Department of Cell and Molecular Biology (CMB),
Karolinska Institutet, Stockholm, Sweden

# Genetic variation in genes involved in Aβ-degradation in Alzheimer disease

Marianne ‘Mia’ Blomqvist


Karolinska Institutet

Stockholm 2009

Published by Karolinska Institutet. Printed by Larserics Digital Print AB, Sundbyberg, Stockholm, Sweden


ISBN 978-91-7409-737-5

*´It takes a Fool to Remain Sane'*
*The Ark, 2000*

*Till Robban: Du är den*
*starkast strålande stjärnan nu.*

# ABSTRACT

Alzheimer disease (AD) is a neurodegenerative disorder that causes dementia among mainly elderly people and is increasing in prevalence in industrialized societies from 1 to 3 % in 60-65 year-olds to as much as 50 % by age 95. This disease is complex. Multiple genes are involved, and they potentially affect each other. The strongest known genetic risk factor for developing AD is the carriage of the ε4-allele of the gene coding for apolipoprotein E (*APOE*) (Strittmatter *et al*., 1993).

This thesis aims to improve the understanding of the genetic basis for AD by comparative re-sequencing of candidate genes for AD in order to discover sequence variants between individuals with or without the disease in the studied population. Comparing such potential differences with known variants is one approach to begin to understand the roles of genes involved in complex diseases. There are today several genes that compete for being the new gene of interest, in addition to *APOE*, as risk factor for AD. One such candidate gene is alpha-3 catenin (*CTNNA3*), which was tested for association with AD in paper I. I genotyped selected markers in connection with *CTNNA3*, and the results indicated an association with AD. Using these markers in a Swedish and a Scottish AD material, we found no association. More tests would however be needed in order to entirely eliminate this gene.

Paper II provides a special focus on a part of chromosome 10q where the gene encoding the protein insulin degrading enzyme (*IDE*) is positioned. The purpose of this study was to investigate if polymorphisms within an LD (linkage disequilibrium) block encompassing *IDE* (Prince *et al*., 2003), might also influence Parkinson's disease (PD), since the quantitative trait age at onset (AAO) has previously shown linkage to *IDE* in both AD and PD (Li *et al*., 2002), suggesting pathogenic alleles influencing both disorders. Our findings pointed towards a linkage disequilibrium (LD) block on chromosome 10q that harbors alleles influencing AD and PD in a similar way, containing the three genes *IDE*, *KIF11* (kinesin family member 11), and *HHEX* (hematopoietically expressed homeobox).

In paper III, the aim was to further investigate an earlier (Prince *et al*., 2003) association between DNA variation in *IDE* and AD pathology and severity. More specifically, the association of *IDE* variants with the density of senile plaques and neurofibrillary tangles was observed, and I therefore attempted

to replicate these findings in an autopsy population in order to measure cerebral amyloid beta (Aβ) levels. This resulted in an observation of significant effects on plasma Aβ levels, which together with no association with direct disease risk in AD case-control samples of one specific single nucleotide polymorphism (SNP), suggested that IDE may have a role as modifier of severity of disease rather than risk.

The aim of paper IV was to examine whether abundant data could exist that might help resolve the role of gene(s) in complex phenotypes, but that many reasonably powered studies producing negative results are not being published. A literature study of association studies based on candidate genes in AD confirmed a publication bias. We also tested 62 genetic markers for association with AD risk in addition to possible effects upon quantitative indices of AD severity, resulting in only modest signals, i.e. the overall results were mostly negative.

Identification of the underlying genes for complex human diseases can strengthen our knowledge about disease causes and possibly suggest potential treatment methods. This strategy could also increase our knowledge about gene function at a molecular level by providing new information about changes in DNA-sequences that can influence function of genes. Ultimately, the identification of disease contributing genes could clear the way for more personally formulated methods of treatment.

Key words: Alzheimer disease, SNPs, genotyping, association study, haplotype, linkage disequilibrium, association mapping

# LIST OF PUBLICATIONS

This thesis is based on the following original publications, referred to in the text by their Roman numerals.

**I.** **Blomqvist ME**, Andreasen N, Bogdanovic N, Blennow K, Brookes AJ, Prince JA. Genetic variation in CTNNA3 encoding alpha-3 catenin and Alzheimer's disease. Neurosci Lett. 2004. 358(3):220-2.

**II.** **Blomqvist ME**, Silburn PA, Buchanan DD, Andreasen N, Blennow K, Pedersen NL, Brookes AJ, Mellick GD, Prince JA. Sequence variation in the proximity of IDE may impact age at onset of both Parkinson disease and Alzheimer disease. Neurogenetics. 2004. 5(2):115-9.

**III.** **Blomqvist ME**, Chalmers K, Andreasen N, Bogdanovic N, Wilcock GK, Cairns NJ, Feuk L, Brookes AJ, Love S, Blennow K, Kehoe PG, Prince JA. Sequence variants of IDE are associated with the extent of beta-amyloid deposition in the Alzheimer's disease brain. Neurobiol Aging. 2005. 26(6):795-802.

**IV.** **Blomqvist ME**, Reynolds C, Katzov H, Feuk L, Andreasen N, Bogdanovic N, Blennow K, Brookes AJ, Prince JA. Towards compendia of negative genetic association studies: an example for Alzheimer disease. Hum Genet. 2006. 119(1-2):29-37.

# POPULÄRVETENSKAPLIG SAMMANFATTNING

## Bakgrund

Alzheimers sjukdom är en komplex sjukdom där områden i hjärnan bryts ned med demens som följd, framför allt hos äldre. Den hittills starkast kända genetiska riskfaktorn för sporadisk Alzheimers är när någon bär på ε4-allelen (variant) av genen som kodar för apolipoprotein E (*APOE*) i sitt genom (Strittmatter *et al.,* 1993). Det genetiska bidraget till Alzheimers måste till stor del bekräftas genom studier av ännu oidentifierade så kallade loci (plats där genen finns i ett genom). Den starkaste indikatorn för att flera gener utöver *APOE* är inblandade i sporadisk Alzheimers kommer från en studie av hur olika gener nedärvs (kopplingsstudier) i sent uppträdande Alzheimers (>65 år) hos syskon (Kehoe *et al*., 1999). Denna studie, liksom andra, har kraftigt indikerat att olika områden på kromosom 10 ökar risken för Alzheimers.

Den maternellt respektive paternellt nedärvda dubbla uppsättningen av varje gen i ett genom gör att vi antingen kan ha två lika, eller två olika uppsättningar alleler för varje gen. En individs genotyp utgörs av specifika uppsättningar av alleler. Dessa två uppsättningar av en gen som finns i en normal kroppscell kan alltså vara olika. Individer med två likadana varianter av en gen sägs vara homozygoter med avseende på denna gen. Individer med två olika varianter av genen sägs vara heterozygoter.

För att finna gener med inblandning i komplexa sjukdomar hos människan är det nödvändigt att undersöka stora populationer med hjälp av markörer som förekommer ofta i genomet. Single nucleotide polymorphisms (SNPs) är exempel på sådana markörer där en enda nukleotid har bytts ut mot en annan nukleotid på samma ställe mellan två lika DNA-sekvenser. SNPs är den vanligast förekommande formen av genetisk variation i det mänskliga genomet. Om SNPs fysiska platser är kända, kan man använda dessa som biologiska markörer vid skapandet av genetiska kartor i associationsstudier.

## Mål

Målet är att öka förståelsen för genetiska variationer bakom Alzheimers och även eventuella kopplingar till Parkinsons sjukdom med ett fokus på *insulin degrading enzyme* genen (*IDE*) belägen på kromosom 10. Strategin baseras

på genotypning av SNPs i DNA-material med Alzheimerspatienter och kontroller. För detta ändamål användes DASH-tekniken (dymamic allele-specific hybridization) utvecklad på vårt labb (Prince *et al*., 2001) och TaqMan (Livak *et al*., 1995; Livak, 1999) (ej publicerat material). Via statistiska analyser av givna genotyper samt haplotyper (mönster av genotyper), kartläggs så kallad linkage disequilibium (LD) kring en intressant gen/er för att så kunna styrka eller förkasta exempelvis *IDE* som en sådan gen.

## Forskningsstrategi

Avhandlingen inkluderar framställande av LD-kartor kring intressanta gener genom genotypning av SNPs. LD-kartor ger information om hur olika gener, belägna tillräckligt nära i genomet, inbördes påverkar varandra. Ett projekt med centrering på en gen/region av kromosom 3 där membrane metallo-endopeptidase genen (*MME*) är belägen har också gjorts (ej publicerat). Andra studier har där visat association med Alzheimers (Shi *et al*., 2005; Helisalmi *et al*., 2004) och jag gjorde ett försök att verifiera dessa samband med relevanta markörer i de material vi har tillgång till, dock med negativt resultat.

I mitt fortsatta vetenskapliga arbete kommer fokus att flyttas till ett eventuellt samband mellan depression, stress och AD. Detta samband kan ligga i degraderingsprocesser i hjärnan med anknytning till Alzheimers där bland annat en förhöjd kortisolnivå hos AD-patienter genom utsöndring av glukokortikoider samt stress kan vara involverat (Catania *et al*., 2009). Samtidigt kommer parallella studier över *IDE* att fortsätta, liksom studier över andra potentiella kandidatgener (Blomqvist *et al*., 2004; Kehoe *et al*., 2004).

Haplotyper relateras till studerade fenotyper (d v s proteinnivåer kopplade till Alzheimers i DNA-material, genetiskt bestämda av den dominanta allelen i ett allelpar nedärvda från föräldrarna) genom den så kallade partition ligation algoritmen (HAPLOTYPER; Niu *et al*., 2002). Det unika i denna avhandling, förutom DASH-tekniken, innefattar tillgång till flera kliniska/epidemiologiska resurser redovisade i mina publikationer.

## Resultat

I en studie baserad på sekvensvariation under inflytande av *IDE* (Blomqvist, *et al*., 2004), erhölls signifikant association ($P<0.05$) med ålder vid vilken sjukdomen först påvisats (påslagsålder) i en australiensk fall-kontrollstudie. Förutom denna studie har jag även genotypat utvalda markörer i anslutning till genen som kodar för alpha-3 catenin (*CTNNA3*), från vilken indikationer getts om association med Alzheimers. Efter test med dessa markörer i ett svenskt respektive skotskt Alzheimersmaterial (fall-kontroll), fann vi ingen association. Fler test skulle dock behövas för att helt utesluta denna gen. Jag har också testat 32 markörer i och i nära anslutning till *IDE* för några olika kvantitativa karaktäristika. Strategin i denna studie var att mäta LD och därefter testa ett urval av dessa markörer för olika fenotyper, för att på så vis få en uppfattning om association med Alzheimers. Resultatet från denna studie indikerade att olika alleler av *IDE* bidrar till variationen i amyloid betanivåer i hjärnan på individer diagnostiserade med Alzheimers och att detta förhållande kan ha betydelse för dysfunktionsgrad hos dessa individer.

## Signifikans

Identifiering av de bakomliggande generna för existerande mänskliga sjukdomar kan stärka vår kunskap om deras ursprung och föreslå behandlingsformer. Denna strategi skulle också öka vår kunskap om hur gener fungerar på en molekylär nivå och ge ny information om DNA-sekvensförändringar som kan påverka genfunktioner. Ultimat kan identifiering av sjukdomsbidragande gener och uppsättningar av sådana bana väg för mer personligt utformade behandlingsmetoder.

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AAO | age at onset |
| Aβ | amyloid beta |
| ABCA1 | ATP-binding cassette, sub-family A member 1 |
| ACE | angiotensin converting enzyme |
| AD | Alzheimer disease |
| AFLP | amplified fragment length polymorphisms |
| ANOVA | analysis of variance |
| APOE | apolipoprotein E |
| APP | amyloid precursor protein |
| C99 | C-terminal fragment 99 |
| CNV | copy number variation |
| CTNNA3 | alpha-3 catenin |
| CSF | cerebrospinal fluid |
| CpG | 'p' indicates a phosphodiester bond between a C and a G nucleotide |
| DASH | dynamic allele-specific hybridization |
| dbSNP | databaseSNP – NCBI's SNP database |
| DJ-1 | (or PARK7) Parkinson disease (autosomal recessive, early onset) 7 |
| DNA | deoxyribonucleic acid |
| ε2/ε3/ε4 | epsilon 2/epsilon 3/epsilon 4 |
| ECE | endothelin converting enzyme |
| EOAD | early onset Alzheimer disease |
| FAD | familial Alzheimer disease |
| GC | glucocorticoids |
| HHEX | hematopoietically expressed homeobox |
| HWE | Hardy-Weinberg equilibrium |
| IDE | insulin degrading enzyme |
| KIF11 | kinesin family member 11 |
| LD | linkage disequilibrium |
| LOAD | late onset Alzheimer disease |
| MAPT | microtubule-associated protein tau |
| MME | membrane metallo-endopeptidase |
| MMSE | mini-mental state examination |
| MRI | magnetic resonance imaging |
| NEP | neprilysin |
| NOS1 | nitric oxide synthase 1 |

| | |
|---|---|
| PARK2 | Parkinson disease (autosomal recessive, juvenile) 2, parkin |
| PET | positron emission tomography |
| PCR | polymerase chain reaction |
| PD | Parkinson disease |
| PSEN1 | presenilin 1 |
| PSEN2 | presenilin 2 |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| SNCA | synuclein, alpha (non A4 component of amyloid precursor) |
| SNP | single nucleotide polymorphism |
| SPECT | single photon emission computed tomography |
| SP-NFT | senile plaque-neurofibrillary tangles |
| SSR | simple sequence repeat |
| UCHL-1 | ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) |

# 1 BACKGROUND

All humans are unique! Despite this fact, there is amazingly little that actually differs between us, or even between us and certain other species, at least genetically speaking. In order to maintain a living organism, there are some essential functions that need to work properly. The ability to transmit and store specific instructions for how to construct the physical body is such an essentiality. All living organisms are also dependent on interactions with the surrounding environment. To succeed in this, an organism needs to have knowledge on how to do so. This information is stored in the organism's deoxy-ribonucleic acid (DNA), a chemical structure consisting of four building blocks; the purine nucleotides adenine (A) and guanine (G) together with the pyrimidine nucleotides cytosine (C) and thymine (T). In 1953, Watson and Crick worked out the structure of DNA, a discovery for which they were later honored with the Nobel Prize.

The four nucleotide bases are connected with strong covalent bonds along a single strand, with weaker hydrogen bonds pairing C with G, and A with T, thus connecting the two strands. Oriented in opposite directions, each single strand has two different ends referred to as 5' and 3'. The DNA makes up a strand-like structure that is rolled up like a rope on *histones* in eukaryotic organisms. The histones can be organized into *chromosomes* in the shape of X's (figure I) during mitosis or meiosis.

