of cortisol and prolactin, which play a significant role in stress and anxiety (anxiogenic). Other disorders such as depression are associated with stress and elevated cortisol levels [272]. It has been proposed that htr2c may be involved in stress [273] and may be up-regulated in depression [274]. Consistent with these studies, *htr2c*-KO mice are less anxious [275], which is supported by pharmacological evidence that *HTR2C* antagonists are anxiolytic [276][277]. Likewise *m*CPP increases depressed mood and tension in patients with depression [278]. Evidently, the serotonin transporter inhibitor, fluoxitine has a similar affinity to 5-HTR2C, where it exerts an antagonistic effects [279] and other antidepressants have been found to influence the response of mice under the swim test [280].

A unique feature of the X-linked 5-HT2C receptor [281] is the editing of its pre-mRNA in which adenosine residues are converted to inosines by double stranded RNA adenosine deaminase. The coding of amino acids is altered and the resulting structural change in the second intracellular loop effects the affinity to G proteins and the activity [282, 283]. The distribution of these receptor isoforms is altered in the brains of suicide victims with a history of major depression. This does not occut in those treated with fluoxitine [284]. These lines of evidence suggest that *HTR2C* is a valid candidate for depression.

Linkage of bipolar disorder to Xq24-25 [285] and the discovery of a common *G-C* transversion (*Cys*23*Ser*) in the third exon of the human *HTR2C* gene [286] sparked a large number of studies into the role of *HTR2C* genetic variation in susceptibility to a range of affective, behavioral and psychiatric disorders and the corresponding responses to treatments. Although the *Ser* allele is expected to have no effect on function [287], studies were initiated under the premise that the polymorphism was functional or in LD with the functional variants. However, studies on affective disorder [288] biopolar disorder [289, 290, 291], depression [292] obsessive compulsive, BMI [293], bulimia nervosa [294], binge eating [294] are either inconclusive or negative. One study found the *Ser* allele to be significantly associated with a higher rate of anorexia in a cohort of teenage girls [295]. Similarly, other associations studies have shown that the *Ser* allele correlates with reduced hypophagia effect of the response to *m*CPP [296]. Findings of associations of the conserved promoter microsatellite polymorphism with panic disorder [297] and bipolar disorder [291, 298] were not confirmed.

However, the promoter SNPs -995*G*>*A*/-759*C*>*T* were found to be associated with obesity and type two diabetes. Different haplotypes also showed variability in expression [299]. Smaller weight gain was observed with the -759 *T* allele in the use of clozapine [300, 301, 301, 302] ,while the *C* allele was found to associate with obesity and heterozygotes lose the least amount of weight [303]. Thus, important associations may have been missed previously due to the unknown LD between SNPs in the promoter and the *Cys*23*Ser*.

12.1.2 Monoamine Oxidase A and B

The monoamine oxidase enzymes MAOA and MAOB are located on the outer membrane of mitochondria [304]. They function in the oxidization of biogenic and xenobiotic amines but have distinct neurotransmitter and inhibitor specificity. Clorgyline is a drug that inhibits the deamination of 5-HT and norepinephrine by MAOA, while MAOB preferentially deaminates benzylamine and beta-phenylethylamine but is inhibited by deprenyl. Both enzymes contain a covalently bound flavin adenine dinucleotide co-factor (FAD) for activity [305].

MAO isoenzyme expression is differential in the CNS but seems conserved between humans and primates. MAOA is found in catecholamine containing cells like substantia nigra, nucleus coerulus, and PVN of hypothalamus. MAOB is expressed in serotonin regions: DRN, histaminergic neurons and astrocytes [306, 307]. Beyond the CNS, MAOA and MAOB are both expressed in lymphocytes and liver [307] whereas MAOA is expressed in the placenta [308] and MAOB in platelets [307].

MAOA and *MAOB* are transcribed by separate genes [309] located on the X-chromosome, Xp11, [305], and oriented in a tail to tail fashion [310]. *MAOA* is over 70kb and has 15 exons where exon 12 houses the FAD domain and potentially has two different polyadenylation sites [311]. *MAOB* is 50 kb away from *MAOA*, has a similar exon-intron structure as *MAOA* but is almost twice as large (115kb) [312]. Little variation has been reported for the two genes, and sequence screens indicate that the patterns of known MAO variation result from a bottleneck [313] and/or selection [314]. Because of the similarity in gene structure and 70% homology in amino acid sequence, the MAOs are believed to have arisen from the duplication of a common ancestor potentially of bacterial origin [1]. The genes have different promoter structures possibly reflecting their cell and tissue specificity [315].

The normal range of platelet-MAOB and skin fibroblast-MAOA activities are highly variable and MAO activities are 20% higher in females than in males. The function of the two deaminases is also influenced by environmental exposures such as smoking, which lowers the activity [316]. Nonetheless, platelet MAO activity does seem to have a large heritable component ($\sim$0.77) as indicated by a number of family and twin studies [317, 318].

Given the actions of these enzymes in the CNS, the determinants of MAO-activity has been of central interest for understanding the role of different neurotransmitter systems in neurological disorders, including depression. Early studies on serotonin turnover indicated a positive correlation between platelet-MAO activity and the 5-HT metabolite 5-HIAA in normal individuals. However, correlations with a depressive phenotype have been inconclusive [319]. Platelet activity appears to be high in depressed individuals, especially females, suggesting the involvement of other factors that influence 5-HIAA levels in the cerebro spinal fluid (CSF).

Although platelet and frontal cortex MAOB have the exact same amino acid composition [320], there is no clear correlation between their activities. Likewise, there is no correlation between the activity of MAOA in skin fibroblasts and MAOA in the CNS of the same individuals.. The use of monoamine oxidase inhibitors (MAOIs) in the treatment of depressive disorders increases 5-HT in the synapses [321]. It has been proposed that both lipid environment and common genetic factors underlie differences in MAO activity between the CNS and peripheral tissues.

A Dutch kindred with aggressive and violent behaviors has been described, where the affected have a defect in urine MAOA metabolites, low 5-HIAA levels, but normal platelet-MAOB activity [321]. A point mutation in exon 8 of *MAOA* has been found to terminate translation of the fully functional oxidase [322]. In a *MAOA*-knockout mouse model increases in 5-HT and noradrenalin were observed along with anxiety and aggression. 5-HT synthesis inhibitors reverse these symptoms. This potentially reflects the importance of *MAOA* in serotonergic development.

It has been hypothesized that low MAO-activity early in life and exposure to stressful events could render people vulnerable to behavioral and psychiatric disorders [323]. This hypothesis was perhaps best exemplified by Caspi [324], who showed that a significant number of children maltreated early in life that grew up to have anti-social behavior had a low expressing variant of the *MAOA* promoter tandem repeat [325]. It is therefore possible that high-MAO activity, such as that seen in depression could affect behavioral development. Studies have demonstrated that *MAOA* is not X-inactivated [326], while other studies have shown that a *C*pG island near the VNTR in intron 1 of *MAOA* is methylated on an X-inactivated chromosome [327]. Considering that X-inactivation skewing with age appears to also have a heritable component [328], this could possibly explain MAO activity increases with age and increases in vulnerability to depression. Therefore the monoamine oxidases are considered candidate genes for depression.

12.2 Hypothesis Driven and Positional Candidate Genes For Alzheimers Disease

Progressive memory loss, cognitive decline and an increasing necessity for daily assistance are clinical characteristics of dementia and Alzheimers disease (AD). Characterization of neuropathological lesions in the AD brain known as plaques and neurofibrillary tangles [329] identified the accumulation of the amyloid precursor protein (APP) cleavage product, amyloid-β (Aβ), and tau protein, respectively [330]. Determining the mechanisms of Aβ synthesis contributed to the identification of mutations in the APP gene [159] as well mutations in genes required for cleavage presenilin-1 [331] and presenilin-2 [332]. These mutations have high penetrances with an early onset of AD (EOAD) and follow a Mendelian mode of inheritance. However, familial AD (FAD) comprises less than 5% of all AD cases and the majority of patients have late onset AD (LOAD) with no clear mode of inheritance.

Alipoprotein E, which is believed to be involved in the clearance of Aβ_{40} and the insoluble form of Aβ_{42}, is the only gene that repeatedly shows strong associations with LOAD. Alone, it does not cause AD but the isoform $\epsilon4$ influences the risk for the disease in an additive manner. In other words, those with two copies of $\epsilon4$ have a greater risk for AD at an early age than those with one and no $\epsilon4$. Considering that the prevalence of AD rises steadily from 2.8% per 1000 people between the ages of 65-69 years to 56% per 1000 over 90 years of age [333, 334] and that a 50% percent increase is expected in people over 65 by 2025, determining more components in the etiology of AD is critical for efficient early diagnosis, prevention and therapy.

The insulin-degrading enzyme (IDE) has been considered to play an important role in the progression and severity of Alzheimers disease. Expression of IDE was first observed in the in the liver, testes, muscle and brain. Its primary substrate is insulin [335] and a decrease in IDE activity is considered to be a major contributor to hyperinsulinaemia and pre-clinical Type II diabetes (T2D) as indicated in the model mouse for T2D, GK, which has mutations in the ide gene [336]. However IDE has also been found to interact with Aβ and function as one of the main degraders of Aβ [337]. In the brain, extra-cellular soluble fractions of the AD abundant Aβ_{40} and Aβ_{42} are degraded by IDE, but in the hippocampus of AD brains the of degradation of Aβ is reduced. Interestingly, in a diet induced model of T2D, increases in gamma secretase activity and Aβ concentrations are observed in the presence of lowered IDE activity and concentration (347).

An important feature of the IDE-KO mice is a 50% increase in Aβ [338]. Because insulin and Aβ are competing substrates for IDE, it has been hypothesized that hyperinsulinaemia or elevated Aβ saturates IDE, or that decreased IDE activity could contribute to the accumulation of Aβ and vulnerability to Alzheimers Disease.

If has been shown that IDE activity decreases with age and that individuals who develop T2D have twice the risk to develop Alzheimers disease compared to those who do not. Insulin levels are is elevated in Alzheimers patients, and this also correlates with cognitive impairment. A limitation in locating the genetic contributions to LOAD is sample collection. However, a number of genome linkage scans have identified regions of chromosome 10q, including the *IDE* gene, as regions potentially housing genes involved in the susceptibility to AD. Two linkage peaks on chromosome 10q have been observed for *APOE* ϵ4 positive and *APOE* ϵ4 negative, respectively [339]. Subsequent studies show specific associations 195kb from *IDE* [340], while others found maximum linkage peaks proximal and distal of *IDE* [341], one of which is a quantitative trait locus for plasma Aβ [342].

The allele specific effects observed in the linkage scan were replicated in a case contol setting in which *APOE* ϵ4 status had a modifying effect [343]. Follow up studies showed LD to be strong across *IDE*. However, associations with IDE SNPs and LOAD were only observed when tests controlled for *APOE* ϵ4 [344, 345, 346]. Subsequent analysis of a more dense set of markers demonstrated that *IDE* and two other genes *KIF11* and *HHEX*, were encased in an extensive LD block of approximately 276kb. Certain haplotypes showed protective (*H2*, *H5*) and predisposing (*H1*, *H4*) associations with quantitative traits, such as the Mini-Mental State Examination (MMSE), CSF-tau and age of onset. In an independent sample the effects of these haplotypes were replicated with LOAD and Aβ plasma levels. Based on linkage analysis and associations the genes *IDE*, *KIF11* and *HHEX* in the haplotype block identified by Prince et al (2003) and prior functional analysis IDE is considered a candidate gene for increasing the risk for Alzheimers disease.

Chapter 13

Present Investigations

13.1 Aims of this Thesis

This thesis serves to demonstrate the importance of re-sequencing both hypothesis-driven and positional candidate genes as a strategy for enhancing our knowledge of gene variation patterns and for improving association study design. To exemplify this, the following goals were set:

1. Re-sequence the *HTR2C* promoter and intragenic regions in order to understand the relationship between polymorphisms utilized for direct and indirect association studies. To use the sequence data and patterns of variation in determining the forces which could have shaped these relationships

2. Carry out a SNP discovery effort in the promoters of the Monoamine oxidase *A* and *B* genes, which show low levels of variation, and validate polymorphisms obtained from dbSNP in the Swedish population

3. Identify polymorphisms in a haploblock, linked to and associated with Alzheimers Disease, through comparatively re-sequencing the promoters, exons and conserved regions of three genes: *IDE*, *HHEX*, and *KIF11* in Alzheimers patients and controls

4. Utilize the polymorphisms of these genes and their LD in association studies with disease phenotypes;

    - *HTR2C*: Obesity and Depression
    - *MAOA* and *MAOB*: Depression
    - *IDE*, *HHEX*, *KIF11*: Alzheimers Disease

13.2 Methods

13.2.1 Sequencing

The Sanger dideoxy-terminator method was used to obtain DNA sequence reads from PCR products. Overlapping amplicons were designed to span the gene or region of interest and optimized PCR reaction conditions eliminated the requirement for product purification. PCR reactions were performed in parallel using a 96-well format that allowed direct transfer of amplicons to DNA sequencing reactions following inspection on agarose gels.

High throughput separation and detection of sequence reads was performed on the MegaBace™1000. This is a 96-capillary instrument with a FRET-based recognition system using an argon laser. Each ddNTP is labeled with flourescein and one of four other rhodamine dyes (rodamine 110, rodamine 6G, rodamine X or tetramethyl rhodamine). The maximum excitation wavelength of fluoresceins is at 488nm, at which it transfers energy to the flourophores for detection through four separate channels by two photomultiplier tubes (PMT).