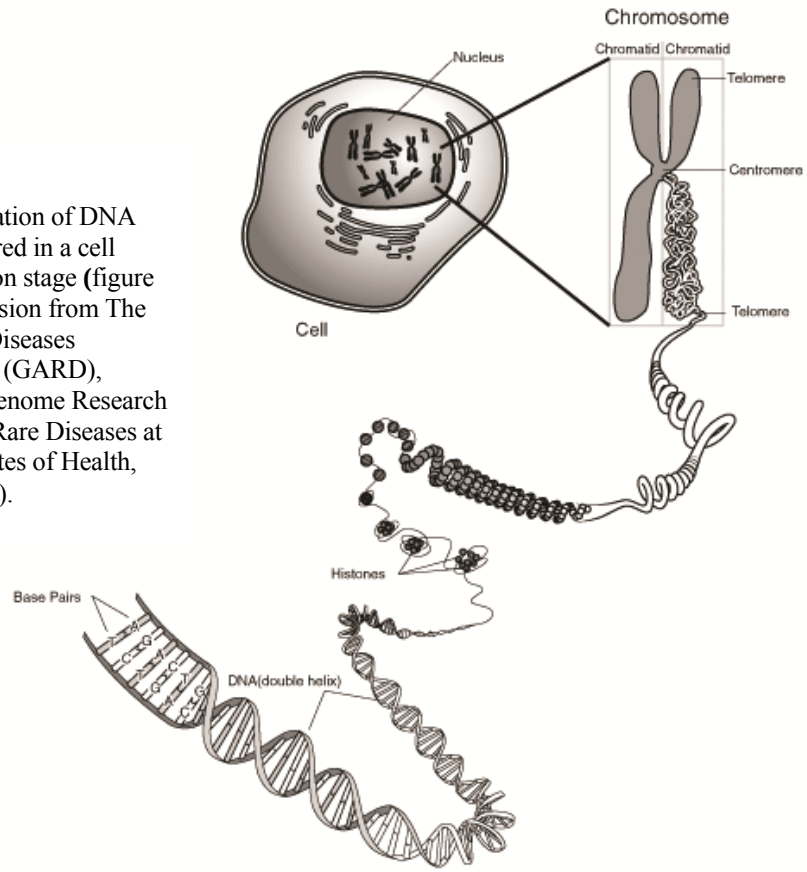


**Figure I.**
Schematic visualization of DNA structure that is stored in a cell during a cell division stage **(**figure is used with permission from The Genetic and Rare Diseases Information Center (GARD), National Human Genome Research Institute/Office of Rare Diseases at the National Institutes of Health, Gaithersburg, USA).

Chromosomes hold genetic information in the form of *genes* where each gene is a set of biochemical instructions that tells a cell how to assemble one of many different proteins (figure II). Almost all genes contain alternating regions called exons, including protein-coding sequences, and introns, separating neighboring exons in a gene.

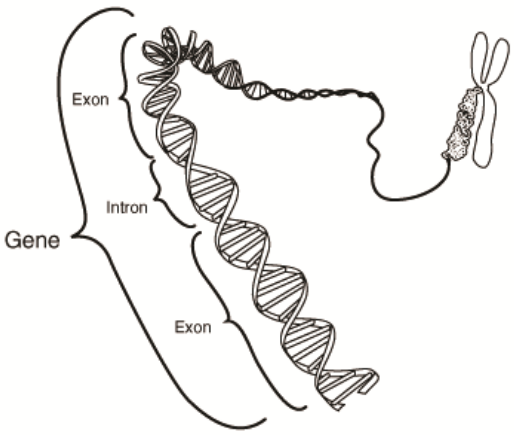


**Figure II.**
Schematic visualization of a gene **(**figure is used with permission from The Genetic and Rare Diseases Information Center (GARD), National Human Genome Research Institute/Office of Rare Diseases at the National Institutes of Health, Gaithersburg, USA).

A *locus* is a place on a specific chromosome where a gene is found and an *allele* is a specific variant of a gene on a chromosome. Since each person receives two alleles of a gene, one from each parent, alleles provide two forms of the same gene. Two or more alleles can shape each human trait. In a healthy individual, the human genome consists of twenty-two pairs of chromosomes and one pair of sex-chromosomes, one inherited from the mother, the other from the father, giving most individuals two copies of every gene. The chromosomes contain between 20,000 and 25,000 protein-coding genes ((19,599 has been confirmed by consortium researchers) Int Hum Genome Seq Consortium, Nature, 2001; 2004).

Differences between alleles are all related to their building blocks (i.e. nucleotides), being substituted, added, or excluded. Due to the double sets of each gene within a genome, we could either have two equal, or two different sets of alleles for every gene. Specific sets of alleles make up an individual's genotype. Having the same sets of alleles, two individuals have the same genotype for a particular locus.

With the Human Genome project (HGP), led by the National Institutes of Health, the Department of Energy, USA, and the Wellcome Trust, as well as the International HapMap Project (Phase II), led by the International HapMap Consortium (i.e. a collaboration among scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States (The International HapMap Consortium, 2003; 2005)) the spotlight was put on the importance of knowledge about naturally occurring variation in the genome, including single nucleotide polymorphisms (SNPs). SNPs involve a single substitution of a nucleotide with a different nucleotide in the same position between two equivalent DNA sequences. SNPs are the most common form of genetic variations in the human genome and therefore very usful as genetic markers when one wants to search for genes potentially involved in e.g. Alzheimer disease (see also *SNP genotyping*, pp 29-31).

There are multiple ways to determine the genotype of an individual. One such method to detect SNPs, Dynamic Allele-Specific Hybridization (DASH), was developed in our laboratory. This method was employed in this thesis.

To make a diagnosis of a severe and multi-factorial complex disease such as AD, or Parkinson disease (PD), is difficult and can be complicated. Previously, the only way to verify AD in a patient was through autopsy of

the brain and looking for plaque formation in a microscope. Today, the final verification of AD is still performed in this way, but with the addition of a battery of sophisticated dementia tests, measuring of particular protein levels in cerebrospinal fluid samples (i.e. tau and amyloid beta (Aβ) proteins), and PET (positron emission tomography) or SPECT (single photon emission computed tomography) scanning. Diagnosing both AD and PD always starts with a review of the patient's disease history, often together with a relative of the patient. This is done in order to rule out other scenarios such as exhaustive depression or major depression. Clinical tests can be performed for both AD and PD, which in the latter case involves testing muscle tension, reflexes, and balance.

The result presented in this thesis aims to provide better understanding of the genetic basis for AD, and possibly connections to PD. A special focus is placed on a part of chromosome 10q where the gene encoding the protein insulin degrading enzyme is located (figure XIII). The project involved, in addition to genotyping with the DASH method (Prince *et al*., 2001), the former genotyping facility KISNP (The Karolinska Institutet Single Nucleotide Polymorphism Research Centre), and access to a large number of human DNA samples.

# 2 ALZHEIMER DISEASE

## 2.1. Characteristics

Alzheimer disease is a neurodegenerative disorder which causes dementia in mainly elderly people and was first described by the German physician Alois Alzheimer in 1907 (Alzheimer, 1907; Jarvik, *et al*., 1987). AD increases in prevalence in industrialized societies from 1 to 3 % in 60-65 year-olds to as much as 50 % by age 95 (Fratiglioni *et al*., 1991). Multiple genes are involved in causing the disease and these genes may have mutual effects on each other (Blomqvist *et al*., 2004; Blomqvist *et al*., 2005).

AD is an amyloidosis, a group of illnesses resulting from the failure of certain proteins to fold and function properly, resulting in toxic accumulations of these proteins. Characteristic neuropathological hallmarks of AD are changes in brain tissue, including (senile) plaques, i.e. deposits of the peptide β-amyloid (Glenner *et al*., 1984; Masters *et al*., 1985), and

neurofibrillary tangles (Terry, 1963), hyperphosphorylated forms of the microtubular protein tau that clog up the nerve cells in the brain cortex, causing these cells to deteriorate. An AD brain is physically decreasing in size during the course of the disease (figure III). The clinical characteristics involve a gradual memory loss with word-finding difficulty and forgetfulness as early signs.

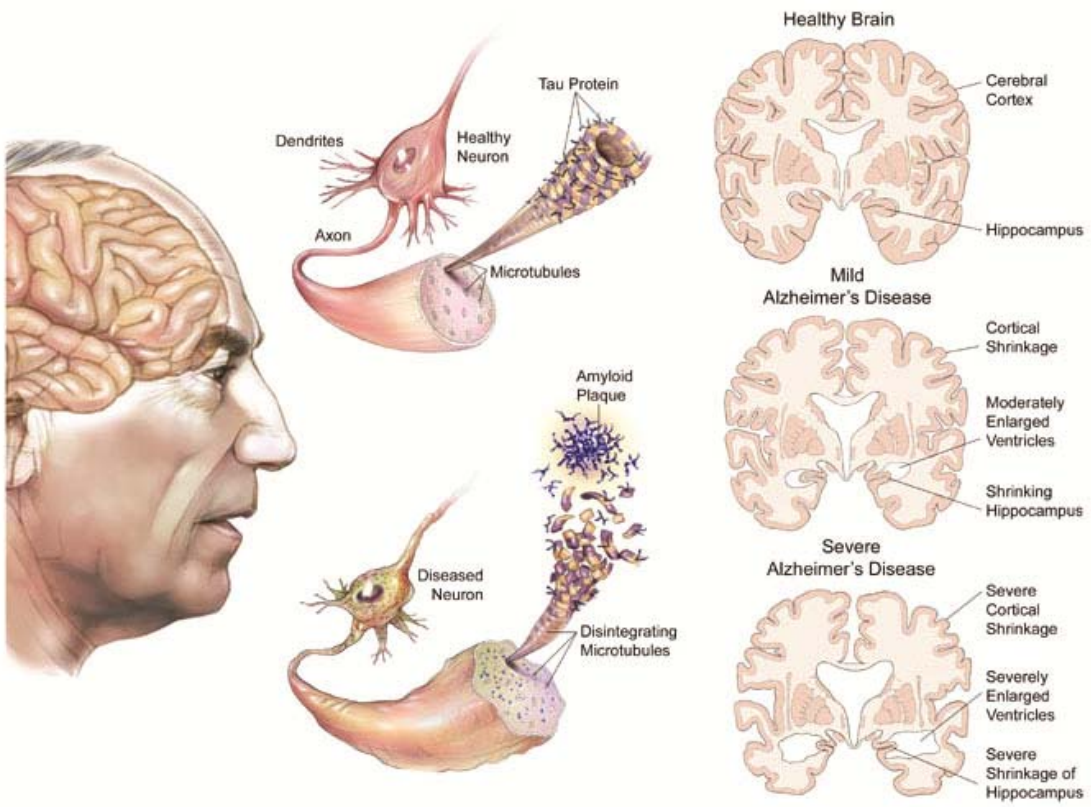


**Figure III.** Two of the most significant neuropathological hallmarks of AD are neurofibrillary tangles, caused by the protein tau, and amyloid plaques (also called senile plaques), caused by the aggregation of amyloid beta. The AD affected brain is degenerated by long-time effect of those plaques and tangles. © Robert F. Morreale, CMI.

## 2.2. Sporadic and familial AD

There are two basic types of AD, sporadic and familial (FAD). The latter form is the more rare one, affecting less than 2 % of patients with AD, all with an early-onset (<65 yr) of the disease (Finckh *et al*., 2000; Campion *et al*., 1999). Genetic mutations involved in FAD are situated on chromosome

1 (Levy-Lahad, *et al*., 1995; Rogaev *et al*., 1995), 14 (Sherrington *et al*., 1995), and 21 (Levy *et al*., 1990; Goate *et al.*, 1991). These genes are presenilin 1 and 2 (*PSEN1, PSEN2*), and amyloid precursor protein (*APP*), respectively. Late onset AD (LOAD) (usually >65 yr) has so far not shown any specific inheritance pattern, while FAD shows an autosomal dominant inheritance pattern. Thus, if one of the parents have FAD, their offspring has a 50 % risk of developing FAD (Farrer *et al*., 1990).

## 2.3. Risk factors

### *2.3.1. Genetic*

AD is a multi-factorial complex disease where both genetic and environmental factors can be involved. More than one gene mutation from genes on multiple chromosomes are involved. The only widely recognized genetic risk factor to date, even if it is neither necessary nor sufficient to cause the disease, is the ε4 allele of the APOE (apolipoprotein E) gene situated on chromosome 19, one of several different forms, or alleles, of this gene (Strittmatter *et al*., 1993). With a relative risk exceeding 2, this common variant has an allele frequency >10 % (Corder *et al*., 1993). Three other genes have also been shown to play a role in the development of AD, namely *APP*, *PSEN1*, and *PSEN2* (Levy *et al*., 1990, Goate *et al*., 1991; Sherrington *et al*., 1995; Levy-Lahad, *et al*., 1995, Rogaev *et al*., 1995). These genes have been demonstrated to be responsible for autosomal dominant AD, particularly prominent in EOAD.

All four genes share influence on the metabolism of amyloid beta (Aβ), together with promoting the accumulation of Aβ in senile plaques (Tanzi *et al*., 2001). *APOE*, located on chromosome 19, codes for a protein involved in cholesterol transport in the bloodstream (Poirier, 1994). *APOE* has three common alleles; ε2, ε3, and ε4. Thus, the combinations that can exist in one individual are ε2/2, ε2/3, ε2/4, ε3/3, ε3/4, or ε4/4. An individual with two copies of the ε4 allele belongs to the highest risk group. The most frequently occurring allele form in the general population is the ε3, suspected to play a neutral role in AD. Finally, the ε2 allele seems to be associated with a lower risk of AD (Strittmatter *et al*., 1993). In well-defined populations, the ancestral ε4 allele has a frequency varying between 0.052-0.407 in Europeans, and 0.089-0.280 in Native Americans. There seems to be a higher frequency in populations that still have an economy based on

foraging, or where food supplies were recently scarce, for example aborigines in Malaysia (0.240) and Australia (0.260), or Lapps (0.310). The ε2 allele frequency shows no specific patterns (0.145-0.02), and this allele is even absent in e.g. Native Americans (Corbo *et al*., 1999).

In addition, there are now multiple suggestions of other chromosomes and genes involved in AD, among which are chromosome 9 with ATP-binding cassette, sub-family A member 1 (*ABCA1*) (Katzov *et al*., 2004), chromosome 10 with insulin degrading enzyme (*IDE*) (Kurochkin *et al*., 1994), and chromosome 12 with nitric oxide synthase 1 (*NOS1*) (Galimberti *et al*., 2005). Other examples of proteases and putative candidates for the process where Aβ is degraded and eliminated are neprilysin (NEP) (Iwata *et al*., 2000), endothelin converting enzyme (ECE) (Eckman *et al*., 2001), and angiotensin converting enzyme (ACE) (Hu *et al*., 2001).

### *2.3.2. Amyloid beta plaque, neurofibrillary tangles, and proteolytic processing of the amyloid precursor protein*

To date, there is still no known normal function of the membrane protein amyloid precursor protein (APP). This protein is synthesized in the endoplasmatic reticulum, and after post-translational modifications, it matures in the Golgi apparatus. APP is transported from the cell body to the nerve endings by axonal transport in neurons (Koo *et al*., 1990).

Amyloid plaques are, together with neurofibrillary tangles, one of the neuropathological hallmarks of AD in an AD-affected brain (figure III). Amyloid plaques, or senile plaques, mainly consist of 39-42 amino acid long Aβ peptides, and are found in the spaces between nerve cells in the brain.

Aβ is the product formed after sequential cleavage of APP by either α-, β-, or γ-secretases. In a pathogenic pathway (amylogenic), proteolytic cleavage is facilitated by β-secretase, releasing the extracellular part of APP, whereafter γ-secretase cleaves within the transmembrane region of APP, rapidly releasing the C-terminal end of the Aβ peptide (figure IV). In addition to Aβ, a soluble APPβ (sAPPβ) is formed and released. The third secretase, α-secretase, is only involved in a benign pathway (non-amylogenic), and it cleaves, together with γ-secretase, the portion of APP that has the potential to become Aβ. These cleavages release a small segment called p3 and the N-terminal part of APP, termed soluble APP

alpha (sAPPα). These cleavages eliminate the Aβ peptide and at the same time any potential buildup of plaques. The two most common isoforms of Aβ are $A\beta_{40}$ and $A\beta_{42}$. While the former is the more common form, $A\beta_{42}$, being hydrophobic, is more fibrillogenic and therefore associated with disease (Hardy, Allsop, 1991).

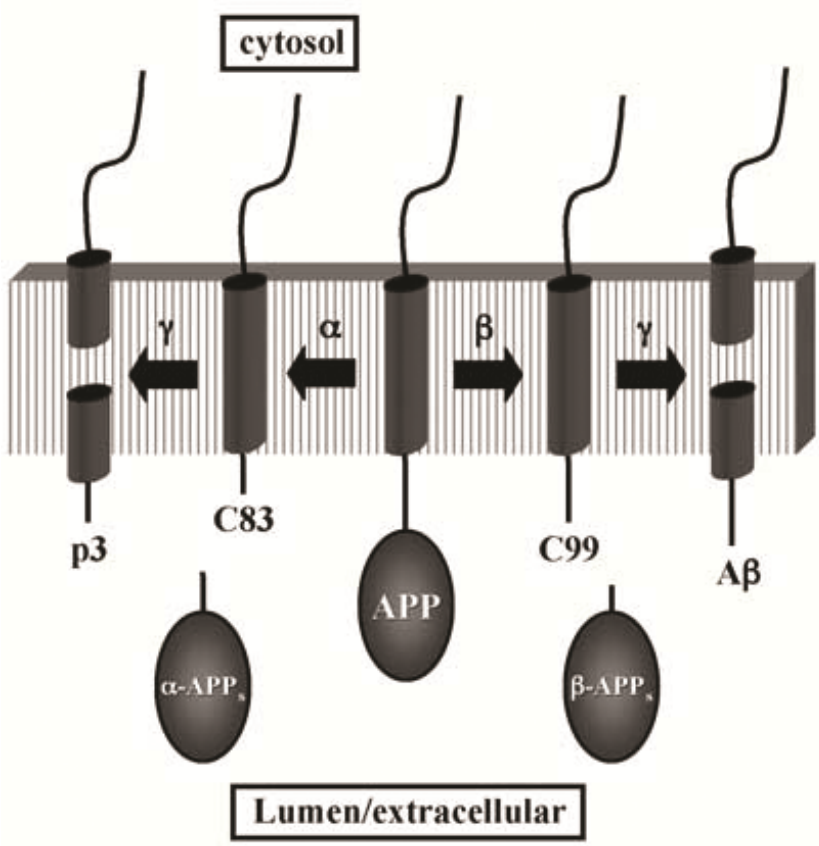


**Figure IV.** Processing of APP. There are two possible pathways when APP is being cleaved by secretases – either by α- and γ-secretases, resulting in a non-amylogenic outcome (p3), or by β- and γ-secretases, resulting in a toxic amylogenic outcome (Aβ). s=soluble. ©NCBI/NLM/NIH/gov/bookshelf.

Neurofibrillary tangles are deposits that are mainly composed of the cytoskeletal protein tau (figure III). Tau proteins normally have a number of phosphate molecules attached to them, and they together have a stabilizing effect when binding to microtubules. In the case of mutations in tau, an abnormally large amount of phosphate molecules are attached to the tau protein, which causes hyperphosphorylation and disengaging from microtubules. In addition, tau proteins then seem to cluster with other tau threads, which results in the formation of neurofibrillary tangles. The

microtubules can subsequently disintegrate and cause the internal transport system of the neurons to collapse (Tiraboschi, *et al*., 2004).

### *2.3.3. Environmental*

In addition to genetic risk factors, there are some indications that environmental factors could potentially contribute to AD. Examples of factors that have been proposed to contribute to AD are head trauma (Roberts *et al*., 1994), dietary cholesterol and/or fat (Michikawa, 2003; Notkola *et al*., 1998; Kivipelto *et al*., 2001; Levin-Allerhand *et al*., 2002), aluminum (Crapper *et al*., 1973; Flaten, 2001)), viral infections (Grant *et al*., 2002), and extremely low frequency magnetic field(s) (ELF-MF) (Huss *et al*., 2009).

Recently, some interesting studies have been published that present a possible connection between depressive symptoms and risk of dementia. One of these studies point to a potential relationship with AD through an association between depression with an early onset and AD (Geerlings *et al*., 2008), while another, focusing on changes in the symptoms of depression during the prodromal phase of AD, showed no such changes (Wilson *et al*., 2008).

Geerlings *et al*. (2008) looked at whether AD risk is mediated by structural brain changes. They examined whether the history of depressive episodes and the occurrence of depressive symptoms were associated with smaller hippocampal and amygdalar volumes and with increased risk for incident AD, using three-dimensional MRI (magnetic resonance imaging). Their conclusion was that smaller amygdalar or hippocampal volumes did not mediate the risk for AD, and neither did the presence of depressive symptoms.

Wilson *et al*. (2008) also looked at how depressive symptoms change during the evolution of dementia and set up a study based on the hypothesis that there is an increase in depressive symptoms during the prodromal phase of AD. As mentioned above, they found no change in the symptoms of depression, but they did notice a slight increase among females relative to men in those with a higher premorbid level openness and a lower level of premorbid agreeableness.

Stress induced release of glucocorticoids (GC) is an additive potential risk factor for AD, either contributing to the development of the disease or to the maintenance of it (Catania *et al*., 2009). In line with this hypothesis, Catania *et al*. showed that GC and stress provoke misprocessing of amyloid precursor protein in rat prefrontal cortex and hippocampal areas. The levels of peptide C-terminal fragment 99 (C99) were thereby increased with further proteolytic cleavage and generation of Aβ as a result. Additionally, they showed that exogenous Aβ can reproduce the effects of stress and GC on C99 production and that a history of stress noticeably potentiates the C99-inducing effects of Aβ and GC.

# 3 GENETIC VARIATION

There are four main primary sources of genetic variation:

1. mutations
2. gene flow
3. recombination
4. sex

*Mutations* are changes at the DNA level, while *gene flow* is movement of genes from one population to another. *Recombination* is the process where a strand of genetic material, DNA or RNA, is broken and subsequently joined to a different DNA molecule. In eukaryotes this commonly occurs during meiosis as a chromosomal crossover between paired chromosomes. *Sex* can introduce new gene combinations into a population and is a form of genetic shuffling. Shuffling can bring together new combinations of genes and is therefore important for evolution (Leder, 1982).

The human genome differs approximately 0.01 % between different individuals in homologous regions, while the corresponding between humans and the ape species *Pan Troglodytes* (chimpanzee) is approximately 1 % (Fujiyama *et al*., 2002). This difference is due to genetic variation that was either inherited from a parent, or acquired during the lifetime of an individual. Mutations that are passed from parent to the child are called hereditary or germline mutations. The genetic differences are obviously very important, from both a biological and a medical perspective. Without any differences, all humans would look the same (which would be rather boring

in the long run) and without any mutations there would not be any evolution (i.e. mutations are not all bad!).

More generally, a mutation is a permanent genetic change in a DNA sequence that makes up a gene and that exists in an individual (or a family) with a frequency <0.01, while a polymorphism is a mutation that has persisted through the generations and spread in a population(s), with a frequency >0.01. Scientists study genome variation in order to improve the knowledge about who we are; how complex human traits have arisen, like our personalities, height, eye color, and susceptibility to diabetes, heart disease, or AD.

## 3.1. Sequence variation

As mentioned above, DNA sequence variation has arisen in pace with the evolution of *Homo sapiens*. The gametes, i.e. sperm and ova, are haploid, containing one of each chromosome. Where all ova have the chromosomal complement 23, X, sperm have either 23, X or 23, Y. The fusion of sperm and ovum to a zygote produces a restored diploid chromosomal complement. The human genome contains about 3 billion bases in its haploid form (Venter *et al*., 2001).

There are several types of DNA sequence variants, which, according to Strachan *et al*. (2003), can be sorted by criteria defined by the physical nature of the sequence variation, by the effect on formation of protein, and by any associated susceptibility to disease. The complete human genome is the diploid complement, i.e. most human cells are diploid with all 46 chromosomes. This can be written as 46, XX in females, and 46, XY in men. Changes of allelic frequencies within a population that are directed occur through the natural selection of a favored genotype. Such changes often lead to homozygosity at a particular locus, but a heterozygote can also be more suited to a given environment than either of the homozygotes. This will lead to a balanced polymorphism. The amount of heterozygosity at a locus in a population, given by the total frequency of heterozygotes at a locus, is a measure of genetic variation (Griffiths *et al*., 2000). Natural selection is an important contributor to human evolution. This process determines which sequence variants that remain over time (Bustamante *et al*., 2005).

## 3.2. Single Nucleotide Polymorphism (SNP)

Single nucleotide polymorphisms, or SNPs, are DNA sequence variants that occur when a single nucleotide base pair has been replaced with another single base pair (figure V). An SNP can be characterized by its allele frequency within a population. To be defined as a polymorphism and fulfill the criteria for an SNP, the frequency of the least common allele in a tested specific population must be at, or above 0.01. SNPs make up approximately 90 % of all genetic variation in the human genome. An SNP occurs at an average spacing of one every 290 base pairs in the human genome (Kruglyak *et al*., 2001). The number of known SNPs, i.e. those with a unique position, was 17,804,034 as of April 30, 2009 (Build130) (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi).

SNPs can have three or four alleles, but the vast majority is diallelic, i.e. having two common alternative bases. One transition C⇔T (A⇔G), and three transversions C⇔A (G⇔T), C⇔G (G⇔C), and T⇔A (A⇔T) (the opposite strand base in parenthesis), make up the four possible types of variants at a single position. The SNP type C⇔T is the most frequent. An example of this is the alteration of the DNA sequence AAACGGA to AAATGGA, where the first 'C' has been replaced with a 'T'. About 2/3 of all SNPs are of this type, while the other three types are found at equal frequencies to comprise the remaining ~1/3. One possibility why C⇔T (A⇔G) SNPs are more common than the other types is the high frequency of 5-methylcytosine deamination reactions that occur in the genome, especially at CpG[1] dinucleotides (Brookes, 1999; Feuk, 2002). The definition of the locus for a polymorphism, or an SNP, is important. In the example shown in figure V the individual is heterozygotic for A/C, or T/G, at the highlighted locus (SNP). It is also important to state the surrounding sequence if the locus is the same at allele 1 and allele 2 in order to define the polymorphism, or SNP, of interest.

Most SNPs occur in non-coding regions of the genome, but some are found in regions that code for proteins, which make them particularly interesting, since these are more likely to alter the biological function of a protein, even if mutations in non-coding regions can affect function by altering regulatory sequences. SNPs that occur in protein-coding regions are either non-

[1] regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide in the linear sequence of bases along its length

synonymous or synonymous (silent), depending on if they do or do not modify the amino acid sequence of the gene product. When SNPs occur in coding sequences, they are more likely to affect the function or availability of a protein than other SNP classes (Carlson *et al*., 2004). 'Hot-spot' regions of the genome are regions where variation is more likely to arise with hundreds of different variants of a sequence. Other parts of the genome are more stable, which means that very little variation between different individuals is present. If a genetic change occurs within a gene, there is a greater risk of disease. Finally, variations in 'extra' or 'junk' DNA regions, other than regulatory elements, are less likely to be harmful since they do not affect any characteristics of an individual. SNPs have also become important tools for disease diagnosis in recent years.

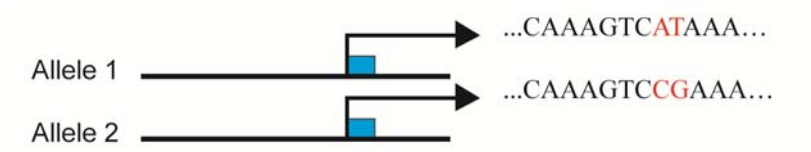


**Figure V.** A dinucleotide polymorphism where the nucleotides AT in one single position on allele 1 has changed into CG on allele 2 in the polymorphic site ((red color = current nucleotides; blue color = SNP site) the figure is modified from Brown, 1999).

## 3.3. Insertions, deletions, and microsatellites

Insertions, deletions, and microsatellites are examples of other important mutations, in addition to SNPs. Insertions are mutations where additional base pairs have been added into a sequence (from TACGGA to TACGGTCCA as an example), while deletions are mutations where a sequence has been lost. Microsatellites, or simple sequence repeats (SSRs), are stretches of DNA (often polymorphic loci) consisting of a short sequence, typically 1-6 bp, repeated several times, i.e. tandemly repeated motifs (example TAATAATAATAA), that can be used as genetic markers (Tautz, Renz, 1984).

Most individuals are heterozygous at any given such locus and microsatellites are variable to a high degree. The repeated sequence can be amplified by PCR and assayed. Fragment length analysis in the form of RFLP (restriction fragment length polymorphism), or AFLP (amplified

fragment length polymorphism), are examples of DNA fingerprinting tools that can be used in order to measure how many repeat units an individual has at a specific locus. In a gel electrophoresis, since each individual inherits one length of nucleotide repeats from its father and one length from its mother, two bands can be seen for most individuals. In the case where only one band is seen in a gel, that individual has inherited the same length from both the mother and the father. In contrast to the frequently occurring SNPs, microsatellite sequences are rare in coding regions of the genome. They occur with a frequency of about 1 per 2 kb (approximately 3 %) in both coding and non-coding regions (International Human Genome Sequencing Consortium, 2001).

## 3.4. Recombination

One definition of recombination is the production of a DNA molecule with segments that has been derived from more than one parent DNA molecule. This is frequently achieved in eukaryotes by the mutual exchange of DNA between non-sister chromatids within a homologous pair of chromosomes during prophase of the first meiotic division (http://www.fao.org/biotech/index_glossary.asp). There are two mechanisms for recombination in eukaryotes; homologous recombination, where a damage in the sequence is repaired with a similar DNA sequence, and non-homologous end joining wherein two non-homologous DNA ends are joined together. Recombination includes the process of crossing over, 1-2 crossing over events occur per meiosis and chromosome. When crossing over events occur, alleles that lie together on a parental chromosome can be split, resulting in e.g. alleles originally situated on chromosomes from different grandparents ending up on the same chromosome. Crossing over between two homologous chromosomes results in chromosomes with some paternal genes and some maternal genes. Recombination provides a useful tool in gene mapping, since such maps are based on the frequency of recombination events. The longer the distance between two genes, the higher the probability of an odd number of crossovers, to a maximum of 50 % between two generations. An indication of how far apart two genes are on a chromosome is the recombination fraction, which is the proportion of meioses resulting in a recombination event (Burton *et al*., 2005).