For base calling and assembly of DNA reads, the Poly/Phred and Phrap software were used, respectively. Consed was used to visually inspect the sequence reads and polymorphisms were determined from deviations from the consensus sequence in a read. Each PCR product was sequenced in both directions for the purposes of polymorphism identification and validation.

13.2.2 Genotyping

In papers I and III, SNP detection was performed using Dynamic Allele Specific Hybridization (DASH). This is a PCR-based hybridization assay that utilizes a biotinylated primer to bind double stranded PCR products to streptavidin-coated microtiter plates. A washing step with sodium hydroxide destabilizes the double strand structure and only the single strand bound by biotin remains. An allele specific probe is subsequently allowed to hybridize to the single strand. Sybr Green 1, an intercalating dye used to detect the presence of a double stranded structure over a temperature range of 35°Cto 85°C. The loss of fluorescent signal indicates that the probe has melted from the template, and this will occur earlier for mismatches than for full complement matches. The negative derivative of this signal is determined to illustrate the distinct peaks for analysis of the allelic state. Following washing steps to clean the bound PCR product, the process can be repeated with a probe specific for the opposite allele, which permits determination of the genotypic state of the SNP.

In paper II, Pyrosequencing was the genotyping platform for allele detection in the SATSA (Swedish Adoption /Twin Study of Aging) sample set. This is also a PCR based method with either reverse or forward biotinylated primers. PCR products are bound to streptavidin-coated sepharose beads and washed with NaOH for denaturation, which renders them single stranded. A primer is annealed adjacent to the SNP of interest and incubated with a reaction mixture containing DNA polymerase, ATP sulfurylase, firefly luciferase, and apyrase and each dNTP is added cyclically. Only if the correct dNTP is present will DNA polymerase elongate the primer by one nucleotide. Pyrophosphate is released from the added nucleotide, which is then used to convert adenosine 5' phosphosulfate (APS) to ATP by ATP-sulfurylase. Luciferase utilizes the ATP in the production of oxyluciferin from luciferin and in the process emits light. A charge-coupling device is used to detect the signal is which is relative to the amount of ATP and determines the SNP allele and adjacent nucleotides.

TaqMan ®Allelic Discrimination was chosen for genotyping a larger twin sample set in Paper IV for the study of *HTR2C* and *MAOA*/*B* polymorphisms in depression. The principles of the TaqMan assay are also based on hybridization. However, in comparison to other methods in the same category, TaqMan is a homogenous reaction that requires no washing steps, which increases high throughput capacity. The reaction is PCR-based and takes advantage of the 5 exonuclease activity of Taq DNA polymerase. During the production of amplicons, probes bind specifically to their complementary alleles. These probes are labeled with a 5-fluorophore (FAM or VIC) and a 3'-non-flourescent quencher (TAMRA)/-minor groove binder complex. If the specific probe is stably bound, Taq will cleave the 5'-flourophore from the probe permitting it to fluoresce. Both probes are present in the reaction and the fluorescence ratio is used to determine the genotype of the SNP.

13.3 Results

13.3.1 Paper I

The *HTR2C* mouse knockout model expressing a hyperphagic-obese phenotype, as well as the identification of a *G* to *C* transversion nsSNP (*rs*6318, *Cys*23*Ser*) in the third exon, spurred a large amount of studies of serotonin receptor 2Cs role in modulating food intake and associations with eating disorders. Work on *HTR2C* agonists and antagonists in animal models continued to reveal significant effects. However, association studies failed to demonstrate, with conviction, that *rs*6318 is directly involved in disorders and it was considered that other polymorphisms in other regions of the gene were contributing to the effect in such disorders. Studies published around the beginning of this study indicated that promoter polymorphisms could be important in obesity related disorders.

Approximately 1,282 bp of the *HTR2C* promoter and 5,223 bp of intragenic sequence, encompassing coding exons one and two, were sequenced in 64 males resulting in 1.16 Mbp of sequence for analysis. A greater degree of nucleotide and haplotype diversity was found in the promoter than in the intragenic region. Neither region indicated a deviance from the standard neutral model. The analysis of haplotype diversity did suggest potential balancing selection or population substructure in the intragenic regions. This could not be observed in the promoter, recombination could serve to increase haplotype diversity, thereby potentially removing the signs of balancing selection or substructure. As recombination is an important factor in the design of association studies, this possibility was explored further.

The Four Gamete Test indicated that recombination had occurred in the proximity of *rs*6813 and between the regions but not in the promoter. Conversely, the population recombination parameter was elevated in the promoter while very low across the other regions. Although $\mid D' \mid$ demonstrated strong to perfect LD, r^2 gave a better indication of potential crossover events such as gene conversion in the promoter. When the background rate of recombination is low, gene conversion over small distances has been shown to elevate C. This will be reflected in r^2 because of the relationship with the population recombination parameter. Therefore gene conversion and recombination were considered to be a driving force in shaping nucleotide, haplotype and LD diversity in *HTR2C*.

Common haplotypes also showed variability in expression of a luciferase reporter gene. The second most frequent male haplotype showed significantly lower expression than the most frequent. At position -697, which is located in a strong transcription start site, the *C* allele appeared to decrease expression in comparison to the wild-type haplotypes. Variation in the expression of this promoter could be influential in *HTR2C* related functions and disease; however, the power to detect associations with these promoter polymorphisms using *rs*6813 as a marker is likely to be weak due to recombination and gene conversion.

Concerning obesity, no associations were found with high BMI $\geq$ 30kgm^{-2} for any of the promoter SNPs at the allele, genotype and haplotype levels. However, when the microsatellite data was included, associations and increased risks for obesity were observed for the haplotype *TA13GG* (OR = 4.42, CI = 2.25-8.68, 2.25, *P*<0.0001), and the diplotype *TA13GG*/*TA16GG* (OR = 5.8, CI = 2.3,14.6, *P* = 0.0006), which are relatively rare in the control population (1%). Individuals heterozygous for promoter SNP -995*G*>*C* were associated with high serum-leptin levels%body fat (Mean Serum = 0.663±0.358, *P*=0.03). Diplotypes heterozygous at -995*G*>*C* and/or the microsatillite exhibited a similar trend of elevated serum-leptin/%Body Fat. This was most noticeable for *TA13GG*/*TA16GG*, however was insignificant after Bonferroni correction.

13.3.2 Paper II

The function of MAOA and MAOB in the central nervous system, correlations between thrombocyte-MAO activity and number of behavioral and psychiatric disorders, and their targeting in the treatment of these diseases, suggest that *MAOA* and *B* genes potentially participate in susceptibility to neurological disorders. However, few polymorphisms have been associated with such diseases with conviction. The lack of variation across the genes complicates the validation of previous *MAOA* and *B* associations with disease states is. Approximately 4.5 kb were sequenced from the promoter region of each gene to determine the extent of genetic variation in the Swedish population. Additionally, 12 SNPs from dbSNP, and two previously reported in the Swedish population were selected for genotyping and validation in the sample set.

With a sample size of 148 X-chromosomes, the power to find SNPs with a frequency between 1% and 3% was calculated to be 77% to 100%, respectively. Even with enough power to find SNPs with a frequency under 1%, little variation. No SNPs were found in the *MAOB* promoter, while three were found in an intronic region from the *MAOA* gene. One of the *MAOA* polymorphisms was previously documented (*rs*3788863) and selected for further study. Surprisingly, of the 14 SNPs selected a priori for validation, six were monomorphic in the sample subset. This included two SNPs from introns 3 and 10 of *MAOB* that had been previously found by resequencing in a Swedish sample group. Of the SNPs that were polymorphic, heterogeneity was observed in the minor allele frequency of *MAOB* SNPs, which was reflected in the haplotype and LD structure. Conversely *MAOA* SNPs were similar in frequency, had stronger allelic correlations and only two haplotypes pdominated the sample set.

Gender differences were observed in trbc-MAO activity. Males and smokers showed significantly less trbc-activity, while females with depressed state had much higher activity. Examination of the gender stratified data, revealed that *MAOA* SNP *rs*979605 genotypes *C*/*C* and *C*/*T* were associated with a significant decrease in trbc-acitivity in females (-2,9; CI 95%: -5,2 -0,6 and -2,4; CI 95%: -4,7 -0,1 respectively). Depressive state was associated with the *A*-allele of *MAOB* SNP *rs*1181252 in males (OR = 4,5; CI 95%: 1,0 21,7) and both *GG* and *GA* of *rs*766117 (OR = 2,2; CI 95%: 1,1 4,3) in females. No associations were observed with *MAOB* haplotypes and trbc-MAO activity, while a decrease was associated with two *MAOA* haplotypes, *A1* and *A3*. Haplotypes from neither gene associated with depressive state but female

MAOA/B homozygotes showed an increase in risk.

13.3.3 Paper III

Fifteen individuals from Sweden and 15 from Scotland were selected for SNP discovery on whether they carried protective haplotypes *H2* or *H5*, or high-risk haplotypes *H1* or *H4*, which span the haploblock containing the three genes *IDE*, *KIF11* and *HHEX* on chromosome 10q. In total 48 variants were found, the majority of which were SNPs. One sixth were insertion/deletion polymorphisms. No LD analysis was performed on the sequencing data. SNPs weighted on the biological relevance of the genes to Alzheimers disease, were selected (9 from *IDE*, 4 from *KIF11*) for genotyping in Alzhiemers disease cases (N = 121) and controls (N = 152).

One predominant haplotype accounted for approximately 60% of the *IDE* haplotype diversity, while 2 haplotypes explained 70% of the haplotype variation in *KIF11*. LD analysis indicated that the genotyped SNPs were in strong LD, however r^2 yielded contrary results. This may be due to the rarity of some of the SNPs selected for genotyping. Suggestive associations were observed with two polymorphisms in *KIF11* 20 and 32. However after Bonferroni correction these observations were no longer significant.

13.3.4 Paper IV

Following the results in Paper I and Paper II, selected SNPs from the *HTR2C* promoter and the *MAO* loci were genotyped in a large sample composed of three sub cohorts of the Swedish Twin Registry (Swedish Adoptee/Twin Study of aging (SATSA), GENDER and OCTO), to test for association with depressive state (N = 1563).

The distribution of allele frequencies and patterns of LD were similar to those observed previously in Paper I and II. The llele frequencies were consistent in *MAOA* and *HTR2C* but variable in *MAOB*, which was reflected in the LD pattern. $\mid D' \mid$ was strong across the regions, but r^2 was more complex, especially for *HTR2C*. Not surprisingly, little LD existed between *HTR2C* and the *MAO* loci, which are located on the opposite arm of the X-chromosome. Interestingly, the *MAOA* and *MAOB* SNPs indicated a trend towards HWD from the 5' to the 3' ends of each gene, This has been suggested to reflect deviations from neutrality, and more recently potential indel copy polymorphisms.

No associations were found with any SNP, genotype and haplotype in the male sample set. A suggestive, but not significant, association was observed for the *HTR2C* promoter SNP *rs*498207 (*G*, *P*= 0.061) between cases and controls. This appeared to confer an increased risk for depressive state in an additive manner. Haplotype analysis suggested significant deviations in frequencies between cases and controls especially for *HTR2C-GGCC* (*P* = 0.028) and *MAOA-CT* (*P* = 0.033). Although the risk for depressive state increased with each allele of *HTR2C-GGCC*, it was marginally insignificant (OR = 2.33 *P* = 0.06).

The haplotype effects were studied further by taking the correlations between twins into account using a GEE model and comparing the risk per haplotype allele to the increased risk of homozygotes over heterozygotes. These should complement each other if there are additive effects. Suggestive, but insignificant, additive effects were again observed again for *HTR2C-GGCC* (*P* = 0.06). Consistent with Paper II, a trend toward significance was observed for

MAOA-CC per allele effects for increased risk of depressive state. However, there were no differences between homozygotes and heterozygotes.

13.4 Discussion

13.4.1 Re-sequencing

Empirical studies have demonstrated active roles for the serotonin receptor 2C in obesity and depression, the monoamine oxidases *A/B* in depression and the insulin-degrading enzyme in Alzheimers disease. However, the identities of possible polymorphisms in these genes that contribute to the risk of disease remain difficult to uncover and validate with confidence. Inconsistencies between studies may include, study design, sample selection, sample size, SNP selection and SNP density. To optimize the power of locating disease polymorphisms, an understanding of nucleotide variation and nucleotide correlations at the population level is paramount. Without this information, inappropriate polymorphisms that do not represent the genealogy may be selected as markers and associations may be missed or falsely identified. Therefore, this thesis has focused on regions of hypothesis-driven and/or positional candidate genes to demonstrate the utility of re-sequencing for improving our understanding of nucleotide variation and thereby enhancing association study design.

Re-sequencing the candidate regions revealed differences in the rates of nucleotide polymorphisms between each of the genes studied. Approximately 26 SNPs were observed in the *HTR2C* promoter and intragenic region with little difference in allele frequency (Paper I). However, different types of polymorphisms with variable allele frequencies were found among the 48 *IDE*, *KIF11* and *HHEX* variants (Paper III). In contrast with these genes, the *MAO* regions studied showed very little variation (Paper II). A number of factors can limit the identification of novel polymorphisms. One of the biggest issues is the power to discover polymorphisms $(1-(1-q)^N)$, which is influenced by sample size (N) and allele frequency (q). Polymorphisms that have higher frequencies ($\geq$5%) are easiest to find owing to their large heterozygosity and the reduced probability of two persons having the same allele. Therefore, smaller sample sizes are suitable for discovery of higher frequency polymorphisms. However, the SNP allele distributions show that there are more SNPs with minor alleles with frequencies lower than 5%, than SNPs with minor alleles over 5%. Consequently, many polymorphisms are missed in the re-sequencing of small samples.