It is rare for a crossover to occur between loci at meiosis, and therefore alleles at loci close together on the same chromosome tend to be inherited

together, which is the definition of genetic linkage. The likelihood that a crossover will occur, or recombination, decreases with the distance between the loci. Measurement of the genetic distance can be performed using the recombination fraction, which varies between 0 (tight linkage) and 0.5 (no linkage). Genes with a recombination fraction lower than 0.5 are said to be linked and must necessarily be on the same chromosome (Hartl, Clark, 1997).

## 3.5. Haplotypes

Haplotypes are linear patterns, a set, of closely linked SNP alleles in a region on the same chromosome. Fine mapping of disease genes can be facilitated by exploiting haplotypes when looking for candidate genes. When a new mutation responsible for a genetic disease arises, it always occurs by entering the population within an already existing haplotype, i.e. the ancestral haplotype (Courtesy: National Human Genome Research Institute).

### *3.5.1. Linkage analysis*

Recombination, exchanging materials of DNA, provides a useful tool for mapping disease genes (Morgan, 1911; Simchen, Stamberg, 1969). In linkage analysis (figure VI), DNA samples from disease affected individuals and healthy individuals in a family are collected and compared in order to study the heredity patterns of the disease compared with the heredity patterns of a small fraction of DNA with a known position, i.e. a marker. That is, one follows the phenotype and relates it to recombination or markers. Low recombination frequencies are indications of closely linked sequences, while high frequencies are indications of sequences that are farther apart. By measuring the frequency of recombination between a suspected disease gene and DNA sequences where the location is already known, the position of the disease gene can be determined. For this reason, a haplotype analysis can be useful for identifying recombination events.

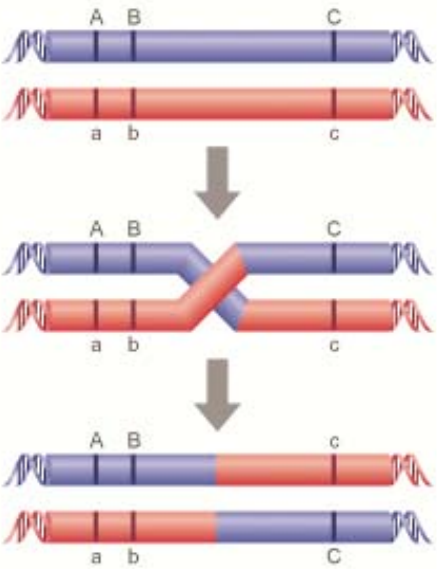


**Figure VI.**
Principle for linkage analysis. The top diagram shows paternal (blue) and maternal (red) chromosomes aligned in a germ cell (a cell that gives rise to eggs or sperm). A, B, and C marks three DNA sequences. The capital letters represent the paternal alleles and the lower case letters represent the maternal alleles. The middle panel shows the physical process of recombination, which involves crossing over of DNA strands between the paired chromosomes. The bottom panel shows what happens when the crossover is resolved. The maternal and paternal alleles are mixed (recombined) and these mixed chromosomes are passed to the sperms or eggs. If A is the disease gene and B and C are genetic markers, recombination is likely to occur much more frequently between A and C than it is between A and B. This allows the disease gene to be mapped relative to the markers B and C (figure used with the permission of The Human Genome, Wellcome trust; © the Wellcome Trust 2004-2005 (http://genome.wellcome.ac.uk).

### *3.5.2. Association analysis*

Using haplotypes in gene mapping involves testing for an association between haplotypes and disease. There are $2^M$ possible haplotypes for M bialleleic markers, which all need to be tested in order to know which haplotypes are associated with a disease. For bialleleic loci with M heterozygous genotypes, there are $2^{M-1}$ possible combinations of phased haplotypes (Palmer *et al*., 2005). In association studies, haplotypes can be used when sick individuals are compared with healthy ones, regarding the frequency of different variants, or alleles, of a specific marker or patterns of such markers, i.e. haplotypes.

The redundancy between markers and marker loci can be used to determine where regions of high linkage disequilibrium (LD) are present in the genome and, ultimately, reduce the cost of association studies. A high LD produces a redundancy among markers and makes it possible for haplotypes as tools for mapping common alleles that can influence the susceptibility to common diseases. This redundancy implies that an association can be detected without genotyping all markers in a region with high LD. Markers that are non-redundant, that detect LD information in a given genomic region are termed haplotype tagging SNPs (figure VII). At low LD, haplotypes can be useful in refinement of SNP-phenotype association if they delineate rare

allele frequencies or if there are significant interactions among the SNPs that affect the trait (Palmer *et al*., 2005). When sites of closely located SNPs are frequently inherited together in blocks, it can be referred to as haplotype blocks. Such blocks harbor a few common haplotypes that are present in a large proportion of chromosomes. Through identification of haplotype blocks, information regarding the extent of LD in the genome can be provided (Gabriel *et al*., 2002).

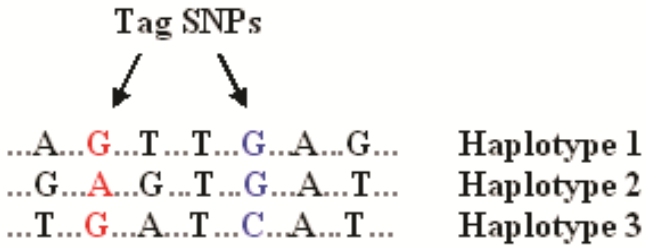


**Figure VII.** Haplotype 1 indicates a chromosome with alleles G and G at the two tag SNPs position, haplotype 2 a chromosome with alleles A and G, and haplotype 3 a chromosome with alleles G and C. SNPs identifying a haplotype block are called tag SNPs (modified from NCBI:http://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Variation/var86.html).

## 3.6. Copy number variation

A copy number variation (CNV) is a DNA segment that has been found to have copy-number differences when comparing two or multiple genomes, where the size of the DNA segment is between one kilobase and several megabases in size (Cook, Scherer, 2008).

In the search for genes underlying common diseases, CNV has recently proved to be important, uncovered during the Human Genome Project (HGP). It is a common and widespread phenomenon in the human genome (Sebat *et al*., 2004; Iafrate *et al*., 2004). Pre-HGP, it was thought that genes were almost always present in two copies in the genome, but this has now been revealed to be more variable (figure VIII). A CNV project is hosted by the Wellcome trust at the Sanger Institute, UK, in co-operation with Hospital for Sick Kids, Toronto, and Harvard Medical School, Boston, and a first-generation CNV map was constructed in 2006, in order to highlight global variation in the human genome (Redon *et al*., 2006). There are now strong

indications that CNVs may be more important than SNPs as the most prevalent form of genetic variation. They will thus be useful for the understanding of human genome variation and evolution.

CNVs occurs mainly in areas of the genome with few or no genes, but they have been found to affect genes that are involved in the immune system, brain activity, and brain development, where the latter two functions have evolved rapidly in humans (Redon *et al*., 2006; Estivill, Armengol, 2007). Association with AD has also been suggested (Redon et al., 2006; Wong *et al*., 2007; Heinzen *et al*., 2009).

According to Dr Scherer, the creation of a global CNV map will transform medical research in four areas, where the most important one is the search for genes involved in common disease. The other three areas are studies of familial genetic conditions, target regions that may be involved in developmental defect caused by chromosomal rearrangements, and to contribute to a more accurate and complete human genome reference sequence (http://www.sickkids.ca/pdfs/4111-CNV_FAQ.pdf).

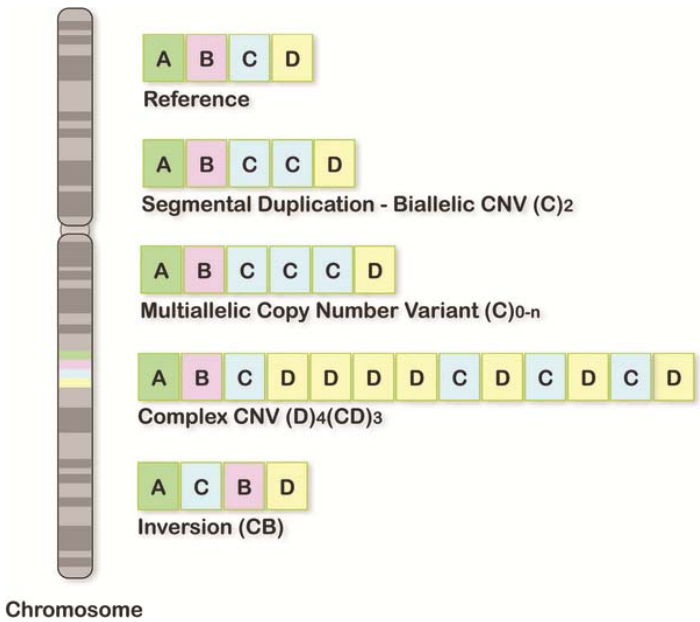


**Figure VIII**.
Genomic structural changes affecting segments of DNA. This leads to deletions, duplications, inversions, and CNV changes (bialleleic, multiallelic, and complex). Segment A is constant, segment B varies in orientation in the inversion, and C and D show different types of variation (Estivill, Armengol, 2007; PLoS Genet 3(10): e190; doi:10.1371/journal.pgen.0030190.g001; © Creative Commons Attribution License, Attribution 2.5 Generic).

# 4 GENETIC STUDIES

## 4.1. Monogenic versus multi-factorial disease

The Austrian monk Gregor Mendel (1822-1884) was the pioneer that first described the principles of inheritance in the mid-nineteenth century (Mendel, 1865). Mendelian diseases are monogenic diseases, where a single mutated gene causes the disease. In multi-factorial diseases environmental factors could play an important role, in addition to mutations in several different genes.

Monogenic diseases are the more rare form of inherited disease and they are often divided by the particular pattern of inheritance they show in affected families, with the most common groups being autosomal dominant disease, autosomal recessive disease, and X-linked disease. In an autosomal disease, the causative gene is on one of the twenty-two chromosomes that do not dictate gender (sex). An autosomal recessive disease requires two copies of the mutant gene, while a dominant only requires one. An X-linked disease has a causative gene that lies on the sex-determining X-chromosome (figure IX).

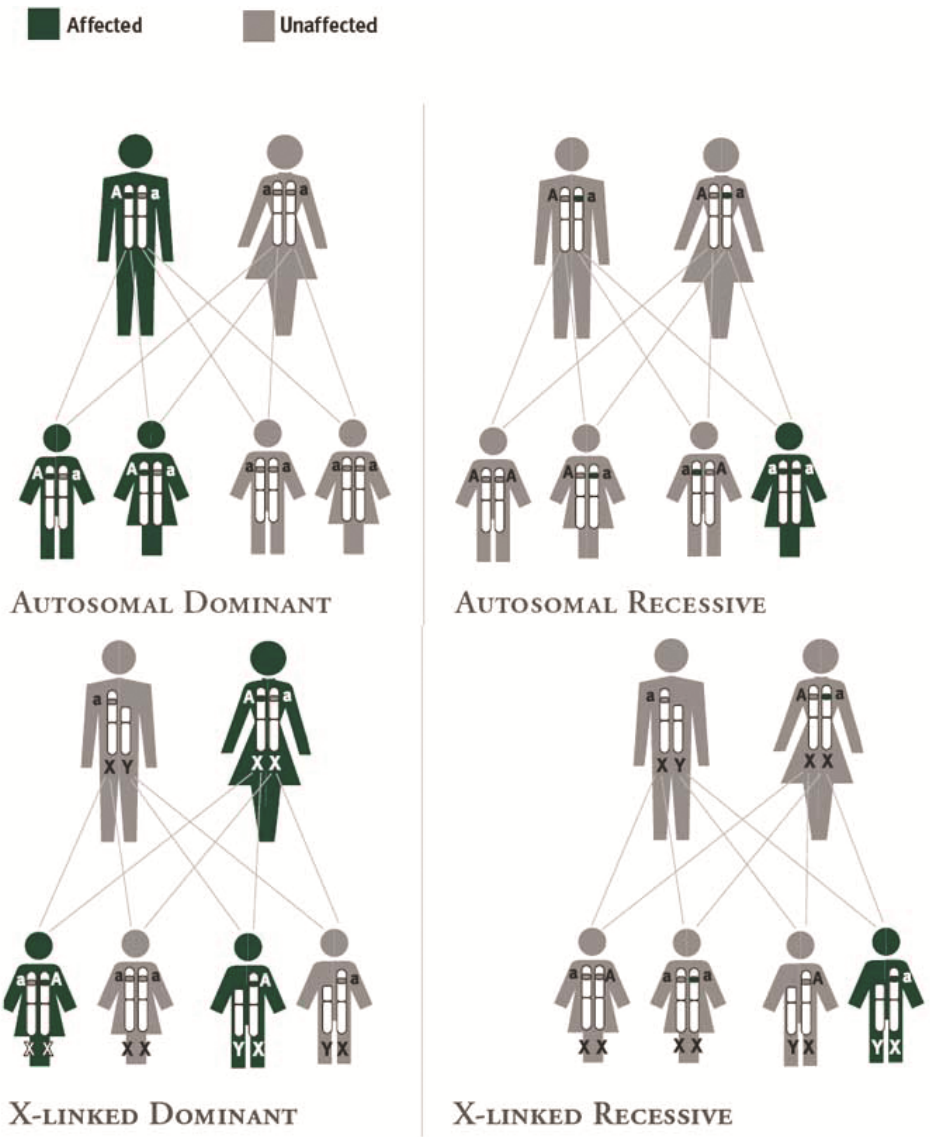


**Figure IX**. Inheritance patterns in pedigree.

**Autosomal dominant**:

- individuals carrying one mutated copy of a gene in each cell will be affected by the disease
- each affected person usually has one affected parent
- tends to occur in every generation of an affected family

**Autosomal recessive**:

- affected individuals must carry two mutated copies of a gene
- parents of affected individual are usually unaffected and each carry a single copy of the mutated gene (carriers)
- not typically seen in every generation

**X-linked dominant**:

- females are more frequently affected than males
- fathers cannot pass X-linked traits to their sons (no male-to-male transmission)

**X-linked recessive**:

- males are more frequently affected than females
- families with an X-linked recessive disorder often have affected males, but rarely affected females, in each generation
- both parents of an affected daughter must be carriers
- only mother must be carrier of affected son (fathers cannot pass X-linked traits to their sons)



Multi-factorial diseases are more frequent, and compared to monogenic diseases, they have a lower tendency to be inherited. Since multi-factorial, or complex diseases often have more than one gene that contributes to the disease, in addition to environmental factors, it is a balance of risks that controls if the individual will get the disease or not. This balance depends on gene variants that may be positive or negative, and on positive and negative environmental factors (Altshuler, 2004).

## 4.2. Linkage disequilibrium

Genetic mutation, the process by which new alleles at a locus appear, for example through errors in replication, is a mechanism that can give rise to LD, an indirect form of association. An appearance of a new allele in a specific individual can lead to a subsequent spread of the allele through the population. Alleles at loci closely linked to the mutated locus will in such a case be in linkage disequilibrium, or associated, with the new allele. An ancestral marker allele, which occurred in a founder with the disease allele, may remain associated with the disease if the pathogenic mutation is close to the marker polymorphism. LD is defined as the non-random assortment of alleles during recombination, or association between chromosomal loci in the general population. This means that the presence of one variant provides information about the other variant, i.e. if you know the allele at one site along a chromosome, you can predict the allele at another site (Weir, 2008).

Comparing linkage, which is an association of gene loci, or markers lying close to each other on the same chromosome using crossing-over, with linkage disequilibrium, which is a more specified type of linkage where alleles occur together more frequently than can be accounted for by chance,

they are both an indication of that markers tend to be inherited together and they both measure a correlation, or co-segregation. More specifically, linkage measures co-segregation in a pedigree, while linkage disequilibrium measures co-segregation in a population. Linkage focuses on a locus; linkage disequilibrium focuses on an allele. The time-span for linkage is 2-3 generations back. LD has a much longer time-span, and is the result of earlier, ancestral, events (Griffits *et al*., 2000).

LD can be influenced by numerous factors, such as the rate of recombination, genetic linkage, mutation rate, non-random mating, genetic drift, and population structure/admixture. Forces that create genetic variation are mutation, random drift, migration, and selection where LD involving the loci underlying the genetic variance for a selected trait, i.e. the genetic variance, is reduced by selection, proportional to the reduction of phenotypic variance of the parents to their entire generation ('Bulmer effect') (Bulmer, 1971). There are two major ways in which selection can affect LD. The first is a type of hitchhiking effect; a haplotype flanking a favored variant can rapidly be swept to high frequency or fixation. The second is through selection for different combinations of alleles at ≥2 loci on the same chromosome (Przeworski, 2002).

LD is decreased by rapid population growth, since this reduces genetic drift. Recombination has a great impact on LD, since it gradually erodes the LD. This also occurs through gene conversion, but to a lesser extent (Ardlie *et al*., 2002). Rates of recombination, a parameter that affects the level of LD, vary in different parts of the human genome (there are 'hot spots', 1-2 kb in length, with concentrated recombination) and are therefore difficult to measure. The distribution and extent of LD can be inflated by demographic factors such as population structure, inbreeding, and 'bottlenecks', i.e. a period when the population was so small that a few ancestral haplotypes gave rise to most of the haplotypes that exist today (Reich *et al*., 2001). The equation below illustrates how LD decays with distance and time:

$$D_t = (1 - \theta)^t D_0 \qquad (4.1)$$

where $D_t$ is the extent of LD t generation later, $D_0$ the extent of disequilibrium at the starting point, and θ is the recombination fraction (Ardlie *et al*., 2002).

## 4.3. Measuring linkage disequilibrium

As mentioned above, LD is an important tool for association analyses, and used by researchers in order to test candidate pathogenic genes using SNP or haplotype blocks (Reish *et al*., 2001; Johnson *et al*., 2001; Jorde, 2000; Rish *et al*., 1996; Daly *et al*., 2001), in addition to population structure studies (Johnson *et al*., 2001; Kruglyak, 1999; Weiss *et al*., 2002). For LD to be useful for gene mapping, knowledge regarding the extent of LD in the areas studied, and the magnitude of haplotypic diversity is needed (Brookes, *et al*., 2005).

In an ideal scenario, ignoring evolutionary forces, and including marker and disease locus mutations, decay in disequilibrium of any kind is due entirely to recombination. Under these conditions, in addition to a short time since the appearance of the disease mutation, the disequilibrium between certain marker loci and the disease allele provided by a curve will exhibit a single maximum at the disease locus. Measuring the amount of LD between a disease allele and markers closely linked to it, could yield information about the location of the disease gene, i.e. in disequilibrium mapping (Devlin, Risch, 1995).

Several ways to measure LD have been developed. Two in particular are commonly used; Lewontin's D, or D', and Hill's and Robertson's r, or $r^2$. The latter form is more appropriate for low allele frequencies since the sample properties are more reliable. If we have two loci, A and B, with two alleles, A, a, and B, b, at each of the two loci, the observed frequency of the haplotype consisting of alleles A and B can be represented as $P_{AB}$. If we assume the independent assortment of alleles at locus A and B, the haplotype frequency to expect is calculated as the product of the allele frequency of each of the two alleles, i.e. $P_A$ x $P_B$ ($P_A$ being the frequency of allele A at the first locus and $P_B$ being the frequency of allele B at the second locus). This leads to the following equation of the simplest measure of disequilibrium for two biallelic loci:

$$D = P_{AB} - P_A P_B \tag{4.2}$$

(Lewontin, 1964).

An illustration of what a simple model of a pairwise LD according to the description above can look like is shown in table 1.

**Table 1.** *Pairwise LD with different haplotype probabilities.*

| alleles | b | B | |
|---|---|---|---|
| a | $P_{ab}$ | $P_{aB}$ | $1\text{-}P_A$ |
| A | $P_{Ab}$ | $P_{AB}$ | $P_A$ |
| | $1\text{-}P_B$ | $P_B$ | |

All four probability cells have the same absolute value for the difference between the observed and expected probabilities under independence, i.e. $|P_{AB}P_{ab} - P_{Ab}P_{aB}|$. Assuming none of the cells can be negative, $D_{max}$ = min[$P_bP_A - P_BP_a$] if D >0, … = min[$P_AP_B - P_aP_b$] if D <0, … where:

$$D = P_{AB} - P_AP_B = P_{ab} - (1 - P_A)(1 - P_B) \tag{4.3}$$

$$D' = proportion\ of\ maximum\ disequilibrium = \frac{D}{D_{max}} \tag{4.4}$$

(Lewontin, 1964).

When $|D'| = 1$ there is complete LD, i.e. no separation has occurred between two markers by recombination, while $|D'| = 0$ equals total independence. $|D'|<1$ is an indication that the complete ancestral LD has been disrupted. If allele frequencies and sample sizes are small, $|D'|$ is scaled in order to remove any effects of different allele frequencies between markers, which creates an upward bias. Comparing D' values between populations with different sample sizes are difficult due to this bias.

The measurement $r^2$, which measures statistical association, can be complementary to D', since $r^2$ is equal to $D^2$ divided by the product of the allele frequencies at the two loci. According to Hill and Robertson, E [$r^2$] = 1/1 + 4Nc, where N is the effective population size and c is the recombination rate measured in Morgans between the two markers. This relationship can also be expressed as:

$$r^2 = \frac{D^2}{\sqrt{P_AP_BP_aP_b}} \tag{4.5}$$

(Hill, Robertson, 1968).

Statistical association can be measured using $r^2$. An inverse relationship between $r^2$ and the sample size is required in order to detect association between susceptibility and SNPs. When only two of the possible four pairwise haplotypes are present, $r^2$ takes the value of 1 (perfect LD). That is, both allele frequencies are identical and there has not been any separation of the two markers by recombination.

Haplotype blocks are patterns of closely located SNPs that are inherited together in blocks. This way of expressing association has been widely used (Daly *et al*., 2001). Regions of the genome that correspond to blocks have a few common haplotypes that account for a large proportion of chromosomes. There are little disequilibrium between blocks and limited haplotype diversity within blocks. The physical extent of the haplotype blocks can be reflected by the D' measure, while haplotype diversity can be reflected by the $r^2$ measure (Cordell, Clayton, 2005). Haplotypes can thus be related to the studied phenotypes (i.e. protein levels connected to AD in a DNA material, genetically determined by the dominant allele in a pair of alleles inherited from the parents) through the partition ligation algorithm[2].

## 4.4. Association studies

There are three major forms of genetic association:

1. direct association
2. indirect association
3. confounded association

*Direct* association targets polymorphisms which in themselves are putative causal variants. The analyses are most simple in this, among the three, most powerful form of association. *Indirect* association uses polymorphisms that are surrogates for any causal locus, which demands typing of multiple surrounding markers in order to catch an indirect association, i.e. association between the allele and a causal variant located nearby. In a *confounded* association, stratification and admixture within a population are used as contributing parameters.

[2] HAPLOTYPER, using partition ligation algorithm in order to reconstruct individual haplotypes from population genotype data ; Niu *et al*., 2002)

Factors that can lead to excess rates of false positives can cause problems in case-control studies. Positive confounding can be the case if false findings are generated, while obscure true causal association can give rise to negative confounding, increasing the false negative rate and reducing power. A well-mixed, outbred population is a basic condition in order to minimize confounding issues (Cordell, Clayton, 2005). Another solution to the confounding problem is to use genetic markers that provide ancestry information, i.e. loci with different allele frequencies compared to a founder population (Pritchard *et al*., 2000; Hoggart *et al*., 2003). Confounding can also be avoided if genomic controls are taken into account (Bacanu *et al*., 2000). Confounding is considered a random process that could potentially affect all loci. Even if individuals have been chosen at random and independently in a population-based association test carried out in a way that fits that specific population, the test subjects can still be regarded as dependently chosen and as a sample of all mankind (Cordell, Clayton, 2005).

In association studies, selection bias does not affect the association unless there is population stratification. Population stratification describes a systematic difference in allelic frequency between subpopulations due to different ancestry and it occurs when the total population has been formed by admixture between subpopulations, or when the admixture proportions (the proportion of the genome with ancestry from each subpopulation) vary between individuals (Hoggart, *et al*., 2003). Matching the ethnic backgrounds of patients and controls, as adequately as possibly, is one solution in order to deal with population stratification. When stratification exists in a population, the association statistics should rise both at the candidate marker and at unlinked markers. In such a case, adjusting for stratification could be performed by genotyping unlinked markers in the same set of cases and controls in which the candidate association was detected (Reich, Goldstein, 2001).

In summary, a positive association can be caused by a direct effect of a polymorphism on risk, by linkage disequilibrium, or by population stratification. When a polymorphic SNP used as a marker occurs within a gene, the amino acid sequence of the gene could possibly be altered, or transcription could be affected. Different alleles may in that way cause differences in the nature or quantity of the gene product, which may affect susceptibility to the disease, i.e. a direct effect of polymorphism on risk (direct association).

Adding statistical power to association studies is important in order to produce a good study design. Type I and type II errors are mentioned above (as negative and positive confounding), and these constitute two statistical parameters that need to be reduced. Power in association studies depends primarily upon strength of the LD with the given markers, frequencies of susceptibility and the marker alleles, sample size, and the effect of the size of the susceptibility locus (Palmer *et al*., 2005; Cardon, Bell, 2001).

Searching for a novel candidate gene potentially involved in complex diseases such as AD thus demands careful planning, and several considerations (figure X). Designing an ideal association study is difficult. As mentioned above, the selected populations should be well-mixed, outbred, and be large enough in size to achieve a good statistical power and the variants for typing should be collected with great accuracy. According to Hattersley and McCarthy (2005), there are few publishers that demand that scientist publishing association studies should describe what kind of steps they have taken in order to reduce error in their study (e.g. errors in genotyping or in performance characteristics).

In order to reduce such errors, we have taken the following steps in the papers presented here: we used blank wells as controls in the 384-microwell titerplates, statistical analysis of the expected allele frequencies using genotype or haplotype frequencies, comparing these with previous data from similar ethnic populations, controlled if genotype frequencies were in Hardy-Weinberg equilibrium, duplicated the genotyping using different assay methods, analysis of LD and haplotype structure, confirming the detected genotypes with reference samples as an assay design quality check, retyping where the result is ambiguous or difficult to interpret, and mixing sample sets within batches, i.e. mixing control samples with case samples on the same plates.

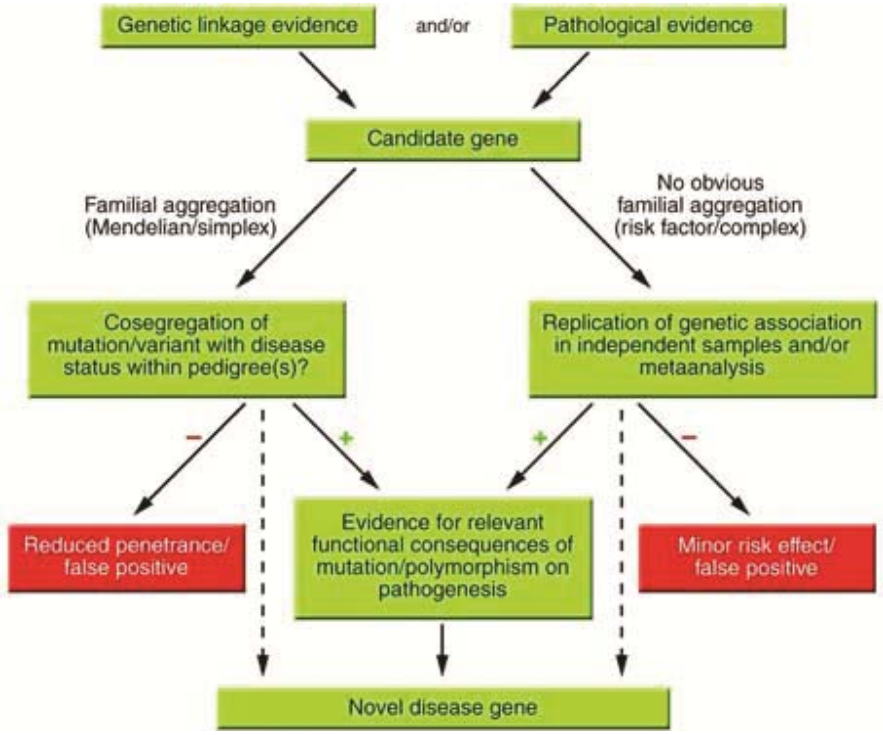


**Figure X.** Flow chart for searching for novel candidate genes. The 'candidate gene approach' includes selection of genes based on genetic linkage data and/or hypothesized or known pathobiological relevance to disease mechanisms. An alternative approach, inherently similar, is referred as 'positional cloning'. This strategy involves detection of formerly unknown genes/proteins according to genetic linkage data. Dashed lines mark 'shortcuts' allowing the definition of a novel disease gene based on the genetic evidence alone, e.g., APOE-ε4 in AD, of which the precise functional consequences remain unknown despite an established genetic role. Red boxes marks examples of genes/mutations with reduced penetrance or minor risk effects within bona fide disease genes (e.g., certain mutations in PSEN1 in AD) (reproduced with permission from the American Society for Clinical Investigation (http://www.jci.org); Bertram, Tanzi, © 2005).

## 4.5. Case-control studies

In a case-control study two sorts of samples are collected and evaluated, the first from affected individuals, the second from unaffected individuals. A chi-square ($\chi^2$) can be calculated for the homogeneity of the allele or genotype frequencies between these two groups of individuals. If there is a significant difference in variant frequency between the groups, i.e. a

significant association, this is consistent with the hypothesis that a typed variant either influences trait susceptibility, or that a second variant in LD with the first one influences that trait susceptibility (Hattersley *et al*., 2005).