Although different samples were used for re-sequencing regions of *HTR2C* and of the haploblock encompassing *IDE*, *KIF11* and *HHEX*, the sample sizes were similar (60-64 chromosomes, Paper I, III). Both studies therefore had similar power to find SNPs of a given frequency. Due to a larger sample size, the power to find SNPs at the same frequencies or less by re-sequencing was enhanced in the *MAOA/B* regions (Paper II). However, there was a polymorphism deficit in the *MAOA/B* regions compared to *HTR2C* and the haploblock. This lack of polymorphism has been noted previously [347, 313] and may be indicative of a population bottleneck or positive selection. Under such models, a number of rare variants would have been expected to occur. The lack of variation in our sample could suggest that many of the *MAOA/B* SNPs documented in dbSNP are sample or population specific, which is further supported by the low validation rate of SNPS (42%) in our sample. Because population and/or selective forces have a large effect on the frequency of nucleotide variation, they not only influence the discovery of SNPs, but are a significant factor in SNP detection and therefore important instrumental for the power to find polymorphisms that are potentially useful for association studies or find the polymorphisms that are important for disease etiology. Therefore, in-depth re-sequencing of genes in different populations will be important in determining the forces that have shaped these patterns of nucleotide diversity.

As gene history has an influence on the frequency of polymorphisms, the forces that shape the genealogy will also impact on the LD between SNPs. Crossovers, influence the power of an association study by breaking down LD between markers and potential disease SNPs. Understanding and identifying these factors that create or breakdown LD are therefore critical for direct and indirect association studies where it is assumed that the SNP being genotyped is the causative SNP or potentially in LD with the marker.

The FGT showed that recombination had occurred between the SNPs identified by re-sequencing of the *HTR2C* promoter and intragenic region (Paper I). Both of these regions contain SNPs that have been used in association studies with serotonin receptor 2C related phenotypes. To date, the $Cys23Ser$ SNP in the third exon has been the most commonly used marker because 1) it causes an amino acid substitution and 2) the minor allele is of high frequency permitting ease of detection using a number of genotyping methods. However, a number of recent studies suggest that promoter polymorphisms are possibly important for the influence of *HTR2C* in response to treatments and weight gain. In Paper I, the promoter polymorphisms affected expression in a luciferase reporter assay. Therefore, if promoter SNPs do factor in *HTR2C*'s contributions to disease, reduced LD between the regions could be a factor in false negative associations or replication studies.

The LD structure is further complicated by gene conversion in the promoter. Studies in Chinese and Japanese populations have previously noted that promoter SNPs -759*C*>*T* and -995*G*>*A* are in strong LD, while the adjacent promoter SNPs -697*G*>*C* and -759*C*>*T* do not. This indicates that gene conversion may have occurred before ancestral migrations to Asia. This could be tested by extending the studies to other global populations and genotyping additional promoter SNPs. Recently, copy-number variation, including gene conversion, has been demonstrated to be an important factor in the interpretation of SNP associations and a necessary consideration for the study design of such studies. However, from the study of the *HTR2C* promoter, $\mid D' \mid$ is probably not an appropriate estimator to detecting gene conversion. r^2 gives a better impression of the event in this sample and since it is a function of the population recombination parameter, it is influenced by the events that shape the genealogy. However, as the length of a gene conversion tract can be quite small, higher SNP density is required, which may not often be the case, and consequently, they may be overlooked. As a result the SNPs selected may not represent the genealogy of that region and potential associations could be missed.

Given that linkage and association studies have identified 10q22 and the haploblock spanning *IDE*, *KIF11* and *HHEX* as potentially housing polymorphisms contributing to the risk of Alzheimers disease, a novel polymorphism rate of 54% in the promoters, exons, flanking and conserved regions demonstrates the utility of re-sequencing genes in regions of strong LD for the discovery of unknown variants that could influence susceptibility to disease. The elevated number of novel polymorphisms in this study may be due to the use of SSCP to discover polymorphisms in *IDE* previously [344] but could be also due to the inclusion of individuals with high risk (*H2*, *H5*) and low risks haplotypes (*H1*,*H4*) . Sample enrichment increase the possibility of discovering polymorphisms common to cases and to the risk of disease. It may in future studies be of interest to compare the discovery rates between cases and controls. A higher number of polymorphisms in the disease class could indicate a greater mutation rate in this group, which would be of importance if the *CD/RV* hypothesis were true. A complementary study may be to examine the patterns of nucleotide/haplotype diversity between cases and controls. An excess of rare variants or common polymorphisms in either group, may represent an elevated mutation rate or, in particular, enrichment for certain polymorphisms that might distinguish the groups and improve SNP selection and power. It should be pointed out that this is speculation and it has to date not been tested. There may be unforeseen biases that may confound such a comparison between cases and controls, but studying candidate genes in large cohorts followed by comparison of the nucleotide distribution between the groups could be also interesting very interesting also.

13.4.2 Application of Variation in Association studies

Obesity and Serum Leptin: Paper I

.

Significant associations between *HTR2C* haplotypes, diplotypes and obesity suggest that promoter polymorphisms collectively act to increase the risk of obesity. However, given the rarity of these haplotypes, replication is necessary to validate their frequencies and also rule out imputation error. Even though expression analysis was performed on the most common haplotypes, it is difficult to extrapolate what effects *TA13GCG* could have on expression in a luciferase reporter vector. If it is assumed that this haplotype has lower expression ability, then its contribution to obesity may be similar, but less extreme, to that exhibited by the *htr2c*-KO mouse model. The degree of expression by this promoter needs to be investigated before any conclusions can be made concerning its relationship to obesity. The *TA13GCG/TA16GCG* diplotype also presented elevated serum leptin levels (corrected for % body fat), which is also similar to the *htr2c*-KO mouse condition. This may suggest that obese individuals are resistant to leptin or that *HTR2C* is an alternative pathway for controlling appetite and phagia. Schizophrenics treated with clozapine show increases in weight, leptin and serum triglycerides [348]. This study is complemented by our study, which shows that -995*G*>*A* heterozygotes have higher serum leptin and work by Pooley et al, which found that -759*C*>*T* heterozygotes have elevated triglycerides [303]. Therefore, there are potential grounds for the heterosis hypothesis [303], which may also be supported by evidence of balancing selection in our sample.

Depression: Paper II & IV

The differences in trbc-MAO activity between genders, smoking, and depressive states were not surprising but compliment previous findings concerning these groups (Paper II). The lack of variation in the sequenced regions and the poor validation rate made interpretation of the association studies difficult. Although an increase in risk for depressive state was observed for the Norrie Disease gene polymorphism *rs*766117, *MAOB* SNP *rs*1181252, and *MAOB* homozygote haplotypes, no correlations with trbc-activity were found. Conversely, decreased trbc-MAO activity was associated with *MAOA* SNP *rs*979605 as well as haplotypes *A1* and *A3*, which may suggest that factors in controlling *MAOB* activity may be located between the genes or in LD with *MAOA* haplotypes. This is further suggested by the observation that departures from HWE become more obvious towards the 3’ end of each gene (Paper IV). Departures from HWE have been used to indicate variation in copy number and it therefore may be interesting to investigate the region between *MAOA* and *MAOB*, either for polymorphisms or for variation in the number of potential regulatory elements that could affect trbc-activity.

The lack of X-inactivation at the *MAOA* locus could be a factor in explaining the increased risk for depression by *A1* homozygotes. However, further reduced trbc-activity by *A1* homozygotes based on haplotype effects, is contrary to what is expected in depression. The activity has been shown to decrease with age and therefore may be an important factor in the observations made in this study. Lower activity is often linked to other behaviors such as alcoholism, aggression and suicide, which could have effects early in life as indicated by other studies [324]. These behaviors may be more common among males than females and, perhaps other factors are involved in male susceptibility to depression [349].

Even increasing the sample size did not help in identifying a role for MAO SNPs or haplotypes in risk for depressive state (Paper VI). However, *HTR2C* SNPs and haplotypes did show a trend towards significance in increasing the risk for depression in females. The *HTR2A* promoter polymorphism -1438*G*>*A*, was previously found to significantly increase the risk in males from the same sample used in this study. Perhaps genotyping the microsatellite may strenghthen this finding of potential sexual dimorphism. The associated *HTR2C* SNPs and haplotypes were shown to reduce luciferase expression in the obesity study (Paper I). On the other hand reduced expression would be expected to mimic the effects of *HTR2C* antagonists such as fluoxitine. However, increased editing of *HTR2C* mRNA at site E, which translates to a receptor (5-HT2C$_{VGV}$) with an increased affinity for 5-HT, is reversed by fluoxitine, so even if there is a reduction in the receptor density, downstream signaling may be increased by the edited isoform. Therefore, these polymorphisms could be important in the response to treatments for depression by influencing the negative feedback of the serotonergic system and enhancing the effects of SSRIs as indicated by animal models [350].

IDE, *KIF11*, *HHEX* and Alzhiemers Disease: Paper III

In contrast to the *MAO* re-sequencing study (Paper II), a wealth of polymorphisms were found in the *IDE*, *KIF11* and *HHEX* haploblock (Paper III). Although in line with the hypothesis driven/positional candidate gene approach, randomly selection of SNPs for association with Alzheimers disease with a bias in the most biologically relevant gene may not have been the best approach. It may have been better to examine LD between the SNPs following the sequencing of the genes and then those that maximize the information across the region. This may indicate a potential flaw in direct association studies and the use of evenly spaced markers for WGA. As observed in Phase I of the HapMap, the selection of tagged SNPs was improved taking the polymorphisms discovered in re-sequenced ENCODE regions into account. Prince et al. found no differences in the genotype frequencies between the same Scottish case-control samples used in Paper III. However, they did find a significant difference in the distribution of the high-risk haplotype *H2* with the MMSE test. Unfortunately, this was not assessed in the present study and perhaps there was a missed opportunity to improve on the previous association. Most of the findings involving *IDE* have been with late onset AD. Therefore, the fact that no association was observed and the patients in this sample had early onset AD may indicate that other strong influences are involved in the predisposition to AD. Nonetheless, the polymorphisms identified in this re-sequencing effort may stand to help with the increasing number of associations with polymorphisms in the 3’ end of *IDE* [351, 352, 353, 354].

13.5 Conclusions and Future Perspectives

Based on the observations made in the papers presented in this thesis, the re-sequencing of candidate genes has great potential for improving association study design and the power to identify polymorphisms that contribute to complex diseases such as obesity, depression and Alzheimers disease. Unraveling the patterns of nucleotide variation presents a greater understanding of the forces that shape the polymorphism patterns in the genomic region of interest. Events such as recombination and gene conversion are readily identified and therefore a better picture of the LD structure can be defined. As observed in much larger re-sequencing efforts, this is a critical step in the selection of appropriate SNPs for association studies. Extensive comparative sequencing will be necessary in loci that prove to be variation poor. This may be assisted by the inclusion of affected individuals, which may lead to the identification of novel polymorphisms that contribute to the phenotype. Therefore re-sequencing candidate genes can be used as a tool for both the *CD/CV* and *CD/RV* models.

An association study is only as strong as its weakest link. Therefore capitalizing on the advantages of re-sequencing requires that other factors such as the sample size and candidate gene motivations are equally optimized. The power to identify all forms of variation is reduced in small samples. Classification of phenotypes, biological markers and sample selection could be important for diseases where *CD/RV* is a factor. A strong basis for studying a candidate gene and functional validation of variation will make interpretation of the genetic contributions to complex disease much more sound. However, the extrapolation from in vitro conditions to an in vivo model requires caution.

Over the last 11 years, advances have been made to improve both statistical and technological approaches for locating disease genes under the premise that complex diseases have a simple genetic architecture shared in different populations. High-throughput advancements in genotyping technologies for whole genome studies sound promising but are limited to the variation that has already been discovered. The completion of the HapMap (Phase II) will be of great benefit for the tailoring of population specific association studies based on common polymorphisms, but the potential to identify low frequency susceptibility alleles with this map remains to be shown empirically. Furthermore, the quest for the 1000-dollar genome has now begun with large grant proposals for the promotion of novel sequencing methods developments or improvements on current technologies. The re-sequencing of genes will not be limited to a selection of single candidates but will include the genes of entire pathways to explore for the genetic components contributing to biological mechanisms and potential variability in disease phenotypes.

Therefore, in the next decade comparative re-sequencing has the potential to replace genotyping as a standard technique. Researchers will be equipped not only for high throughput polymorphism detection but de novo discovery. This will ultimately influence the ability to clarify the role of candidate genes and genetic variation in complex disease.

Chapter 14

Acknowledgments

Before you embark on reading this page and looking for your name, I would like to thank you for being a friend, my family, a colleague or an acquaintance and saving me from finishing this thesis on Chapter 13. You may say a chapter dedicated to acknowledgments really is not professional, bordering on cheating even, and to this I would respond . . . . "well, I haven't been sleeping that much, allow me this much" , politely of course . . . .

First of all, I would like to thank my supervisor **Dr. Björn Andersson**. I am truly grateful for your guidance through the PhD. ropes. There is no doubt, reaching this point has been equally challenging for you. Thanks for your eagerness to make this project work, for your enthusiasm in bettering ones self and for your level headiness at all times. Thank you for your patience in the early stages of my time in Sweden. Coming here meant there was a lot to learn about life and I have really enjoyed the pints, the whiskey and the laughs. I hope the future holds great things in store for you.

I would like to also thank:

My co-supervisor and former prefect , Professor Claes Wahlestedt, for his welcome to CGB, his support and collaborations.

Former CGB prefect, Professor Christer Höög, for his door always being open

Professor Nancy Pedersen, for supporting my ideas and helping me to pursue them

Professor Shea Fanning, for making all of this happen

Catherine Daly, for introducing me to genetics

The collaborators who I have worked with over the years, Ann-Charlotte Lundsedt, Prof. Catharina Svanborg, Dr. Salim Mottagui-Tabar, Prof.Peter Arner, Dr. Bengt Sennblad, Dr. Mrten Jansson, Prof. Anthony Brookes, Dr. Jonathan Prince, Dr. Anja Castensson and Prof Elena Jazin

To the Andersson Group, everyone past and present, thanks for making it fun and a fascinating place to be; Erik, Daniel, Ellen, Anh-Nhi, Esteban, Lena, Yumi, Hamid, Daryoush, Kim, Delal, Martti and Stephen.

I have met so many people at CGB over the years and made some terrific friends that I will never forget. I wish I had enough time (and space) to describe the impact each and everyone of you has had: Mary-Rose, Joacim, David, Lars, Sarah, Bill, Jen, Rob C.,Tim, Aliastair, Abi, Carsten, Lukas, Therese, Kairi, Anna, Marcus, Vivian, Rikard, Camilla, Cecilia, Pär, Mohammad, Omid, Fredrick, Geert, Ivana, Jenny, Hong, Hannah, Kirsty, Joel, Hagit, Mia, Mungwang and Ingrid. Cheers to all pub the pubcrew over the years, may there be many more!

Stephen, Ylva, Gerry, Mauro, and Åsa, thanks for great times and a lot of laughs.

To my Irish pals who have since left for greener pastures, Paddy, Jim, Sarah, Margaret, Kevin and Lou go raibh mile maith agaibh!!

Danke Elke and Uwe for the hundreds of invites.

Daniel, Bruce, Brian, Reginald and Sergey, for all sorts of training.

Carole and Sindy, thanks for taking care of the *kid*.

Marcela, gracias for always listening.

Alan, thanks for being a real friend.

To Mam, Dad, Rob, Lyn, Emma-Lyn and my family in Ireland. I can not even thank you enough for your support throughout this time. You have always been there when I needed encouragement. I am truly blessed to have you. To think that I have been here for five and a half years is always weird for me because I feel close to you. I am eternally grateful for your visits. Home was never far away at all.

To Emily, what would I have done and where would I have ended up without you? Thank you for your support and all your care. Coming to Sweden turned out not just for studying, and life has been so much more with you in it. Wov woo

And to anyone who I have forgotten to mention here, I apologize terribly

Bibliography

[1] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. 0028-0836 (Print) Journal Article.

[2] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291:1304–51. 0036-8075 (Print) Journal Article.

[3] Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–33. 0028-0836 (Print) Journal Article.

[4] Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, et al. (2005) A haplotype map of the human genome. Nature 437:1299–320. 1476-4687 (Electronic) Journal Article.

[5] Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–8. 0028-0836 (Print) Journal Article.

[6] Eddy SR (1999) Noncoding rna genes. Curr Opin Genet Dev 9:695–9. 0959-437X (Print) Journal Article Review.

[7] Zhang Z, Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res 31:5338–48. 1362-4962 (Electronic) Journal Article.

[8] Rovelet-Lecrux A, Hannequin D, Raux G, Meur NL, Laquerriere A, et al. (2006) App locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–6. 1061-4036 (Print) Journal Article.

[9] Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, et al. (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22:164–7. 1061-4036 (Print) Journal Article.

[10] Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in dna. Nature 287:560–1. 0028-0836 (Print) Journal Article.

[11] Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63:474–88. 0002-9297 (Print) Journal Article.

[12] Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–74. 1088-9051 (Print) Journal Article.

[13] Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous snps: server and survey. Nucleic Acids Res 30:3894–900. 1362-4962 (Electronic) Journal Article.

[14] Knight JC (2005) Regulatory polymorphisms underlying complex disease traits. J Mol Med 83:97–109. 0946-2716 (Print) Journal Article Review.

[15] Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence changes. J Med Genet 42:737–48. 1468-6244 (Electronic) Journal Article.

[16] Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297:1007–13. 1095-9203 (Electronic) Journal Article.

[17] Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, et al. (2004) Rescue-ese identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32:W187–90. 1362-4962 (Electronic) Journal Article.

[18] Mullis KB, Faloona FA (1987) Specific synthesis of dna in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335–50. 0076-6879 (Print) Journal Article.

[19] Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–4. 0036-8075 (Print) Journal Article.

[20] Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and dna polymorphisms using the polymerase chain reaction. Genomics 5:874–9. 0888-7543 (Print) Journal Article.

[21] Ganguly A, Rock MJ, Prockop DJ (1993) Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded pcr products and dna fragments: evidence for solvent-induced bends in dna heteroduplexes. Proc Natl Acad Sci U S A 90:10325–9. 0027-8424 (Print) Journal Article.

[22] Myers RM, Maniatis T, Lerman LS (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol 155:501–27. 0076-6879 (Print) Journal Article.

[23] Cotton RG, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci U S A 85:4397–401. 0027-8424 (Print) Journal Article.

[24] Hart JR, Johnson MD, Barton JK (2004) Single-nucleotide polymorphism discovery by targeted dna photocleavage. Proc Natl Acad Sci U S A 101:14040–4. 0027-8424 (Print) Journal Article.

[25] Faham M, Baharloo S, Tomitaka S, DeYoung J, Freimer NB (2001) Mismatch repair detection (mrd): high-throughput scanning for dna variations. Hum Mol Genet 10:1657–64. 0964-6906 (Print) Evaluation Studies Journal Article.

[26] Sanger F, Coulson AR (1975) A rapid method for determining sequences in dna by primed synthesis with dna polymerase. J Mol Biol 94:441–8. 0022-2836 (Print) Journal Article.

[27] Maxam AM, Gilbert W (1977) A new method for sequencing dna. Proc Natl Acad Sci U S A 74:560–4. 0027-8424 (Print) Journal Article.

[28] Gyllensten UB, Erlich HA (1988) Generation of single-stranded dna by the polymerase chain reaction and its application to direct sequencing of the hla-dqa locus. Proc Natl Acad Sci U S A 85:7652–6. 0027-8424 (Print) Journal Article.

[29] Metzker ML, Lu J, Gibbs RA (1996) Electrophoretically uniform fluorescent dyes for automated dna sequencing. Science 271:1420–2. 0036-8075 (Print) Journal Article.

[30] Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, et al. (1986) Fluorescence detection in automated dna sequence analysis. Nature 321:674–9. 0028-0836 (Print) Journal Article.

[31] Innis MA, Myambo KB, Gelfand DH, Brow MA (1988) Dna sequencing with thermus aquaticus dna polymerase and direct sequencing of polymerase chain reaction-amplified dna. Proc Natl Acad Sci U S A 85:9436–40. 0027-8424 (Print) Journal Article.

[32] Kheterpal I, Scherer JR, Clark SM, Radhakrishnan A, Ju J, et al. (1996) Dna sequencing using a four-color confocal fluorescence capillary array scanner. Electrophoresis 17:1852–9. 0173-0835 (Print) Journal Article.

[33] Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. ii. error probabilities. Genome Res 8:186–94. 1088-9051 (Print) Journal Article.

[34] Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. i. accuracy assessment. Genome Res 8:175–85. 1088-9051 (Print) Journal Article.

[35] Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202. 1088-9051 (Print) Journal Article.

[36] Nickerson DA, Tobe VO, Taylor SL (1997) Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25:2745–51. 0305-1048 (Print) Journal Article.

[37] Emrich CA, Tian H, Medintz IL, Mathies RA (2002) Microfabricated 384-lane capillary array electrophoresis bioanalyzer for ultrahigh-throughput genetic analysis. Anal Chem 74:5076–83. 0003-2700 (Print) Journal Article.

[38] Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single dna molecules. Proc Natl Acad Sci U S A 100:3960–4. 0027-8424 (Print) Journal Article.

[39] Metzker ML, Raghavachari R, Richards S, Jacutin SE, Civitello A, et al. (1994) Termination of dna synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. Nucleic Acids Res 22:4259–67. 0305-1048 (Print) Journal Article.

[40] Deamer DW, Branton D (2002) Characterization of nucleic acids by nanopore analysis. Acc Chem Res 35:817–25. 0001-4842 (Print) Journal Article Review.

[41] Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM (2003) Fluorescent in situ sequencing on polymerase colonies. Anal Biochem 320:55–65. 0003-2697 (Print) Journal Article.

[42] Chee M, Yang R, Hubbell E, Berno A, Huang XC, et al. (1996) Accessing genetic information with high-density dna arrays. Science 274:610–4. 0036-8075 (Print) Journal Article.

[43] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–80. 1476-4687 (Electronic) Journal Article.

[44] Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–8. 1061-4036 (Print) Journal Article.

[45] Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, et al. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–47. 1061-4036 (Print) Journal Article.

[46] Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate snps from public est data. Nat Genet 21:323–5. 1061-4036 (Print) Journal Article.

[47] Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, et al. (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet 23:452–6. 1061-4036 (Print) Journal Article.

[48] Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. Genome Res 8:748–54. 1088-9051 (Print) Journal Article.

[49] Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, et al. (1999) Mining snps from est databases. Genome Res 9:167–74. 1088-9051 (Print) Journal Article.

[50] Irizarry K, Kustanovich V, Li C, Brown N, Nelson S, et al. (2000) Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. Nat Genet 26:233–6. 1061-4036 (Print) Journal Article.

[51] Tajima F (1983) Evolutionary relationship of dna sequences in finite populations. Genetics 105:437–60. 0016-6731 (Print) Journal Article.

[52] Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–76. 0040-5809 (Print) Journal Article.

[53] Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–6. 0028-0836 (Print) Journal Article.

[54] Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–73. 0027-8424 (Print) Journal Article.

[55] Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by dna polymorphism. Genetics 123:585–95. 0016-6731 (Print) Journal Article.

[56] Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709. 0016-6731 (Print) Journal Article.

[57] Jeffreys AJ (1979) Dna sequence variants in the g gamma-, a gamma-, delta- and beta-globin genes of man. Cell 18:1–10. 0092-8674 (Print) Journal Article.

[58] Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, et al. (1998) Dna sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–40. 1061-4036 (Print) Journal Article.

[59] Fullerton SM, Bond J, Schneider JA, Hamilton B, Harding RM, et al. (2000) Polymorphism and divergence in the beta-globin replication origin initiation region. Mol Biol Evol 17:179–88. 0737-4038 (Print) Journal Article.

[60] Nickerson DA, Taylor SL, Fullerton SM, Weiss KM, Clark AG, et al. (2000) Sequence diversity and large-scale typing of snps in the human apolipoprotein e gene. Genome Res 10:1532–45. 1088-9051 (Print) Journal Article.

[61] Mateu E, Calafell F, Lao O, Bonne-Tamir B, Kidd JR, et al. (2001) Worldwide genetic analysis of the cftr region. Am J Hum Genet 68:103–17. 0002-9297 (Print) Journal Article.

[62] Subrahmanyan L, Eberle MA, Clark AG, Kruglyak L, Nickerson DA (2001) Sequence variation and linkage disequilibrium in the human t-cell receptor beta (tcrb) locus. Am J Hum Genet 69:381–95. 0002-9297 (Print) Journal Article.

[63] Crawford DC, Akey DT, Nickerson DA (2005) The patterns of natural variation in human genes. Annu Rev Genomics Hum Genet 6:287–312. 1527-8204 (Print) Journal Article Review.

[64] Garg K, Green P, Nickerson DA (1999) Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. Genome Res 9:1087–92. 1088-9051 (Print) Journal Article.

[65] Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27:234–6. 1061-4036 (Print) News.

[66] Kroetz DL, Pauli-Magnus C, Hodges LM, Huang CC, Kawamoto M, et al. (2003) Sequence diversity and haplotype structure in the human abcb1 (mdr1, multidrug resistance transporter) gene. Pharmacogenetics 13:481–94. 0960-314X (Print) Journal Article.

[67] Glatt CE, DeYoung JA, Delgado S, Service SK, Giacomini KM, et al. (2001) Screening a large reference sample to identify very low frequency sequence variants: comparisons between two genes. Nat Genet 27:435–8. 1061-4036 (Print) Journal Article.

[68] Li WH, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513–23. 0016-6731 (Print) Journal Article.

[69] Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, et al. (2004) Pattern of sequence variation across 213 environmental response genes. Genome Res 14:1821–31. 1088-9051 (Print) Journal Article.

[70] Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, et al. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 32:135–42. 1061-4036 (Print) Journal Article.

[71] Hellmann I, Prufer K, Ji H, Zody MC, Paabo S, et al. (2005) Why do human diversity levels vary at a megabase scale? Genome Res 15:1222–31. 1088-9051 (Print) Journal Article.

[72] Miller RD, Taillon-Miller P, Kwok PY (2001) Regions of low single-nucleotide polymorphism incidence in human and orangutan xq: deserts and recent coalescences. Genomics 71:78–88. 0888-7543 (Print) Journal Article.

[73] Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. (2005) The dna sequence of the human x chromosome. Nature 434:325–37. 1476-4687 (Electronic) Journal Article.

[74] Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human dna diversity. Proc Natl Acad Sci U S A 94:4516–9. 0027-8424 (Print) Journal Article.

[75] Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, et al. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–7. 0028-0836 (Print) Journal Article.

[76] Harris EE, Hey J (1999) X chromosome evidence for ancient human histories. Proc Natl Acad Sci U S A 96:3320–4. 0027-8424 (Print) Journal Article.