Genetic association studies, or epidemiological studies, can be family-based or population-based. In both cases, genetic epidemiologists assess the impact of genes in the population studied. As for all types of association studies, there are mainly three forms of biases that need to be taken into consideration, namely *misclassification*, *selection*, and *confounding*. These biases diminish as the sample size increases. Confounding is a function of the complex relationships between environmental factors and disease. Examples of biases in each of these main groups are provided in a classic paper from 1979 by Sackett.

Advantages with case-control studies are that they do not require any follow-ups and that they also provide estimates of exposure effects. On the other hand they require careful selection of controls in addition to the potential for confounding in terms of population stratification (Cordell *et al*., 2005).

# 5 SNP GENOTYPING

## 5.1. General

In total, we have approximately 20,000 genes in our genome. In order to find disease genes, it is often useful to narrow down the search to one or a few candidate genes (figure X). The selected gene can be studied using genetic markers in the candidate gene, or by selecting genes with known traits that are close to the candidate gene of interest. Such markers are often inherited together with the gene of interest due to their closeness. SNPs are often used in genetic association studies due to their many advantages as genetic markers. These include that SNPs occurs frequently throughout the genome, appearing in introns, exons, promoters, enhancers, and intergenic regions and may in some cases themselves be functional. They are less mutable than other forms of polymorphism, and can through this stability provide more consistent estimates of gene-phenotype associations. Groups of adjacent SNPs can exhibit patterns of correlation, which enhances gene mapping. Differences in SNP frequencies between populations can be used in population-based genetic studies (Palmer *et al*., 2005).

SNPs are useful tools for the determination of the likelihood of an individual to develop a specific disease. Studies of DNA stretches harboring an SNP associated with a disease trait can provide researchers with a hint regarding which genes may be associated with a disease. Through association studies using SNPs as markers, genetic maps can be created from knowing the physical location of these markers and the positions relative to other markers, or genes. According to Palmer and Cardon (2005), there is a minimum of six major areas of potential application for SNP technologies that will improve our understanding of complex disease. These areas are hypothesis-free gene discovery and mapping, association-based candidate polymorphism testing, pharmacogenetics, diagnostics and risk profiling, prediction of response to non-pharmacological environmental stimuli, and homogeneity testing and epidemiological study design (Schork, *et al*., 2000).

SNP genotyping can be used in comparative DNA analyses. SNPs are also locus specific and are excellent long term investments (advantages with using SNPs in the rather inexpensive DASH-method). SNP detection is also easily automated, and SNPs stand out as 'high-throughput' markers. SNP markers can be used to pinpoint functional polymorphisms, and in SNP assays, very small amounts of DNA - typically 50 to 100 ng per individual - are required.

What are the possible disadvantages of using SNPs as markers? One negative aspect is the requirement of special, often expensive, instrumentation. Most SNPs are biallelic and therefore less informative than multiallelic markers such as RFLPs and SSRs (Butler, 2005).

A high density of SNPs near a complex disease gene can provide an association. In addition, a haplotype based statistical analysis can contribute with respect to significance tests as well as fine localization of complex disease genes (Martin *et al*., 2000).

When testing genotyping assays using SNPs as markers, one way to control that the results are reasonably accurate is to examine the genotype data for consistency, or inconsistency with Hardy (Hardy, 1908) -Weinberg (Weinberg, 1908). The most adequate way in a case-control study is to only use the controls when examining the consistency with Hardy-Weinberg (HWE) proportions since this group is more likely to reflect the population.

## 5.2. Genotyping technologies

According to Howell (2003), a useful way to understand the similarities and differences between different SNP technologies is to divide the technologies using three basic concepts. First there is the reaction principle, where the alleles are distinguished, the second where the reaction format is defined, and third - the detection mechanism, where the visualization of the allele specific products occurs. To convert the DNA into an appropriate genotyping form, the reaction principle usually includes polymerase chain reaction (PCR) of the DNA sample.

Several reaction principles have been used for SNP genotyping. These include enzymatic cleavage, single nucleotide primer extension, oligonucleotide ligation, and hybridization with allele-specific probes (Syvänen, 2001; Syvänen, 2005). Enzymatic cleavage relies on endonucleolytic cleavage at the base of stem-looped structures formed during the cooling of single stranded DNA without re-annealing. Examples of genotyping methods based on this technology are restriction site analysis and the invader assay. In single nucleotide primer extension methods, template-directed primer extension is detected by e.g. fluorescence polarization. Examples here are pyrosequencing (Nyren *et al*., 1993; Alderbom *et al*., 2000) and microarray primer extension (Pastinen *et al*., 2000). Oligonucleotide ligation is ligase-mediated and used in methods such as OLA (oligonucleotide ligation assay) (Nickerson *et al*., 1990; Samiotaki *et al*., 1994; Grossman *et al*., 1994) and rolling circle amplification (Lizardi *et al*., 1998). Hybridization methods are based on hybridization of an oligonucleotide probe to the target DNA, where examples of methods include DASH (Prince *et al*., 2001) and TaqMan (Livak *et al*., 1995; Livak, 1999).

For distinguishing between alleles with hybridization assays, the involvement of a single stranded target DNA and an allele-specific oligonucleotide probe is required. 100 % complementarity between target and probe results in a stable DNA duplex in contrast to a non-complimentary, or a mismatched base, which gives a less stable duplex. Of the four methods mentioned above, hybridization is the only one that does not use an enzymatic step.

## 5.3. Hybridization with DASH

Dynamic allele specific hybridization is a method for scoring SNPs or deletions/insertions. The thermodynamic stability differs between two double-stranded DNA duplexes with allele-specific oligonucleotide (probes) that represent two different alleles, and include the immediately surrounding sequence. The hybridization utilizes this difference in thermodynamic stability. The two alleles are separated by affinity for the oligo sequence. One of the alleles is complementary to the oligo, and forms a more stable DNA duplex compared to the other allele, which is non-complementary and thereby introduces a mismatch. Hybridization methods scores alternative alleles by measuring and comparing these stabilities.

The reaction principle in DASH is hybridization, the reaction format is solid-phase and the detection mechanism is fluorescent detection (Howell, 1999). The DASH methodology consists of some basic steps, including amplification of the region surrounding the SNP of interest by PCR, binding of the PCR product to a streptavidin-coated membrane and denaturation of the dsDNA, addition of a fluorescent allele-specific probe for detection to form a target-probe duplex, heating of the duplex while monitoring the fluorescence, and plotting of the negative derivative of the allele-specific decrease in fluorescence against temperature. One of the two PCR primers is biotinylated at the 5'-end, which enables the binding of the PCR product to streptavidin-coated membranes (Jobs *et al*., 2003) or to microtiter plates (Howell *et al*., 1999). Thus the PCR product is bound to a streptavidin-coated 96-wells microtiter plate, or to a 384-wells plate in the improved DASH II technology. This approach provides the advantage and simplicity of using a solid-phase format and of genotyping separate DNA samples from 96 or 384 different individuals in one run.

When the biotinylated PCR product strand has been bound to the microtiter plate, the non-labeled strand is rinsed off with alkali. A one-allele-specific oligonucleotide probe is subsequently hybridized to the target at low temperature. In this step, a duplex DNA region is formed, which can interact with a double strand-specific intercalating dye (figure XI).

The dye emits fluorescence upon excitation. The fluorescence is proportional to the amount of double-stranded DNA (i.e. the probe-target duplex) present. The duplex is heated steadily, and the fluorescence output from the dye vs. temperature is measured. This procedure provides a

complete denaturation profile. Distinguishing between the different target sequences alleles can easily be performed thanks to their difference in melting temperature ($T_m$).

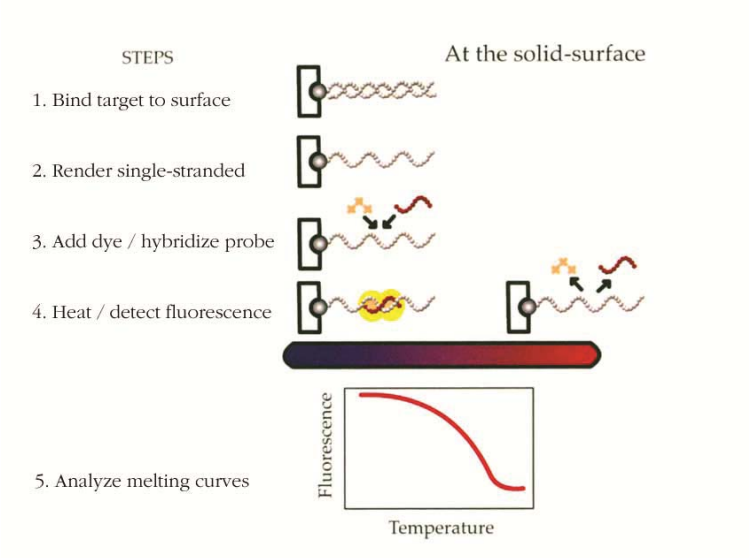


**Figure XI.** The principle for DASH. In the first step, the target (biotinylated PCR product) binds to the streptavidin coated well surface (immobilization). Addition of an alkali solution denatures the PCR product and removes byproducts from PCR together with non-labeled strand. A one- allele-specific probe (ROX-labeled in the present study) is added and hybridized in the presence of SYBR Green I dye. Heating of the sample while monitoring the fluorescence and finally plotting of data is performed with collected fluorescence vs. temperature, and scoring of genotypes by comparative analysis of the melting curves generated. (The figure is used with permission from Howell, © 2002).

In case of a homozygous sample containing the mismatching allele, the fluorescence intensity decreases at a relatively low temperature, while a homozygous sample containing the matching allele results in a fluorescence intensity drop at a higher temperature. When the first negative derivative of the denaturation curves is calculated, interpretable peaks of the probe-target duplexes at the melting point are produced, and these can easily be displayed. A curve showing a double peak with two distinct drops in fluorescence intensity after the first negative derivative, indicates a heterozygous sample, while a single high or low temperature peak represents a matching or a mismatching sample, respectively (figure XII).

Plotting the temperature on the X-axis and the fluorescence on the Y-axis indicates a mismatch between the probe and the template if the $T_m$ is low. If the $T_m$ is high, the probe matches the template and thus the probe-target duplex is more stable (figure XII).

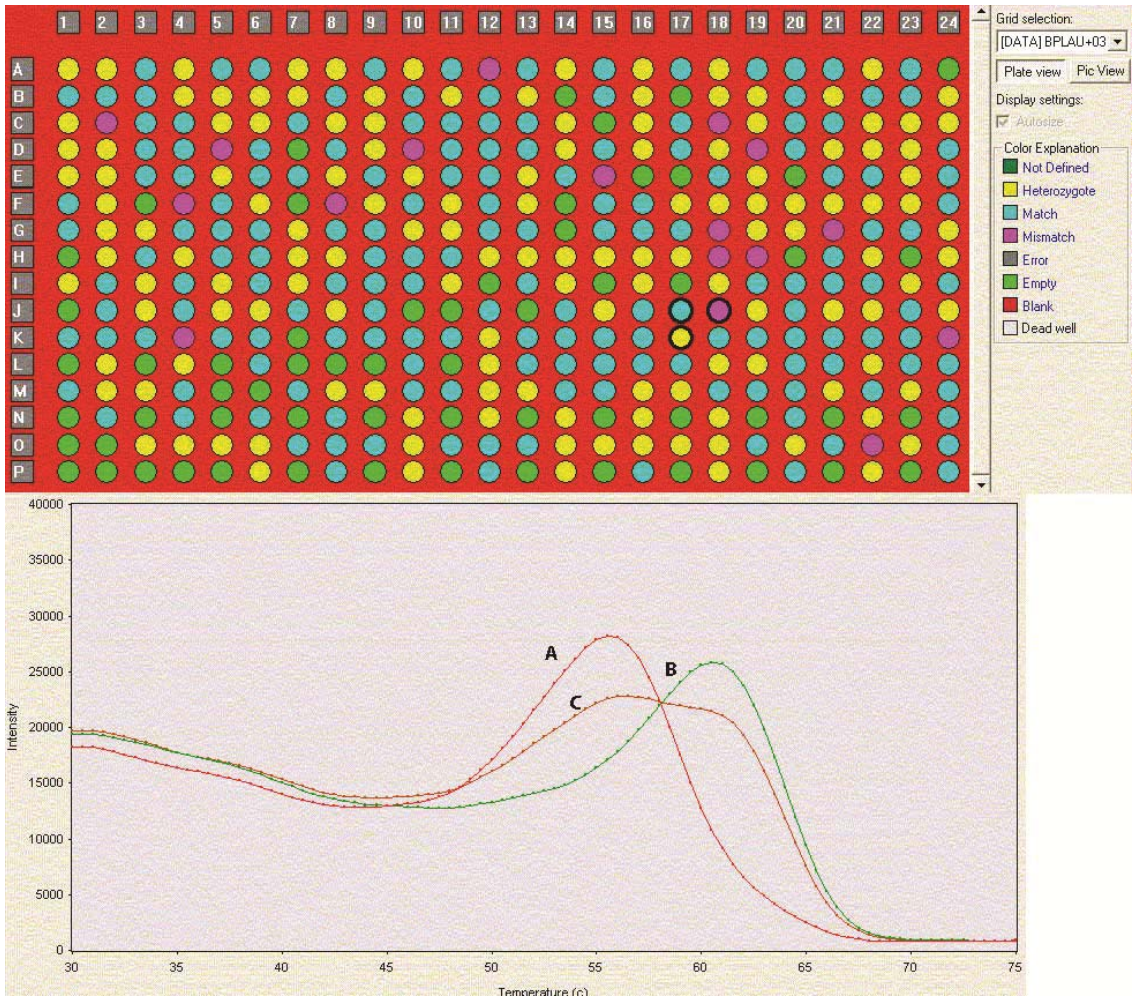


**Figure XII.** Example showing the negative derivative of the denaturation curves from a DASH-run displaying the peaks at the melting temperature of the probe-target duplexes. The Y-axes shows the intensity of fluorescence ($-dI/d_T$) whilst the X-axes shows the temperature in °C. Curve A represents a homozygous mismatch, curve B a homozygous match, and curve C a heterozygous genotype. The 384-plate is shown in the top of the figure, with the three shown samples marked with black circles.

To make a DASH-assay as optimal as possible, some simple design rules need to be observed. Short PCR products work best, two 22-23 bp PCR primers amplifying a 50-60 bp DNA fragment (Brookes, 1999; Howell, 1999).

## 5.4. Hybridization with TaqMan

Both DASH and TaqMan are hybridization methods, but TaqMan is also classified as an enzyme-based method, since the Taq DNA polymerase 5'-nuclease activity is used. While the DASH method uses a solid support reaction format, the TaqMan method uses a homogeneous reaction. The detection method in DASH is fluorescence, and in TaqMan FRET (fluorescence resonance energy transfer).