[77] Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, et al. (1991) The structure of human mitochondrial dna variation. J Mol Evol 33:543–55. 0022-2844 (Print) Journal Article.

[78] Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial dna. Science 253:1503–7. 0036-8075 (Print) Journal Article.

[79] Lewontin RC (1964) The interaction of selection and linkage. i. general considerations; heterotic models. Genetics 49:49–67.

[80] Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309. 1471-0056 (Print) Journal Article Review.

[81] Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19–24. 0168-9525 (Print) Journal Article.

[82] Lewontin RC (1988) On measures of gametic disequilibrium. Genetics 120:849–52. 0016-6731 (Print) Journal Article.

[83] Hill W, Roberston A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–239.

[84] Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14. 0002-9297 (Print) Journal Article Review.

[85] Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, et al. (2001) Linkage disequilibrium in the human genome. Nature 411:199–204. 0028-0836 (Print) Journal Article.

[86] Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press.

[87] Ewans W (1982) On the concept of effective population size. Theor Popul Biol 21:373–378.

[88] Strobeck C (1987) Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics 117:149–153.

[89] Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–25. 0016-6731 (Print) Journal Article.

[90] Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–44. 1061-4036 (Print) Journal Article.

[91] Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A, Slatkin M, et al. (1995) The distribution of linkage disequilibrium over anonymous genome regions. Hum Mol Genet 4:887–94. 0964-6906 (Print) Journal Article.

[92] Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435–8. 1061-4036 (Print) Journal Article.

[93] Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci U S A 96:15173–7. 0027-8424 (Print) Journal Article.

[94] Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. Genetics 152:1711–22. 0016-6731 (Print) Journal Article.

[95] Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, et al. (2002) A first-generation linkage disequilibrium map of human chromosome 22. Nature 418:544–8. 0028-0836 (Print) Journal Article.

[96] Eisenbarth I, Striebel AM, Moschgath E, Vogel W, Assum G (2001) Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 mb region of human chromosome 22. Hum Mol Genet 10:2833–9. 0964-6906 (Print) Journal Article.

[97] Smith AV, Thomas DJ, Munro HM, Abecasis GR (2005) Sequence features in regions of weak and strong linkage disequilibrium. Genome Res 15:1519–34. 1088-9051 (Print) Journal Article.

[98] Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229–32. 1061-4036 (Print) Journal Article.

[99] Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–23. 0036-8075 (Print) Journal Article.

[100] Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, et al. (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human xq25 and xq28. Nat Genet 25:324–8. 1061-4036 (Print) Journal Article.

[101] McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, et al. (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581–4. 1095-9203 (Electronic) Journal Article.

[102] Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P (2005) Human recombination hot spots hidden in regions of strong marker association. Nat Genet 37:601–6. 1061-4036 (Print) Journal Article.

[103] Kauppi L, Stumpf MP, Jeffreys AJ (2005) Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human mhc class ii region. Genomics 86:13–24. 0888-7543 (Print) Journal Article.

[104] Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, et al. (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nat Genet 37:429–34. 1061-4036 (Print) Journal Article.

[105] Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) A dynamic programming algorithm for haplotype block partitioning. Proc Natl Acad Sci U S A 99:7335–9. 0027-8424 (Print) Journal Article.

[106] Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, et al. (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–7. 1061-4036 (Print) Journal Article.

[107] Goldstein DB, Ahmadi KR, Weale ME, Wood NW (2003) Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. Trends Genet 19:615–22. 0168-9525 (Print) Journal Article.

[108] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. Science 296:2225–9. 1095-9203 (Electronic) Journal Article.

[109] Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489–93. 0036-8075 (Print) Journal Article.

[110] Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, et al. (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–402. 0002-9297 (Print) Journal Article.

[111] Tishkoff SA, Verrelli BC (2003) Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr Opin Genet Dev 13:569–75. 0959-437X (Print) Journal Article Review.

[112] Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19:135–40. 0168-9525 (Print) Journal Article Review.

[113] Evans DM, Cardon LR (2005) A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. Am J Hum Genet 76:681–7. 0002-9297 (Print) Journal Article.

[114] Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304. 0016-6731 (Print) Journal Article.

[115] Malcom CM, Wyckoff GJ, Lahn BT (2003) Genic mutation rates in mammals: local similarity, chromosomal heterogeneity, and x-versus-autosome disparity. Mol Biol Evol 20:1633–41. 0737-4038 (Print) Journal Article.

[116] Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340. 1527-8204 (Print) Journal Article Review.

[117] Miller RD, Kwok PY (2001) The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. Hum Mol Genet 10:2195–8. 0964-6906 (Print) Historical Article Journal Article Review.

[118] Terwilliger JD, Zollner S, Laan M, Paabo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. Hum Hered 48:138–54. 0001-5652 (Print) Journal Article Review.

[119] Begun DJ, Aquadro CF (1992) Levels of naturally occurring dna polymorphism correlate with recombination rates in d. melanogaster. Nature 356:519–20. 0028-0836 (Print) Journal Article.

[120] Nachman MW (2001) Single nucleotide polymorphisms and recombination rate in humans. Trends Genet 17:481–5. 0168-9525 (Print) Journal Article.

[121] Payseur BA, Nachman MW (2000) Microsatellite variation and recombination rate in the human genome. Genetics 156:1285–98. 0016-6731 (Print) Journal Article.

[122] Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, et al. (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–7. 1061-4036 (Print) Journal Article.

[123] Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of dna sequences. Genetics 111:147–64. 0016-6731 (Print) Journal Article.

[124] Stumpf MP, McVean GA (2003) Estimating recombination rates from population-genetic data. Nat Rev Genet 4:959–68. 1471-0056 (Print) Journal Article Review.

[125] Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, et al. (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. Am J Hum Genet 69:582–9. 0002-9297 (Print) Journal Article.

[126] Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, et al. (2001) Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am J Hum Genet 69:831–43. 0002-9297 (Print) Journal Article.

[127] Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, et al. (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. Am J Hum Genet 66:69–83. 0002-9297 (Print) Journal Article.

[128] Hilliker AJ, Harauz G, Reaume AG, Gray M, Clark SH, et al. (1994) Meiotic gene conversion tract length distribution within the rosy locus of drosophila melanogaster. Genetics 137:1019–26. 0016-6731 (Print) Journal Article.

[129] Andolfatto P, Nordborg M (1998) The effect of gene conversion on intralocus associations. Genetics 148:1397–9. 0016-6731 (Print) Letter.

[130] Schimenti JC (1994) Gene conversion and the evolution of gene families in mammals. Soc Gen Physiol Ser 49:85–91. 0094-7733 (Print) Journal Article Review.

[131] Hayashida H, Kuma K, Miyata T (1992) Interchromosomal gene conversion as a possible mechanism for explaining divergence patterns of zfy-related genes. J Mol Evol 35:181–3. 0022-2844 (Print) Letter.

[132] Zangenberg G, Huang MM, Arnheim N, Erlich H (1995) New hla-dpb1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat Genet 10:407–14. 1061-4036 (Print) Journal Article.

[133] Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, et al. (1994) Complex gene conversion events in germline mutation at human minisatellites. Nat Genet 6:136–45. 1061-4036 (Print) Journal Article.

[134] Sharon D, Gilad Y, Glusman G, Khen M, Lancet D, et al. (2000) Identification and characterization of coding single-nucleotide polymorphisms within a human olfactory receptor gene cluster. Gene 260:87–94. 0378-1119 (Print) Journal Article.

[135] Collier S, Tassabehji M, Sinnott P, Strachan T (1993) A de novo pathological point mutation at the 21-hydroxylase locus: implications for gene conversion in the human genome. Nat Genet 3:260–5. 1061-4036 (Print) Case Reports Journal Article.

[136] Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, et al. (2003) Mutations in sbds are associated with shwachman-diamond syndrome. Nat Genet 33:97–101. 1061-4036 (Print) Journal Article.

[137] Teich N, Nemoda Z, Kohler H, Heinritz W, Mossner J, et al. (2005) Gene conversion between functional trypsinogen genes prss1 and prss2 associated with chronic pancreatitis in a six-year-old girl. Hum Mutat 25:343–7. 1098-1004 (Electronic) Journal Article.

[138] Eikenboom JC, Vink T, Briet E, Sixma JJ, Reitsma PH (1994) Multiple substitutions in the von willebrand factor gene that mimic the pseudogene sequence. Proc Natl Acad Sci U S A 91:2221–4. 0027-8424 (Print) Journal Article.

[139] Sunyaev S, Kondrashov FA, Bork P, Ramensky V (2003) Impact of selection, mutation rate and genetic drift on human genetic variation. Hum Mol Genet 12:3325–30. 0964-6906 (Print) Journal Article.

[140] Przeworski M (2002) The signature of positive selection at randomly chosen loci. Genetics 160:1179–89. 0016-6731 (Print) Journal Article.

[141] Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111. 1471-0056 (Print) Journal Article Review.

[142] Hughes AL, Yeager M (1998) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3:d509–16. 1093-4715 (Electronic) Journal Article Review.

[143] Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, et al. (1997) Archaic african and asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–89. 0002-9297 (Print) Journal Article.

[144] Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, et al. (2002) A strong signature of balancing selection in the 5' cis-regulatory region of ccr5. Proc Natl Acad Sci U S A 99:10539–44. 0027-8424 (Print) Journal Article.

[145] Mead S, Stumpf MP, Whitfield J, Beck JA, Poulter M, et al. (2003) Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. Science 300:640–3. 1095-9203 (Electronic) Historical Article Journal Article.

[146] Goodman M (1962) Evolution of the immunologic species specificity of human serum proteins. Hum Biol 34:104–50. 0018-7143 (Print) Journal Article.

[147] Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–82. 0036-8075 (Print) Journal Article.

[148] Feuk L, Macdonald JR, Tang T, Carson AR, Li M, et al. (2005) Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee dna sequence assemblies. PLoS Genet 1:e56. 1553-7390 (Print) Journal Article.

[149] Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, et al. (1998) Out of africa and back again: nested cladistic analysis of human y chromosome variation. Mol Biol Evol 15:427–41. 0737-4038 (Print) Journal Article.

[150] Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtdna control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–82. 0002-9297 (Print) Journal Article.

[151] Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, et al. (1996) Global patterns of linkage disequilibrium at the cd4 locus and modern human origins. Science 271:1380–7. 0036-8075 (Print) Journal Article.

[152] Hardy J, Pittman A, Myers A, Gwinn-Hardy K, Fung HC, et al. (2005) Evidence suggesting that homo neanderthalensis contributed the h2 mapt haplotype to homo sapiens. Biochem Soc Trans 33:582–5. 0300-5127 (Print) Journal Article Review.

[153] Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, et al. (2002) Molecular evolution of foxp2, a gene involved in speech and language. Nature 418:869–72. 0028-0836 (Print) Journal Article.

[154] Lander ES (1996) The new genomics: global views of biology. Science 274:536–9. 0036-8075 (Print) Journal Article.

[155] Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318. 0002-9297 (Print) Journal Article.

[156] Clerget-Darpoux F (2001) Extension of the lod score: the mod score. Adv Genet 42:115–24. 0065-2660 (Print) Journal Article Review.

[157] Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary dna. Science 245:1066–73. 0036-8075 (Print) Journal Article.

[158] Easton DF (1999) How many more breast cancer predisposition genes are there? Breast Cancer Res 1:14–7. 1465-5411 (Print) Editorial.

[159] Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, et al. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial alzheimer's disease. Nature 349:704–6. 0028-0836 (Print) Journal Article.

[160] Dean M (2003) Approaches to identify genes for complex human diseases: lessons from mendelian disorders. Hum Mutat 22:261–74. 1098-1004 (Electronic) Journal Article Review.

[161] Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Essex, England: Longman, 4th edition. D.S. Falconer and Trudy F.C. Mackay. Quantitative genetics ill.; 24 cm.

[162] Risch N (1990) Linkage strategies for genetically complex traits. i. multilocus models. Am J Hum Genet 46:222–8. 0002-9297 (Print) Journal Article.

[163] Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl:228–37. 1061-4036 (Print) Historical Article Journal Article Review.

[164] Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, et al. (2002) Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 31:33–6. 1061-4036 (Print) Journal Article.

[165] Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 9:578–94. 0958-1669 (Print) Journal Article Review.

[166] Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with snps? Nat Genet 26:151–7. 1061-4036 (Print) Journal Article Review.

[167] Willett WC (2002) Balancing life-style and genomics research for disease prevention. Science 296:695–8. 1095-9203 (Electronic) Journal Article.

[168] Chakravarti A (1999) Population genetics–making sense out of sequence. Nat Genet 21:56–60. 1061-4036 (Print) Journal Article Review.

[169] Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17:502–10. 0168-9525 (Print) Journal Article.

[170] Neel JV (1962) Diabetes mellitus: a ”thrifty” genotype rendered detrimental by ”progress”? Am J Hum Genet 14:353–62. 0002-9297 (Print) Journal Article.

[171] Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H (2003) A polygenic basis for late-onset disease. Trends Genet 19:97–106. 0168-9525 (Print) Journal Article Review.

[172] Pritchard JK, Cox NJ (2002) The allelic architecture of human disease genes: common disease-common variant.or not? Hum Mol Genet 11:2417–23. 0964-6906 (Print) Journal Article Review.

[173] Wang WY, Pike N (2004) The allelic spectra of common diseases may resemble the allelic spectrum of the full genome. Med Hypotheses 63:748–51. 0306-9877 (Print) Journal Article.

[174] Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, et al. (2006) Variant of transcription factor 7-like 2 (tcf7l2) gene confers risk of type 2 diabetes. Nat Genet 0. 1061-4036 (Print) Journal article.

[175] Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, et al. (1996) The ctla-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. belgian diabetes registry. Hum Mol Genet 5:1075–80. 0964-6906 (Print) Journal Article Multicenter Study.

[176] Williams NM, Norton N, Williams H, Ekholm B, Hamshere ML, et al. (2003) A systematic genomewide linkage study in 353 sib pairs with schizophrenia. Am J Hum Genet 73:1355–67. 0002-9297 (Print) Journal Article.

[177] Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–7. 0036-8075 (Print) Journal Article.

[178] Rao DC, Province MA (2001) Genetic dissection of complex traits. San Diego: Academic Press. Edited by D.C. Rao, Michael A. Province. ill.; 24 cm. Advances in genetics; 42. Based on a symposium held in honor of Newton E. Morton on the occasion of his 70th birthday–P. xix. 1. Newton Morton's contributions – 2. Overview and preliminaries – 3. Phenotypes and genotypes – 4. Model-based methods for linkage analysis – 5. Model-free methods for linkage and association analysis – 6. More recent methods – 7. Optimum strategies – 8. Multiple comparisons and significance levels – 9. Challenges for the new milennium.

[179] Brookes AJ (1999) The essence of snps. Gene 234:177–86. 0378-1119 (Print) Journal Article Review.

[180] Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, et al. (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to crohn disease. Nat Genet 29:223–8. 1061-4036 (Print) Journal Article.

[181] Glatt CE, Freimer NB (2002) Association analysis of candidate genes for neuropsychiatric disease: the perpetual campaign. Trends Genet 18:307–12. 0168-9525 (Print) Journal Article Review.

[182] Mackay TF (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–39. 0066-4197 (Print) Journal Article Review.

[183] Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, et al. (2005) Comparative genomics and gene expression analysis identifies bbs9, a new bardet-biedl syndrome gene. Am J Hum Genet 77:1021–33. 0002-9297 (Print) Journal Article.

[184] Vitt U, Gietzen D, Stevens K, Wingrove J, Becha S, et al. (2004) Identification of candidate disease genes by est alignments, synteny, and expression and verification of ensembl genes on rat chromosome 1q43-54. Genome Res 14:640–50. 1088-9051 (Print) Journal Article.

[185] Wang X, Ishimori N, Korstanje R, Rollins J, Paigen B (2005) Identifying novel genes for atherosclerosis through mouse-human comparative genetics. Am J Hum Genet 77:1–15. 0002-9297 (Print) Journal Article Review.

[186] Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. (2004) Genome sequence of the brown norway rat yields insights into mammalian evolution. Nature 428:493–521. 1476-4687 (Electronic) Journal Article.

[187] Giallourakis C, Henson C, Reich M, Xie X, Mootha VK (2005) Disease gene discovery through integrative genomics. Annu Rev Genomics Hum Genet 6:381–406. 1527-8204 (Print) Journal Article Review.

[188] Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, et al. (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 dna: the effect of single base pair mismatch. Nucleic Acids Res 6:3543–57. 0305-1048 (Print) Journal Article.

[189] Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, et al. (1983) Detection of sickle cell beta s-globin allele by hybridization with synthetic oligonucleotides. Proc Natl Acad Sci U S A 80:278–82. 0027-8424 (Print) Journal Article.

[190] Saiki RK, Chang CA, Levenson CH, Warren TC, Boehm CD, et al. (1988) Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified dna and nonradioactive allele-specific oligonucleotide probes. N Engl J Med 319:537–41. 0028-4793 (Print) Journal Article.

[191] Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, et al. (2001) Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (dash): design criteria and assay validation. Genome Res 11:152–62. 1088-9051 (Print) Journal Article.

[192] Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal 14:143–9. Journal Article.

[193] Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14:303–8. 1087-0156 (Print) Journal Article.

[194] Johnson MP, Haupt LM, Griffiths LR (2004) Locked nucleic acid (lna) single nucleotide polymorphism (snp) genotype analysis and validation using real-time pcr. Nucleic Acids Res 32:e55. 1362-4962 (Electronic) Journal Article.

[195] Ross PL, Lee K, Belgrader P (1997) Discrimination of single-nucleotide polymorphisms in human dna using peptide nucleic acid probes detected by maldi-tof mass spectrometry. Anal Chem 69:4197–202. 0003-2700 (Print) Journal Article.

[196] Kuimelis RG, Livak KJ, Mullah B, Andrus A (1997) Structural analogues of taqman probes for real-time quantitative pcr. Nucleic Acids Symp Ser 0:255–6. 0261-3166 (Print) Journal Article.

[197] Jobs M, Howell WM, Stromqvist L, Mayr T, Brookes AJ (2003) Dash-2: flexible, low-cost, and high-throughput snp genotyping by dynamic allele-specific hybridization on membrane arrays. Genome Res 13:916–24. 1088-9051 (Print) Evaluation Studies Journal Article.

[198] Lee LG, Livak KJ, Mullah B, Graham RJ, Vinayak RS, et al. (1999) Seven-color, homogeneous detection of six pcr products. Biotechniques 27:342–9. 0736-6205 (Print) Journal Article.

[199] Landegren U, Kaiser R, Sanders J, Hood L (1988) A ligase-mediated gene detection technique. Science 241:1077–80. 0036-8075 (Print) Journal Article.

[200] Nickerson DA, Kaiser R, Lappin S, Stewart J, Hood L, et al. (1990) Automated dna diagnostics using an elisa-based oligonucleotide ligation assay. Proc Natl Acad Sci U S A 87:8923–7. 0027-8424 (Print) Journal Article.

[201] Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, et al. (1999) Universal dna microarray method for multiplex detection of low abundance point mutations. J Mol Biol 292:251–62. 0022-2836 (Print) Journal Article.

[202] Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, et al. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 19:225–32. 1061-4036 (Print) Journal Article.

[203] Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, et al. (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 21:673–8. 1087-0156 (Print) Evaluation Studies Journal Article Validation Studies.

[204] Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, et al. (2005) Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted snps genotyped in a single tube assay. Genome Res 15:269–75. 1088-9051 (Print) Journal Article.

[205] Wang Y, Moorhead M, Karlin-Neumann G, Falkowski M, Chen C, et al. (2005) Allele quantification using molecular inversion probes (mip). Nucleic Acids Res 33:e183. 1362-4962 (Electronic) Evaluation Studies Journal Article.

[206] Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein e. Genomics 8:684–92. 0888-7543 (Print) Journal Article.

[207] Nikiforov TT, Rendle RB, Goelet P, Rogers YH, Kotewicz ML, et al. (1994) Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. Nucleic Acids Res 22:4167–75. 0305-1048 (Print) Journal Article.

[208] Braun A, Little DP, Koster H (1997) Detecting cftr gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 43:1151–8. 0009-9147 (Print) Journal Article.

[209] Nyren P, Pettersson B, Uhlen M (1993) Solid phase dna minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem 208:171–5. 0003-2697 (Print) Journal Article.

[210] Shumaker JM, Metspalu A, Caskey CT (1996) Mutation detection by solid phase primer extension. Hum Mutat 7:346–54. 1059-7794 (Print) Journal Article.

[211] Lindroos K, Liljedahl U, Raitio M, Syvanen AC (2001) Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. Nucleic Acids Res 29:E69–9. 1362-4962 (Electronic) Journal Article.

[212] Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, et al. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 10:1031–42. 1088-9051 (Print) Journal Article.

[213] Chen J, Iannone MA, Li MS, Taylor JD, Rivers P, et al. (2000) A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. Genome Res 10:549–57. 1088-9051 (Print) Journal Article.

[214] Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable snp genotyping assay using microarray technology. Nat Genet 37:549–54. 1061-4036 (Print) Journal Article.

[215] Steemers FJ, Chang W, Lee G, Barker DL, Shen R, et al. (2006) Whole-genome genotyping with the single-base extension assay. Nat Methods 3:31–3. 1548-7091 (Print) Journal Article.

[216] Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, et al. (1999) Polymorphism identification and quantitative detection of genomic dna by invasive cleavage of oligonucleotide probes. Nat Biotechnol 17:292–6. 1087-0156 (Print) Journal Article.

[217] Wilkins Stevens P, Hall JG, Lyamichev V, Neri BP, Lu M, et al. (2001) Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. Nucleic Acids Res 29:E77. 1362-4962 (Electronic) Journal Article.

[218] Strittmatter WJ, Roses AD (1996) Apolipoprotein e and alzheimer's disease. Annu Rev Neurosci 19:53–77. 0147-006X (Print) Journal Article Review.

[219] Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, et al. (1998) A pro12ala substitution in ppargamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet 20:284–7. 1061-4036 (Print) Journal Article.

[220] Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, et al. (2002) Association of the adam33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426–30. 0028-0836 (Print) Journal Article.

[221] Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, et al. (2001) Association of nod2 leucine-rich repeat variants with susceptibility to crohn's disease. Nature 411:599–603. 0028-0836 (Print) Journal Article.

[222] Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, et al. (2002) Functional snps in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–4. 1061-4036 (Print) Journal Article.

[223] Dahlback B (1997) Resistance to activated protein c caused by the factor vr506q mutation is a common risk factor for venous thrombosis. Thromb Haemost 78:483–8. 0340-6245 (Print) Journal Article Review.

[224] de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, et al. (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–23. 1061-4036 (Print) Journal Article.

[225] Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, et al. (2005) Complement factor h polymorphism in age-related macular degeneration. Science 308:385–9. 1095-9203 (Electronic) Journal Article.

[226] Wang WY, Cordell HJ, Todd JA (2003) Association mapping of complex diseases in linked regions: estimation of genetic effects and feasibility of testing rare variants. Genet Epidemiol 24:36–43. 0741-0395 (Print) Journal Article Validation Studies.

[227] Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, et al. (2002) The swedish twin registry: a unique resource for clinical, epidemiological and genetic studies. J Intern Med 252:184–205. 0954-6820 (Print) Journal Article Review.

[228] Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74:765–9. 0002-9297 (Print) Journal Article.

[229] Nyholt DR (2005) Evaluation of nyholt's procedure for multiple testing correction - author's reply. Hum Hered 60:61–2. 0001-5652 (Print) Comment Editorial.

[230] Salyakina D, Seaman SR, Browning BL, Dudbridge F, Muller-Myhsok B (2005) Evaluation of nyholt's procedure for multiple testing correction. Hum Hered 60:19–25; discussion 61–2. 0001-5652 (Print) Journal Article.

[231] Dudbridge F, Koeleman BP (2004) Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. Am J Hum Genet 75:424–35. 0002-9297 (Print) Journal Article.

[232] Sabatti C, Service S, Freimer N (2003) False discovery rate in linkage and association genome screens for complex disorders. Genetics 164:829–33. 0016-6731 (Print) Journal Article.

[233] Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004. 0006-341X (Print) Journal Article.

[234] Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–81. 0002-9297 (Print) Journal Article.

[235] Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An icelandic example of the impact of population structure on association studies. Nat Genet 37:90–5. 1061-4036 (Print) Journal Article.

[236] Zollner S, Wen X, Hanchard NA, Herbert MA, Ober C, et al. (2004) Evidence for extensive transmission distortion in the human genome. Am J Hum Genet 74:62–72. 0002-9297 (Print) Journal Article.

[237] Sklar P, Gabriel SB, McInnis MG, Bennett P, Lim YM, et al. (2002) Family-based association study of 76 candidate genes in bipolar disorder: Bdnf is a potential risk locus. brain-derived neutrophic factor. Mol Psychiatry 7:579–93. 1359-4184 (Print) Journal Article Multicenter Study.

[238] Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61. 1098-3600 (Print) Journal Article Review.

[239] Cavalleri GL, Lynch JM, Depondt C, Burley MW, Wood NW, et al. (2005) Failure to replicate previously reported genetic associations with sporadic temporal lobe epilepsy: where to from here? Brain 128:1832–40. 1460-2156 (Electronic) Journal Article.

[240] Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of rna polymerase loading. Nat Genet 33:469–75. 1061-4036 (Print) Journal Article.

[241] Laitinen T, Polvi A, Rydman P, Vendelin J, Pulkkinen V, et al. (2004) Characterization of a common susceptibility locus for asthma-related traits. Science 304:300–4. 1095-9203 (Electronic) Journal Article.

[242] Ueda H, Howson JM, Esposito L, Heward J, Snook H, et al. (2003) Association of the t-cell regulatory gene ctla4 with susceptibility to autoimmune disease. Nature 423:506–11. 0028-0836 (Print) Journal Article.

[243] Bai F, Rankinen T, Charbonneau C, Belsham DD, Rao DC, et al. (2004) Functional dimorphism of two hagrp promoter snps in linkage disequilibrium. J Med Genet 41:350–3. 1468-6244 (Electronic) Journal Article.

[244] Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, et al. (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci U S A 97:10483–8. 0027-8424 (Print) Journal Article.

[245] Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, et al. (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612. 0002-9297 (Print) Journal Article.

[246] Winkler C, An P, O'Brien SJ (2004) Patterns of ethnic diversity among the genes that influence aids. Hum Mol Genet 13 Spec No 1:R9–19. 0964-6906 (Print) Journal Article Review.

[247] Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, et al. (2000) Apolipoprotein e variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet 67:881–900. 0002-9297 (Print) Journal Article.

[248] Rapport M, Green A, Page I (1948) Partial purification of the vasoconstrictor in beef serum. J Biol Chem 174:735–738.

[249] Twarog BM, Page IH (1953) Serotonin content of some mammalian tissues and urine and a method for its determination. Am J Physiol 175:157–61. 0002-9513 (Print) Journal Article.