The TaqMan principle is also used for Q-PCR, which was used in my fifth paper (not published), a quantitative, real-time PCR that reflects the quantitative relationship between the amount of PCR product at any given PCR cycle and the amount of starting target DNA. In TaqMan, the probes used contain a reporter dye that is linked to the 5' end of the probe, and a quencher that is non-fluorescent at the 3' end of the probe. Probes that are hybridized to the complementary target are the only ones cleaved by the 5' exonuclease activity of the TaqMan DNA polymerase. This cleavage separates the non-fluorescent quencher from the reporter dye, which gives an increase in reporter dye fluorescence at each PCR cycle. (http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_040198.pdf).

## 5.5. Other methods

### *5.5.1. Sequenom iPLEX*

Sequenom iPLEX is based on single nucleotide primer extension and is used for fine-mapping and SNP validation studies, or any application based on SNP panels. The steps involved in this method are amplification of the DNA, shrimp alkaline phosphatase (SAP)-cleavage of a phosphate from unincorporated dNTPs (deoxynucleoside triphosphates) (which converts them to dNDPs), thus making them unavailable for future reactions (i.e. the iPLEX reaction), adding of a reaction cocktail (including primer, enzyme,

buffer, mass-modified nucleotides), and in a last step visualizing the results by MALDI-TOF (Matrix-assisted laser desorption/ionisation-time of flight) mass spectrometry analysis (detecting the desalted iPLEX reaction product transferred onto SpectroCHIPs) (Copyright © 2005 - 2009 SEQUENOM, Inc.).

### *5.5.2. Golden Gate assay*

The Golden Gate assay is a method based on allele-specific primer extension and oligonucleotide ligation. It involves 8 steps including the making of activated DNA, adding DNA and hybridizing to oligonucleotides, extend, ligate, clean-up, PCR cycle, binding of PCR-product, elute dye-labeled strand, and prepare for hybridization, hybridize to an Array Matrix, washing and drying Array Matrix, and finally imaging of the Array Matrix (© 2006 Illumina, Inc.).

# 6 FOCUS ON IDE

Throughout my studies I have had a particular focus on the gene coding for the insulin degrading enzyme, *IDE*. The evolutionarily conserved, 110-kDa protease IDE (EC 3.4.24.56, insulysin, or insulinase) is a neutral metallopeptidase in the protein family peptidase M16, and is zinc-binding (Authier *et al*.1995, Qiu *et al*., 1998). Besides Aβ, IDE also degrades substrates such as insulin, glucagon, and other polypeptides, such as transforming growth factor α (TGF-α), atrial natriuretic factor (ANF), β-endorphin, and amylin. The principal subcellular location for IDE is primarily cytosolic (Hersh, 2006), but it is also found in extracellular fluid, internal membranes, and the plasma membrane. It has also been observed as a fraction of the enzymes present in peroxisomes (directed to that organelle by a C-terminal peroxisomal targeting signal (Authier *et al*., 1995; 1996)). The enzymatic activity is optimal at pH 6.0-8.5 (i.e., a physiologically relevant pH range) (Duckworth *et al*., 1998).

The gene is situated on chromosome 10q[3] (figure XIII), with a size of 122,412 bases. It stretches between 94,211,441 bp from *pter* to 94,333,852

[3] situated on the q-arm (the longer arm of chromosome 10)

bp (UCSC Genome Browser on Human, Feb. 2009, Assembly (hg19)). In addition to having a particularly robust capacity to degrade Aβ, IDE may also be involved in the catabolism of the β-amyloid precursor protein (APP) intracellular domain (AICD). Adding of the IDE substrate insulin to an in vitro degradation assay efficiently blocked removal of AICD, whereas overexpression of IDE enhanced the degradation of AICD (Edbauer *et al*., 2002).

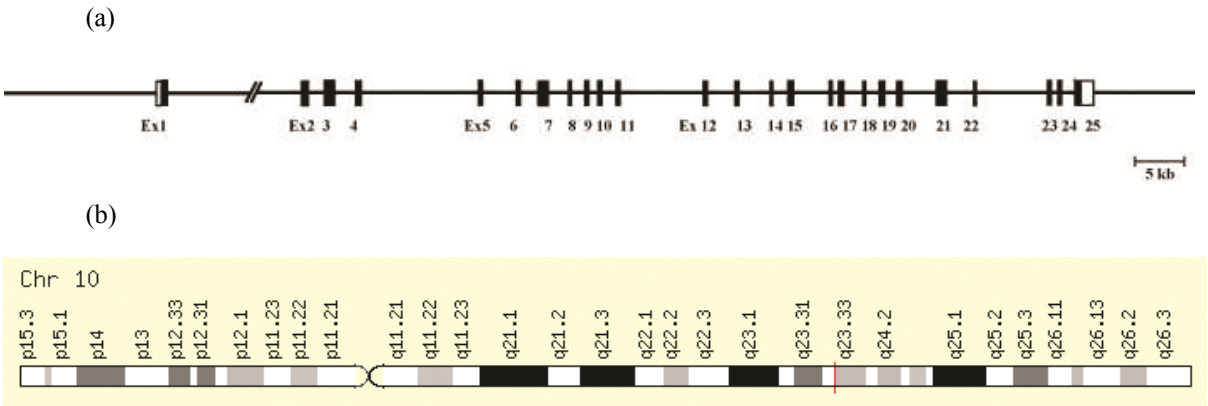


**Figure XIII.** (a) *IDE* gene organization. Map of *IDE* on chromosome 10q23–25: 120kb. Shaded blocks represent coding exons and white blocks represent 5' and 3' UTR. (b) Position of the *IDE* gene on chromosome 10 marked in red.

Even if IDE does not have a known signal peptide or any transmembrane domains, the protease has a membrane-anchored form, which suggests a supporting function in the regulation of insulin signaling at the plasma membrane, in addition to participating in degradation of both membrane-associated and soluble forms of Aβ. There is still some uncertainty regarding the ability of IDE to degrade secreted peptides like Aβ and insulin, since the enzyme primarily exists in a soluble form in the cytoplasm, but since one form of IDE has been found on the cell surface (including in neurons) in addition to the intracellular membranes, the physiological relevance of its degrading capability is more widely accepted (Selkoe, 2001). A recent report suggests that Aβ forms a highly stable complex with IDE, and even forms part of the active site of IDE. This finding supports a novel interaction between IDE and Aβ with potential implications for AD pathogenesis (Llovera *et al*., 2008). Continued studies of the mode of entry of IDE into

membranes and how it attaches to membranes by an anchor are necessary in order to increase the knowledge about this protein.

The interaction between IDE and its substrates has been under the magnifying glass for a long time, and is now a hotter topic than ever after X-ray crystallographic analysis of the free and substrate-bound states (Shen *et al*., 2006; Im *et al*., 2007; Malito *et al*., 2008; Hulse *et al*., 2009). The crystallographic structure published by Shen *et al*., (2006) has opened the door not only for potential drug discovery for e.g. AD, but has also indicated some key factors for the substrate selectivity of IDE (figure XIV).

There are at least five such key factors that have been discussed in recent publications. Since IDE is made up by four domains, where two main domains (IDE-N and IDE-C) are divided into two domains each, with one main domain being mostly neutral or negatively charged (IDE-N; exosite) whereas the other is predominantly positively charged (IDE-C; catalytic site), it enables IDE with the option to select and exclude substrates by charge complementarity and charge repulsion, respectively (Malito *et al*., 2008). Another determining factor is size, since the catalytic chamber of IDE is only large enough to accommodate relatively small peptides (<50 amino acids). When a peptide substrate binds to IDE, it undergoes a substantial conformational change. This flexibility facilitates an effective degradation of them by IDE, as another key factor. This may also play a role in the formation of Aβ plaques (Malito *et al*., 2008).

IDE has a preference for specific substrate sequences when they are bound to the catalytic residues (IDE-C). Mainly hydrophobic residues are bound, but also non-hydrophobic residues fit without a significant energetic penalty. Key specific determinants are favorable binding of the substrate N-terminus and cleavage sites to β-strands within IDE, and proper anchoring of the cleavage site in the catalytic cleft (Shen *et al*., 2006). Peptides without significant positive charges at the C-terminus and that avoid charge-related repulsion from IDE-C are also the best IDE substrates. IDE uses its N-terminus for anchoring to substrates, and binds them to a conserved exosite in the substrate-binding chamber (Malito *et al*., 2008). Finally, for IDE has a unique regulatory mechanism. Its catalytic site is buried in the secondary structure, with a kinetic control of the access to the chamber that uses a closed-open conformational switch (figure XIV) (Shen *et. al*., 2006).

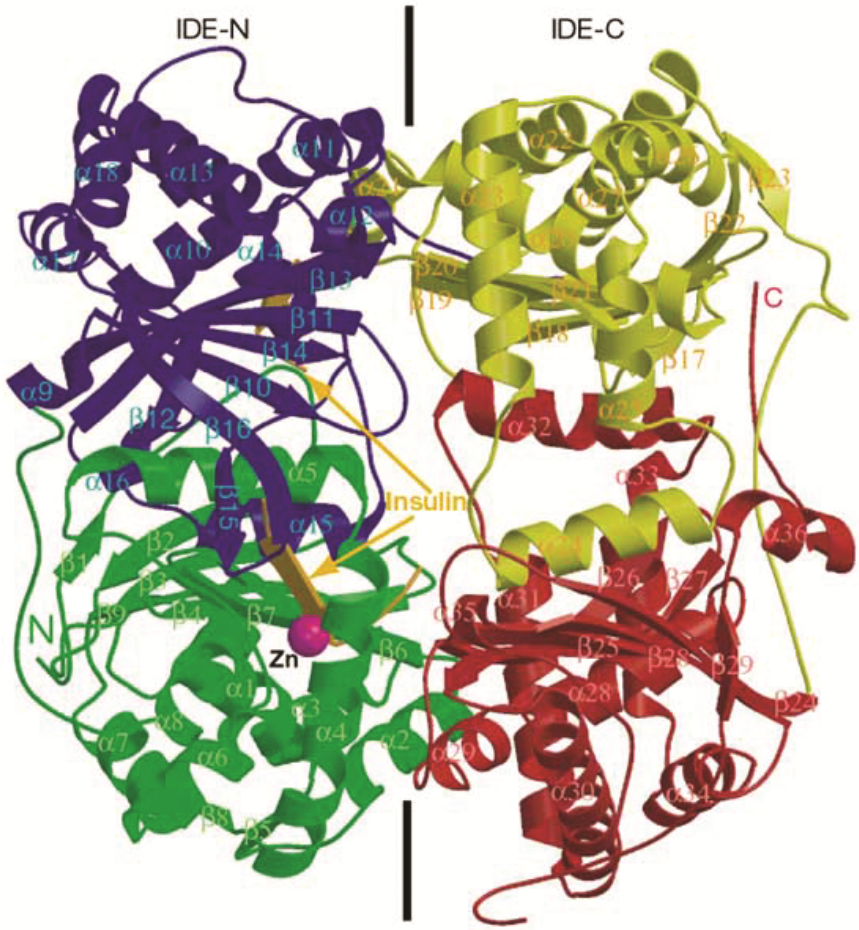


**Figure XIV.** Secondary structure representation of the IDE-E111Q–insulin β-chain complex. Domains 1, 2, 3 and 4 are colored green, blue, yellow and red, respectively. $Zn^{2+}$ and insulin β-chain are colored magenta and orange, respectively. Reprinted with permission from Macmillan Publishers Ltd: [Nature] (Shen *et al.*), © (2006).

# 7 AIMS

## 7.1. Overall aims

In an effort to expand knowledge in the area of complex diseases, more specifically AD, the underlying genetic factors have been a major focus throughout this thesis. SNP based association studies can help pinpoint genes involved in complex diseases. A sufficiently large number of clinical

samples and a reliable high-throughput SNP scoring method are necessary in order to reveal an association.

This thesis work aims to improve the understanding of the genetic basis for AD and related quantitative traits. For this purpose, we used comparative re-sequencing of candidate genes for AD in order to discover sequence differences between individuals with and without the disease in the studied population. Comparing such potential differences with previously known variants is one approach to better understand the role of genes involved in complex diseases. I have focused particular on a part of chromosome 10q, which was previously implicated by linkage findings as harboring one or more potential AD modifying genes. The gene encoding the insulin degrading enzyme (*IDE*) is located in this region. IDE has been studied extensively, with a particular emphasis on comparative re-sequencing of the promoters, exons, and coding regions. In addition, the genes *HHEX*, *KIF11*, and *CTNNA3* were studied in the same way. The goal of this thesis work was to identify polymorphisms in haploblocks linked to, and associated with, AD, and to use these polymorphisms in population-specific association studies.

## 7.2. Specific aims

***7.2.1.*** ***Paper I.*** To investigate underlying linkage signals from polymorphisms in *CTNNA3* for potential involvement in AD, and to compare the linkage signals with those derived from polymorphisms in *IDE*.

***7.2.2.*** ***Paper II.*** To investigate the quantitative trait AAO, which has been claimed to be connected both with AD and PD through a linkage disequilibrium (LD) block on chromosome 10q, in order to explore the possibility that polymorphism within this LD block may also influence PD.

***7.2.3.*** ***Paper III.*** To further investigate the previously examined association between DNA variants in *IDE* and AD pathology and severity with a particular focus on Aβ-deposition in the AD brain.

***7.2.4. Paper IV.*** To highlight a frequently occurring publication bias in scientific journals. Mostly positive association studies are published, even though most studies have negative results. The background is our conviction that abundant data exist that might help resolve the role of gene(s) in complex phenotypes, but that studies producing mostly negative results are being neglected in many journals.

# 8 SIGNIFICANCE

Identifying the genes that underlie prevalent complex human diseases, such as AD, stands to greatly enhance our knowledge of their disease etiology and to suggest treatment. These efforts also increase our understanding of how genes function at the molecular level and they provide novel information regarding the DNA sequence changes that can affect gene function. Ultimately, the identity of the complete repertoire of genes that contribute to an individual's risk for disease can lead to the personalization of therapy and possibly to a cure.

# 9 PRESENT INVESTIGATION

## 9.1. Paper I

### *9.1.1. Aim and background*

There are several genes that compete for being the focus gene, in addition to the known risk factor *APOE*, as a risk factor in AD. Recent studies of genes involved in AD have drawn attention to a region on 10q of chromosome 10 (Bertram *et al*., 2000; Ertekin-Taner *et al*., 2000; Kehoe *et al*., 1999; Li *et al*., 2002; Myers *et al*., 2000). In fine mapping, by measuring linkage distances[4], a distally localized linkage peak (Bertram *et al*., 2000; Li *et al*., 2002) in this region has been found in the vicinity of *IDE*, a gene that has

[4] using the LOD (logarithm (base 10) of odds) score method (Rice *et al*., 2001)

been shown to influence AD (Prince *et al*., 2003). A linkage peak located more centrally, in the proximity of alpha-3 catenin (*CTNNA3*) (Ertekin-Taner *et al*., 2000; Kehoe *et al*., 1999; Myers *et al*., 2000), has also been identified.