[250] Audet MA, Descarries L, Doucet G (1989) Quantified regional and laminar distribution of the serotonin innervation in the anterior half of adult rat cerebral cortex. J Chem Neuroanat 2:29–44. 0891-0618 (Print) Journal Article.

[251] Rubenstein JL (1998) Development of serotonergic neurons and their projections. Biol Psychiatry 44:145–50. 0006-3223 (Print) Journal Article Review.

[252] Jacobs BL, Azmitia EC (1992) Structure and function of the brain serotonin system. Physiol Rev 72:165–229. 0031-9333 (Print) Journal Article Review.

[253] Kosofsky BE, Molliver ME (1987) The serotoninergic innervation of cerebral cortex: different classes of axon terminals arise from dorsal and median raphe nuclei. Synapse 1:153–68. 0887-4476 (Print) Journal Article.

[254] Sodhi MS, Sanders-Bush E (2004) Serotonin and brain development. Int Rev Neurobiol 59:111–74. 0074-7742 (Print) Journal Article Review.

[255] Hoyer D, Clarke DE, Fozard JR, Hartig PR, Martin GR, et al. (1994) International union of pharmacology classification of receptors for 5-hydroxytryptamine (serotonin). Pharmacol Rev 46:157–203. 0031-6997 (Print) Journal Article Review.

[256] Lopez-Gimenez JF, Mengod G, Palacios JM, Vilaro MT (2001) Regional distribution and cellular localization of 5-ht2c receptor mrna in monkey brain: comparison with [3h]mesulergine binding sites and choline acetyltransferase mrna. Synapse 42:12–26. 0887-4476 (Print) Journal Article.

[257] Julius D, MacDermott AB, Axel R, Jessell TM (1988) Molecular characterization of a functional cdna encoding the serotonin 1c receptor. Science 241:558–64. 0036-8075 (Print) Journal Article.

[258] Hoffman BJ, Mezey E (1989) Distribution of serotonin 5-ht1c receptor mrna in adult rat brain. FEBS Lett 247:453–62. 0014-5793 (Print) Journal Article.

[259] Tecott LH, Sun LM, Akana SF, Strack AM, Lowenstein DH, et al. (1995) Eating disorder and epilepsy in mice lacking 5-ht2c serotonin receptors. Nature 374:542–6. 0028-0836 (Print) Journal Article.

[260] Nonogaki K, Strack AM, Dallman MF, Tecott LH (1998) Leptin-independent hyperphagia and type 2 diabetes in mice with a mutated serotonin 5-ht2c receptor gene. Nat Med 4:1152–6. 1078-8956 (Print) Journal Article.

[261] Nonogaki K, Abdallah L, Goulding EH, Bonasera SJ, Tecott LH (2003) Hyperactivity and reduced energy cost of physical activity in serotonin 5-ht(2c) receptor mutant mice. Diabetes 52:315–20. 0012-1797 (Print) Journal Article.

[262] Dourish CT, Clark ML, Fletcher A, Iversen SD (1989) Evidence that blockade of post-synaptic 5-ht1 receptors elicits feeding in satiated rats. Psychopharmacology (Berl) 97:54–8. 0033-3158 (Print) Journal Article.

[263] Fletcher PJ (1988) Increased food intake in satiated rats induced by the 5-ht antagonists methysergide, metergoline and ritanserin. Psychopharmacology (Berl) 96:237–42. 0033-3158 (Print) Journal Article.

[264] Kennett GA, Curzon G (1988) Evidence that hypophagia induced by mcpp and tfmpp requires 5-ht1c and 5-ht1b receptors; hypophagia induced by ru 24969 only requires 5-ht1b receptors. Psychopharmacology (Berl) 96:93–100. 0033-3158 (Print) Journal Article.

[265] Schechter LE, Simansky KJ (1988) 1-(2,5-dimethoxy-4-iodophenyl)-2-aminopropane (doi) exerts an anorexic action that is blocked by 5-ht2 antagonists in rats. Psychopharmacology (Berl) 94:342–6. 0033-3158 (Print) Journal Article.

[266] Vaupel DB, Morton EC (1982) Anorexia and hyperphagia produced by five pharmacologic classes of hallucinogens. Pharmacol Biochem Behav 17:539–45. 0091-3057 (Print) Journal Article.

[267] Rowland NE, Carlton J (1986) Effects of fenfluramine on food intake, body weight, gastric emptying and brain monoamines in syrian hamsters. Brain Res Bull 17:575–81. 0361-9230 (Print) Journal Article.

[268] Sargent PA, Sharpley AL, Williams C, Goodall EM, Cowen PJ (1997) 5-ht2c receptor activation decreases appetite and body weight in obese subjects. Psychopharmacology (Berl) 133:309–12. 0033-3158 (Print) Clinical Trial Journal Article Randomized Controlled Trial.

[269] Walsh AE, Smith KA, Oldman AD, Williams C, Goodall EM, et al. (1994) m-chlorophenylpiperazine decreases food intake in a test meal. Psychopharmacology (Berl) 116:120–2. 0033-3158 (Print) Clinical Trial Journal Article Randomized Controlled Trial.

[270] Heisler LK, Cowley MA, Tecott LH, Fan W, Low MJ, et al. (2002) Activation of central melanocortin pathways by fenfluramine. Science 297:609–11. 1095-9203 (Electronic) Journal Article.

[271] Jorgensen H, Riis M, Knigge U, Kjaer A, Warberg J (2003) Serotonin receptors involved in vasopressin and oxytocin secretion. J Neuroendocrinol 15:242–9. 0953-8194 (Print) Journal Article.

[272] Czeh B, Michaelis T, Watanabe T, Frahm J, de Biurrun G, et al. (2001) Stress-induced changes in cerebral metabolites, hippocampal volume, and cell proliferation are prevented by antidepressant treatment with tianeptine. Proc Natl Acad Sci U S A 98:12796–801. 0027-8424 (Print) Journal Article.

[273] Moreau JL, Jenck F, Martin JR, Perrin S, Haefely WE (1993) Effects of repeated mild stress and two antidepressant treatments on the behavioral response to 5ht1c receptor activation in rats. Psychopharmacology (Berl) 110:140–4. 0033-3158 (Print) Journal Article.

[274] Maes M, Meltzer HY, D'Hondt P, Cosyns P, Blockx P (1995) Effects of serotonin precursors on the negative feedback effects of glucocorticoids on hypothalamic-pituitary-adrenal axis function in depression. Psychoneuroendocrinology 20:149–67. 0306-4530 (Print) Clinical Trial Journal Article Randomized Controlled Trial.

[275] Heisler LK, Chu HM, Brennan TJ, Danao JA, Bajwa P, et al. (1998) Elevated anxiety and antidepressant-like responses in serotonin 5-ht1a receptor mutant mice. Proc Natl Acad Sci U S A 95:15049–54. 0027-8424 (Print) Journal Article.

[276] Eison AS, Eison MS (1994) Serotonergic mechanisms in anxiety. Prog Neuropsychopharmacol Biol Psychiatry 18:47–62. 0278-5846 (Print) Journal Article Review.

[277] Kennett GA, Whitton P, Shah K, Curzon G (1989) Anxiogenic-like effects of mcpp and tfmpp in animal models are opposed by 5-ht1c receptor antagonists. Eur J Pharmacol 164:445–54. 0014-2999 (Print) Journal Article.

[278] Riedel WJ, Klaassen T, Griez E, Honig A, Menheere PP, et al. (2002) Dissociable hormonal, cognitive and mood responses to neuroendocrine challenge: evidence for receptor-specific serotonergic dysregulation in depressed mood. Neuropsychopharmacology 26:358–67. 0893-133X (Print) Clinical Trial Controlled Clinical Trial Journal Article.

[279] Ni YG, Miledi R (1997) Blockage of 5ht2c serotonin receptors by fluoxetine (prozac). Proc Natl Acad Sci U S A 94:2036–40. 0027-8424 (Print) Journal Article.

[280] Clenet F, De Vos A, Bourin M (2001) Involvement of 5-ht(2c) receptors in the anti-immobility effects of antidepressants in the forced swimming test in mice. Eur Neuropsychopharmacol 11:145–52. 0924-977X (Print) Journal Article.

[281] Xie E, Zhu L, Zhao L, Chang LS (1996) The human serotonin 5-ht2c receptor: complete cdna, genomic structure, and alternatively spliced variant. Genomics 35:551–61. 0888-7543 (Print) Journal Article.

[282] Niswender CM, Copeland SC, Herrick-Davis K, Emeson RB, Sanders-Bush E (1999) Rna editing of the human serotonin 5-hydroxytryptamine 2c receptor silences constitutive activity. J Biol Chem 274:9472–8. 0021-9258 (Print) Journal Article.

[283] Niswender CM, Sanders-Bush E, Emeson RB (1998) Identification and characterization of rna editing events within the 5-ht2c receptor. Ann N Y Acad Sci 861:38–48. 0077-8923 (Print) Journal Article Review.

[284] Gurevich I, Tamir H, Arango V, Dwork AJ, Mann JJ, et al. (2002) Altered editing of serotonin 2c receptor pre-mrna in the prefrontal cortex of depressed suicide victims. Neuron 34:349–56. 0896-6273 (Print) Journal Article.

[285] Pekkarinen P, Terwilliger J, Bredbacka PE, Lonnqvist J, Peltonen L (1995) Evidence of a predisposing locus to bipolar disorder on xq24-q27.1 in an extended finnish pedigree. Genome Res 5:105–15. 1088-9051 (Print) Journal Article.

[286] Lappalainen J, Zhang L, Dean M, Oz M, Ozaki N, et al. (1995) Identification, expression, and pharmacology of a cys23-ser23 substitution in the human 5-ht2c receptor gene (htr2c). Genomics 27:274–9. 0888-7543 (Print) Journal Article.

[287] Fentress HM, Grinde E, Mazurkiewicz JE, Backstrom JR, Herrick-Davis K, et al. (2005) Pharmacological properties of the cys23ser single nucleotide polymorphism in human 5-ht2c receptor isoforms. Pharmacogenomics J 5:244–54. 1470-269X (Print) Journal Article.

[288] Lerer B, Macciardi F, Segman RH, Adolfsson R, Blackwood D, et al. (2001) Variability of 5-ht2c receptor cys23ser polymorphism among european populations and vulnerability to affective disorder. Mol Psychiatry 6:579–85. 1359-4184 (Print) Journal Article Multicenter Study.

[289] Oruc L, Verheyen GR, Furac I, Jakovljevic M, Ivezic S, et al. (1997) Association analysis of the 5-ht2c receptor and 5-ht transporter genes in bipolar disorder. Am J Med Genet 74:504–6. 0148-7299 (Print) Journal Article.

[290] Gutierrez B, Fananas L, Arranz MJ, Valles V, Guillamat R, et al. (1996) Allelic association analysis of the 5-ht2c receptor gene in bipolar affective disorder. Neurosci Lett 212:65–7. 0304-3940 (Print) Journal Article.

[291] Gutierrez B, Arias B, Papiol S, Rosa A, Fananas L (2001) Association study between novel promoter variants at the 5-ht2c receptor gene and human patients with bipolar affective disorder. Neurosci Lett 309:135–7. 0304-3940 (Print) Journal Article.

[292] Frisch A, Postilnick D, Rockah R, Michaelovsky E, Postilnick S, et al. (1999) Association of unipolar major depressive disorder with genes of the serotonergic and dopaminergic pathways. Mol Psychiatry 4:389–92. 1359-4184 (Print) Journal Article.

[293] Lentes KU, Hinney A, Ziegler A, Rosenkranz K, Wurmser H, et al. (1997) Evaluation of a cys23ser mutation within the human 5-ht2c receptor gene: no evidence for an association of the mutant allele with obesity or underweight in children, adolescents and young adults. Life Sci 61:PL9–16. 0024-3205 (Print) Journal Article.

[294] Burnet PW, Smith KA, Cowen PJ, Fairburn CG, Harrison PJ (1999) Allelic variation of the 5-ht2c receptor (htr2c) in bulimia nervosa and binge eating disorder. Psychiatr Genet 9:101–4. 0955-8829 (Print) Journal Article.

[295] Westberg L, Bah J, Rastam M, Gillberg C, Wentz E, et al. (2002) Association between a polymorphism of the 5-ht2c receptor and weight loss in teenage girls. Neuropsychopharmacology 26:789–93. 0893-133X (Print) Journal Article.

[296] Quested DJ, Whale R, Sharpley AL, McGavin CL, Crossland N, et al. (1999) Allelic variation in the 5-ht2c receptor (htr2c) and functional responses to the 5-ht2c receptor agonist, m-chlorophenylpiperazine. Psychopharmacology (Berl) 144:306–7. 0033-3158 (Print) Clinical Trial Letter Randomized Controlled Trial.

[297] Deckert J, Meyer J, Catalano M, Bosi M, Sand P, et al. (2000) Novel 5'-regulatory region polymorphisms of the 5-ht2c receptor gene: association study with panic disorder. Int J Neuropsychopharmacol 3:321–325. 1461-1457 (Print) Journal article.

[298] Meyer J, Saam W, Mossner R, Cangir O, Ortega GR, et al. (2002) Evolutionary conserved microsatellites in the promoter region of the 5-hydroxytryptamine receptor 2c gene (htr2c) are not associated with bipolar disorder in females. J Neural Transm 109:939–46. 0300-9564 (Print) Journal Article.

[299] Yuan X, Yamada K, Ishiyama-Shigemoto S, Koyama W, Nonaka K (2000) Identification of polymorphic loci in the promoter region of the serotonin 5-ht2c receptor gene and their association with obesity and type ii diabetes. Diabetologia 43:373–6. 0012-186X (Print) Journal Article.

[300] Reynolds GP, Zhang Z, Zhang X (2003) Polymorphism of the promoter region of the serotonin 5-ht(2c) receptor gene and clozapine-induced weight gain. Am J Psychiatry 160:677–9. 0002-953X (Print) Journal Article.

[301] Reynolds GP, Zhang ZJ, Zhang XB (2002) Association of antipsychotic drug-induced weight gain with a 5-ht2c receptor gene polymorphism. Lancet 359:2086–7. 0140-6736 (Print) Clinical Trial Journal Article.

[302] Miller del D, Ellingrod VL, Holman TL, Buckley PF, Arndt S (2005) Clozapine-induced weight gain associated with the 5ht2c receptor -759c/t polymorphism. Am J Med Genet B Neuropsychiatr Genet 133:97–100. 1552-4841 (Print) Journal Article.

[303] Pooley EC, Fairburn CG, Cooper Z, Sodhi MS, Cowen PJ, et al. (2004) A 5-ht2c receptor promoter polymorphism (htr2c - 759c/t) is associated with obesity in women, and with resistance to weight loss in heterozygotes. Am J Med Genet B Neuropsychiatr Genet 126:124–7. 1552-4841 (Print) Journal Article.

[304] Erwin VG, Hellerman L (1967) Mitochondrial monoamine oxidase. i. purification and characterization of the bovine kidney enzyme. J Biol Chem 242:4230–8. 0021-9258 (Print) Journal Article.

[305] Bach AW, Lan NC, Johnson DL, Abell CW, Bembenek ME, et al. (1988) cdna cloning of human liver monoamine oxidase a and b: molecular basis of differences in enzymatic properties. Proc Natl Acad Sci U S A 85:4934–8. 0027-8424 (Print) Journal Article.

[306] Donnelly CH, Murphy DL (1977) Substrate- and inhibitor-related characteristics of human platelet monoamine oxidase. Biochem Pharmacol 26:853–8. 0006-2952 (Print) Journal Article.

[307] Thorpe LW, Westlund KN, Kochersperger LM, Abell CW, Denney RM (1987) Immunocytochemical localization of monoamine oxidases a and b in human peripheral tissues and brain. J Histochem Cytochem 35:23–32. 0022-1554 (Print) Journal Article.

[308] Chen K, Wu HF, Shih JC (1993) The deduced amino acid sequences of human platelet and frontal cortex monoamine oxidase b are identical. J Neurochem 61:187–90. 0022-3042 (Print) Journal Article.

[309] Levy ER, Powell JF, Buckle VJ, Hsu YP, Breakefield XO, et al. (1989) Localization of human monoamine oxidase-a gene to xp11.23-11.4 by in situ hybridization: implications for norrie disease. Genomics 5:368–70. 0888-7543 (Print) Journal Article.

[310] Chen ZY, Powell JF, Hsu YP, Breakefield XO, Craig IW (1992) Organization of the human monoamine oxidase genes and long-range physical mapping around them. Genomics 14:75–82. 0888-7543 (Print) Journal Article.

[311] Chen ZY, Hotamisligil GS, Huang JK, Wen L, Ezzeddine D, et al. (1991) Structure of the human gene for monoamine oxidase type a. Nucleic Acids Res 19:4537–41. 0305-1048 (Print) Journal Article.

[312] Grimsby J, Chen K, Wang LJ, Lan NC, Shih JC (1991) Human monoamine oxidase a and b genes exhibit identical exon-intron organization. Proc Natl Acad Sci U S A 88:3637–41. 0027-8424 (Print) Journal Article.

[313] Balciuniene J, Syvanen AC, McLeod HL, Pettersson U, Jazin EE (2001) The geographic distribution of monoamine oxidase haplotypes supports a bottleneck during the dispersion of modern humans from africa. J Mol Evol 52:157–63. 0022-2844 (Print) Journal Article.

[314] Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K (2002) Evidence for positive selection and population structure at the human mao-a gene. Proc Natl Acad Sci U S A 99:862–7. 0027-8424 (Print) Journal Article.

[315] Shih JC, Zhu QS, Grimsby J, Chen K (1994) Identification of human monoamine oxidase (mao) a and b gene promoters. J Neural Transm Suppl 41:27–33. 0303-6995 (Print) Journal Article.

[316] Fowler JS, Volkow ND, Wang GJ, Pappas N, Logan J, et al. (1996) Brain monoamine oxidase a inhibition in cigarette smokers. Proc Natl Acad Sci U S A 93:14065–9. 0027-8424 (Print) Journal Article.

[317] Nies A, Robinson DS, Lamborn KR, Lampert RP (1973) Genetic control of platelet and plasma monoamine oxidase activity. Arch Gen Psychiatry 28:834–8. 0003-990X (Print) Journal Article.

[318] Pedersen NL, Oreland L, Reynolds C, McClearn GE (1993) Importance of genetic effects for monoamine oxidase activity in thrombocytes in twins reared apart and twins reared together. Psychiatry Res 46:239–51. 0165-1781 (Print) Journal Article.

[319] Asberg M, Bertilsson L, Martensson B (1984) Csf monoamine metabolites, depression, and suicide. Adv Biochem Psychopharmacol 39:87–97. 0065-2229 (Print) Journal Article Review.

[320] Kochersperger LM, Parker EL, Siciliano M, Darlington GJ, Denney RM (1986) Assignment of genes for human monoamine oxidases a and b to the x chromosome. J Neurosci Res 16:601–16. 0360-4012 (Print) Journal Article.

[321] Brunner HG, Nelen MR, van Zandvoort P, Abeling NG, van Gennip AH, et al. (1993) X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. Am J Hum Genet 52:1032–9. 0002-9297 (Print) Journal Article.

[322] Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase a. Science 262:578–80. 0036-8075 (Print) Journal Article.

[323] Buchsbaum MS, Coursey RD, Murphy DL (1976) The biochemical high-risk paradigm: behavioral and familial correlates of low platelet monoamine oxidase activity. Science 194:339–41. 0036-8075 (Print) Journal Article.

[324] Caspi A, McClay J, Moffitt TE, Mill J, Martin J, et al. (2002) Role of genotype in the cycle of violence in maltreated children. Science 297:851–4. 1095-9203 (Electronic) Journal Article.

[325] Sabol SZ, Hu S, Hamer D (1998) A functional polymorphism in the monoamine oxidase a gene promoter. Hum Genet 103:273–9. 0340-6717 (Print) Journal Article.

[326] Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in x-linked gene expression in females. Nature 434:400–4. 1476-4687 (Electronic) Journal Article.

[327] Black GC, Chen ZY, Craig IW, Powell JF (1991) Dinucleotide repeat polymorphism at the maoa locus. Nucleic Acids Res 19:689. 0305-1048 (Print) Journal Article.

[328] Kristiansen M, Knudsen GP, Bathum L, Naumova AK, Sorensen TI, et al. (2005) Twin study of genetic and aging effects on x chromosome inactivation. Eur J Hum Genet 13:599–606. 1018-4813 (Print) Journal Article Twin Study.

[329] Joachim CL, Morris JH, Selkoe DJ (1989) Diffuse senile plaques occur commonly in the cerebellum in alzheimer's disease. Am J Pathol 135:309–19. 0002-9440 (Print) Journal Article.

[330] Wood JG, Mirra SS, Pollock NJ, Binder LI (1986) Neurofibrillary tangles of alzheimer disease share antigenic determinants with the axonal microtubule-associated protein tau (tau). Proc Natl Acad Sci U S A 83:4040–3. 0027-8424 (Print) Journal Article.

[331] Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, et al. (1995) Cloning of a gene bearing missense mutations in early-onset familial alzheimer's disease. Nature 375:754–60. 0028-0836 (Print) Journal Article.

[332] Levy-Lahad E, Wijsman EM, Nemens E, Anderson L, Goddard KA, et al. (1995) A familial alzheimer's disease locus on chromosome 1. Science 269:970–3. 0036-8075 (Print) Journal Article.

[333] Kukull WA, Bowen JD (2002) Dementia epidemiology. Med Clin North Am 86:573–90. 0025-7125 (Print) Journal Article Review.

[334] Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, et al. (2002) Dementia and alzheimer disease incidence: a prospective cohort study. Arch Neurol 59:1737–46. 0003-9942 (Print) Journal Article.

[335] Kirschner RJ, Goldberg AL (1983) A high molecular weight metalloendoprotease from the cytosol of mammalian cells. J Biol Chem 258:967–76. 0021-9258 (Print) Journal Article.

[336] Fakhrai-Rad H, Nikoshkov A, Kamel A, Fernstrom M, Zierath JR, et al. (2000) Insulin-degrading enzyme identified as a candidate diabetes susceptibility gene in gk rats. Hum Mol Genet 9:2149–58. 0964-6906 (Print) Journal Article.

[337] Kurochkin IV, Goto S (1994) Alzheimer's beta-amyloid peptide specifically interacts with and is degraded by insulin degrading enzyme. FEBS Lett 345:33–7. 0014-5793 (Print) Journal Article.

[338] Farris W, Mansourian S, Chang Y, Lindsley L, Eckman EA, et al. (2003) Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain in vivo. Proc Natl Acad Sci U S A 100:4162–7. 0027-8424 (Print) Journal Article.

[339] Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, et al. (1999) A full genome scan for late onset alzheimer's disease. Hum Mol Genet 8:237–45. 0964-6906 (Print) Journal Article.

[340] Bertram L, Blacker D, Mullin K, Keeney D, Jones J, et al. (2000) Evidence for genetic linkage of alzheimer's disease to chromosome 10q. Science 290:2302–3. 0036-8075 (Print) Journal Article.

[341] Myers A, Holmans P, Marshall H, Kwon J, Meyer D, et al. (2000) Susceptibility locus for alzheimer's disease on chromosome 10. Science 290:2304–5. 0036-8075 (Print) Journal Article.

[342] Ertekin-Taner N, Graff-Radford N, Younkin LH, Eckman C, Baker M, et al. (2000) Linkage of plasma abeta42 to a quantitative locus on chromosome 10 in late-onset alzheimer's disease pedigrees. Science 290:2303–4. 0036-8075 (Print) Journal Article.

[343] Ait-Ghezala G, Abdullah L, Crescentini R, Crawford F, Town T, et al. (2002) Confirmation of association between d10s583 and alzheimer's disease in a case–control sample. Neurosci Lett 325:87–90. 0304-3940 (Print) Journal Article.

[344] Abraham R, Myers A, Wavrant-DeVrieze F, Hamshere ML, Thomas HV, et al. (2001) Substantial linkage disequilibrium across the insulin-degrading enzyme locus but no association with late-onset alzheimer's disease. Hum Genet 109:646–52. 0340-6717 (Print) Journal Article Multicenter Study.

[345] Boussaha M, Hannequin D, Verpillat P, Brice A, Frebourg T, et al. (2002) Polymorphisms of insulin degrading enzyme gene are not associated with alzheimer's disease. Neurosci Lett 329:121–3. 0304-3940 (Print) Journal Article.

[346] Edland SD, Wavrant-De Vriese F, Compton D, Smith GE, Ivnik R, et al. (2003) Insulin degrading enzyme (ide) genetic variants and risk of alzheimer's disease: evidence of effect modification by apolipoprotein e (apoe). Neurosci Lett 345:21–4. 0304-3940 (Print) Journal Article.

[347] Tivol EA, Shalish C, Schuback DE, Hsu YP, Breakefield XO (1996) Mutational analysis of the human maoa gene. Am J Med Genet 67:92–7. 0148-7299 (Print) Journal Article.

[348] Atmaca M, Kuloglu M, Tezcan E, Ustundag B (2003) Serum leptin and triglyceride levels in patients on treatment with atypical antipsychotics. J Clin Psychiatry 64:598–604. 0160-6689 (Print) Journal Article.

[349] Jansson M, Gatz M, Berg S, Johansson B, Malmberg B, et al. (2003) Association between depressed mood in the elderly and a 5-htr2a gene variant. Am J Med Genet B Neuropsychiatr Genet 120:79–84. 1552-4841 (Print) Journal Article Twin Study.

[350] Cremers TI, Giorgetti M, Bosker FJ, Hogg S, Arnt J, et al. (2004) Inactivation of 5-ht(2c) receptors potentiates consequences of serotonin reuptake blockade. Neuropsychopharmacology 29:1782–9. 0893-133X (Print) Journal Article.

[351] Blomqvist ME, Chalmers K, Andreasen N, Bogdanovic N, Wilcock GK, et al. (2005) Sequence variants of ide are associated with the extent of beta-amyloid deposition in the alzheimer's disease brain. Neurobiol Aging 26:795–802. 0197-4580 (Print) Clinical Trial Journal Article.

[352] Blomqvist ME, Silburn PA, Buchanan DD, Andreasen N, Blennow K, et al. (2004) Sequence variation in the proximity of ide may impact age at onset of both parkinson disease and alzheimer disease. Neurogenetics 5:115–9. 1364-6745 (Print) Journal Article.

[353] Karamohamed S, Demissie S, Volcjak J, Liu C, Heard-Costa N, et al. (2003) Polymorphisms in the insulin-degrading enzyme gene are associated with type 2 diabetes in men from the nhlbi framingham heart study. Diabetes 52:1562–7. 0012-1797 (Print) Journal Article.

[354] Gu HF, Efendic S, Nordman S, Ostenson CG, Brismar K, et al. (2004) Quantitative trait loci near the insulin-degrading enzyme (ide) gene contribute to variation in plasma insulin levels. Diabetes 53:2137–42. 0012-1797 (Print) Journal Article.