In 2003, Ertekin-Taner *et al*. reported an extensive LD mapping study of the *CTNNA3* region. Fifty-six SNPs were genotyped, among which two intronic variants[5] showed exceptionally strong evidence of association. In paper I, we report on an attempt to replicate these findings by genotyping the two most strongly associated markers, rs7070570 and marker 4783[6], in a relatively large sample of AD cases and controls.

### *9.1.2. Materials and methods*

The samples consisted of a Scottish and a Swedish populations as previously described (Prince *et al*., 2003) and an additional sample set consisting of 158 AD patients for quantitative trait analyzes (Andreasen *et al*., 2001). A total of 1006 individuals were genotyped using the DASH technology (Prince *et al*., 2001) (table II).

By including a number of quantitative traits with connection to AD pathology, we increased the power of the study. The traits included were age at onset (AAO), mini-mental state examination (MMSE) scores, cerebrospinal fluid (CSF) $A\beta_{42}$ and tau levels, and senile plaque-neurofibrillary tangles (SP-NFT) scores. Since the two tested markers were in very high LD ($r^2$=0.9), we excluded the 4783 marker in subsequent analyses.

### *9.1.3. Results*

No significant evidence of association was observed between the markers and any of the studied quantitative traits. We also tested for interaction effects in relation to age, gender, and ε4 status with second order factorial ANOVA models, by examining MMSE scores, $A\beta_{42}$, or tau levels. Only one isolated positive finding emerged from this analysis (interaction between

[5] alleles located in an intronic sequence of a gene
[6] without any dbSNP entry when the paper was submitted; rs12357560

rs7070570 and age in relation to tau levels ($P$=0.028)). We therefore summarized our results as negative, i.e., we have not been able to replicate the previous finding that sequence variation in *CTNNA3* contributes to AD (Ertekin-Taner *et al*., 2003).

### *9.1.4. Discussion*

In summary, we have not been able to confirm that sequence variation in *CTNNA3* contributes to AD. CTNNA3 is an uncommonly large gene, which appears to cover several LD blocks, and the identification of possible functionally relevant sequence variants will likely be very difficult, especially since the sequence effort by Ertekin-Taner *et al*. (2003) did not reveal any coding region variants that give rise to amino acid changes.

## 9.2. Paper II

### *9.2.1. Aim and background*

Parkinson disease (PD) and AD are the two most common neurodegenerative illnesses associated with aging in humans and there are several links between them. Epidemiological, neuropathological, and genetic similarities can be found, and common characteristics include abnormal protein aggregation in the central nervous system, with α-synuclein in PD (Spillantini *et al*., 1997) and β-amyloid in AD (Masters *et al*., 1985). As was partially mentioned earlier, genes known to underlie rare mendelian forms of AD are *APP*, *PSEN1*, *PSEN2*, and *SNCA* (synuclein, alpha (non A4 component of the amyloid precursor)), *PARK2* (Parkinson disease (autosomal recessive, juvenile) 2, parkin), and *DJ-1* (*or PARK7*) (Parkinson disease (autosomal recessive, early onset) 7) for PD (Hardy, 2003). There are few known genetic risk factors for the more-common, sporadic forms. In AD, there is the ε4 allele of *APOE*, and for idiopathic[7] PD *MAPT* (microtubule-associated protein tau) (Martin *et al*., 2001), and *UCHL-1* (ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)) (Maraganore *et al*., 2004) even if the evidence for the latter two is not completely conclusive.

[7] of unknown cause; a disease that is of uncertain or unknown origin

The purpose of this study was to investigate if polymorphisms within an LD block[8] encompassing *IDE*, which we reported on earlier (Prince *et al*., 2003), were associated with PD, since the quantitative trait AAO has shown linkage to *IDE* in both AD and PD (Li *et al*., 2002), which suggested that there were pathogenic IDE alleles that influenced both disorders.

### *9.2.2. Materials and methods*

SNPs that delineated multilocus haplotypes from this region, with frequencies above 5 %, were used in an Australian PD case-control study. A previous association with AD was replicated in an additional AAO study using two independent Swedish AD case samples. The Australian study sample consisted of 306 PD cases and 321 controls, while the first Swedish sample consisted of 139 AD cases, and the second of 151 AD cases (table II). The *IDE* markers were genotyped using DASH (Prince *et al*., 2001).

26 SNPs from an LD block that spans around *IDE*, and extends across *KIF11* (kinesin family member 11) and *HHEX* (hematopoietically expressed homeobox) were identified in an earlier study (Prince *et al*., 2003). These SNPs were used in a haplotype 'tagging' strategy (Johnson *et al*., 2001) where SNPs where chosen by their capacity to delineate common haplotypes (>5 %). Five SNPs from the *IDE* region were genotyped in the PD samples in order to search for evidence of association. Three of these were previously defined tag markers, rs2251101, rs1832196, and rs1544210, and two additional markers, rs967878, and rs1887922 were identified. The LD structures were found to be similar in the Australian and Swedish populations. We also tested markers rs7412 and rs429358, since they defined the ε2, ε3, and ε4 alleles of *APOE*.

### *9.2.3. Results*

We explored three association models in the PD sample. The first one involved comparing haplotype distributions between cases and controls using haplotypes defined by the three tag-SNPs. No association was found in this test. In a second test we tested for effects on AAO in PD using a sample

[8] regions of low haplotype diversity separated by recombinational 'hot spots' (Goldstein, 2001)

set that was stratified into early and late-onset groups (median trait value set to 60 years). This experiment tested the possibility that early onset PD has a stronger genetic component due to an enrichment of detrimental alleles and a deficit of protective alleles (Tanner *et al*., 1999). Significant association was observed when both the early and late onset groups were included, and when comparing the early onset group with unaffected controls. In the third model, haplotypes defined by the three tag markers were tested in two AD case samples (sets B and C in paper II). We attempted to replicate earlier finding, an association with AAO in AD (Prince *et al*., 2003). We observed significant evidence of association in one of these two samples, with haplotype effects equivalent to those observed previously.

### *9.2.4. Discussion*

In summary, our findings indicated that an LD block on chromosome 10q harbors alleles that influence AD and PD in a similar way. There are several factors indicating that *IDE* is a candidate for being responsible for this effect. Among these is the fact that this gene is one of the principal proteases responsible for the degradation of Aβ (Qiu *et al*., 1998). Others include *IDE*-deficient mice that exhibit increased cerebral accumulation of Aβ (Farris *et al*., 2003), and a genetic association with neuritic plaque density in AD with maximum effects in the 3' region of IDE, a large block of linkage disequilibrium where both SNPs and haplotypes demonstrated significant disease association (Prince *et al*., 2003). There are fewer indications of the involvement of this gene in PD, but Aβ and α-synuclein have been shown to interact (Yoshimoto *et al*., 1995; Higuchi *et al*., 1998; Nussbaum, Ellis, 2003; Ferreon *et al*., 2009) and transgenic mouse studies have revealed that accumulation of α-synuclein can be enhanced by the presence of amyloid precursor protein and Aβ (Masliah *et al*., 2001). A reduction in IDE activity caused by increased β-amyloid levels could enhance the cerebral aggregation of α-synuclein and thereby promote PD pathology.

## 9.3. Paper III

### *9.3.1. Aim and background*

The aim of this study was to further investigate earlier examined an association between genetic variants in *IDE* and AD pathology and severity.

Four independent studies (Bertram *et al*., 2000; Myers *et al*., 2000; Ertekin-Taner *et al*., 2000; Li *et al*., 2002) have described linkage of AD to chromosome 10q, and two of these studies report maximum LOD scores in the vicinity of *IDE* (Bertram *et al*., 2000; Li *et al*., 2002). As mentioned above, we have previously determined and tested haplotypes in an LD block spanning *IDE* (Prince *et al*., 2003), and one of the most intriguing observations was the association of *IDE* variants with the density of senile plaques and neurofibrillary tangles in an autopsy AD population. It was therefore considered of great importance to try to replicate these findings in autopsy populations in which specific measurements of cerebral Aβ levels could be performed.

A large series of potentially functional variants (in the region and vicinity of *IDE*) was tested against the studied phenotypes and in cases where association could be demonstrated we attempted to reproduce the haplotype effects in independent materials. Using this strategy, we focused primarily on replication of significant findings, since multiple testing became a major issue with the number of markers tested.

### *9.3.2. Materials and methods*

Materials from four populations were used; set A, C, and D consisted of 93, 111, and 186 late-onset AD cases from Swedish populations, respectively, where set A had an autopsy diagnosis and the remaining sets had clinical diagnoses. Set B consisted of 72 neuropathologically confirmed AD cases from an English population (Chalmers *et al*., 2003) (table II). In three of these four sets, significant haplotype effects were observed for several markers in tight LD in the 3' region of *IDE*.

From sets A, C, and D, genotypes were obtained for 32 markers, and in set B 5 additional markers were used. The 32 markers included 31 SNPs and a 2-bp insertion/deletion polymorphism. Genotyping assay quality and verification of the SNPs as polymorphic was performed by an each-marker-test in a pilot set of 16 Swedish control samples. DASH was used for the genotyping (Prince *et al*., 2001).

An LD map was calculated using the 32 markers in combined sets, A, C, and D. From this map, an LD block between markers rs2149632 and rs1544210 was evident, in addition to modest LD between several markers within and

outside this putative block (figure XV). LD between marker pairs was estimated using the $r^2$ metric (Hill, 1974).

Four SP-NFT traits were studied in set A, including frontal lobe and hippocampus SP scores ($r^2 = 0.187$, $P = 0.0001$), as well as hippocampus SP and NFT scores ($r^2 = 0.07$, $P = 0.02$). In set B, three Aβ traits from frontal, parietal, and temporal lobes were studied. These were all found to be strongly correlated ($r^2 > 0.4$ and $P < 0.0001$). Sets C and D were tested for MMSE scores. Indices for SP and NFT density were used in a histopathological scoring system that was applied to the autopsy cases (Alafuzoff *et al*., 1987).

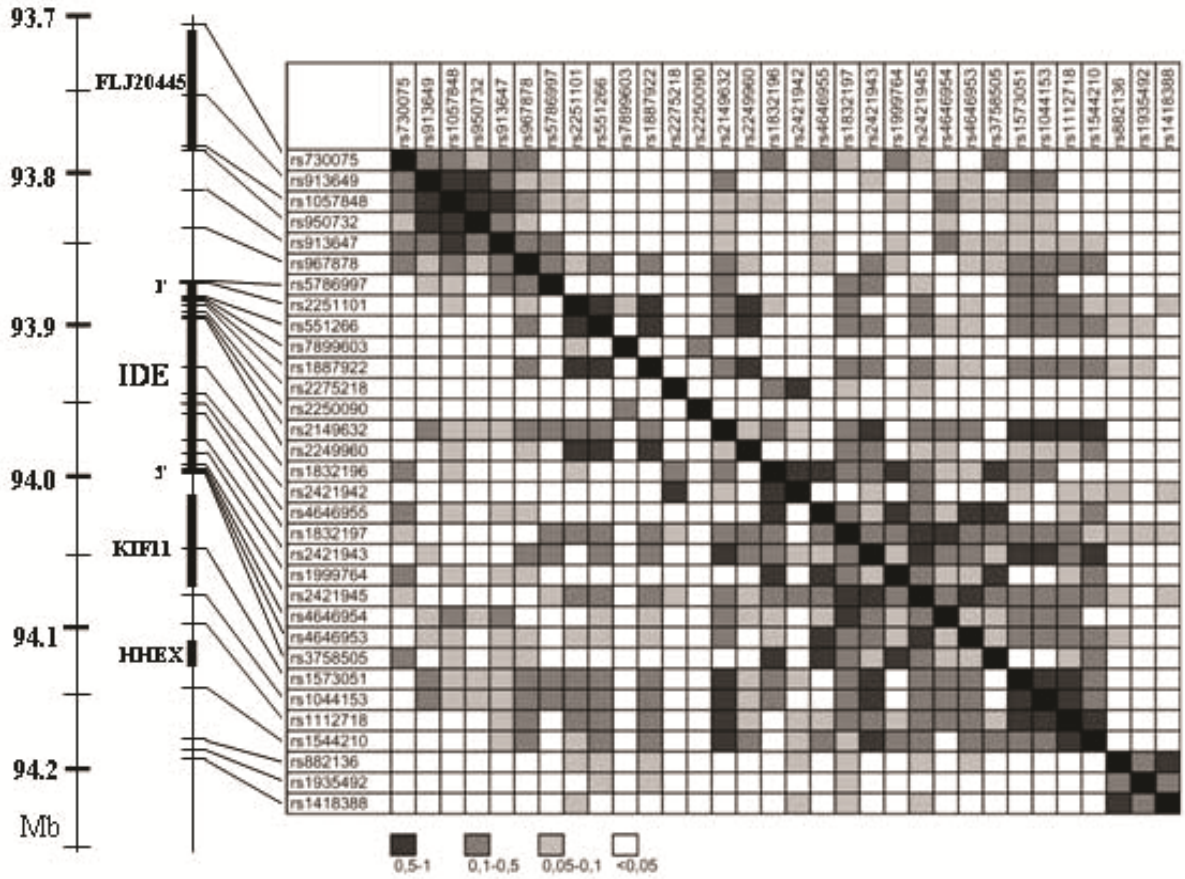


**Figure XV.** Pairwise LD values (estimated using the $r^2$ metric) for 32 genetic markers spanning IDE are shown. Markers are ordered from *pter* to *qter* along chromosome 10 with the distance (Mb) indicated to the left. IDE extends from 3' to 5' between markers rs2251101 and rs3758505. Data above and below the diagonal are equivalent. The LD map was constructed using data from all studied Swedish individuals (sets A, C, and D). For $r^2$-values, white: <0.05, light-gray: 0.05–0.1, medium-gray: 0.1–0.5, and dark gray: 0.5–1.0.

### *9.3.3. Results*

In set A, several markers were found to be significantly associated with SP-NFT scores in the hippocampus, but not in the frontal lobe. Among these markers, rs551266 and rs1887922 showed the strongest effects, i.e. the common allele at these sites was associated with higher average scores. A stepwise, forward selection ANOVA model was used in order to test for

potential individual marker effects and allelic heterogeneity with hippocampus SP-NFT values. The rs1887922 marker was the only one that showed a significant independent effect in a final main-effects model.

The 32 markers were tested against MMSE scores in set C and D. This resulted in a maximum effect for markers rs2251101, rs1887922 and rs2249960 in set C, but no compelling evidence of association in set D. An ANOVA test, see above, for markers exhibiting maximum effects was performed for set C. Marker rs187922 showed the only significant effect. The common allele of this marker was associated with lower MMSE scores, which was in agreement with the data from set A. An attempt to replicate these findings was carried out using set B, via five markers and immunohistochemical measurements of cerebral Aβ load. The markers were chosen to represent the allelic diversity in *IDE* (Johnson *et al*., 2001). The strongest effect was seen for marker rs1887922 in relation to temporal Aβ load, with the common allele being associated with higher Aβ levels, which was also in agreement with the data from sets A and C.

Marker rs187922 was also tested for disease risk association in a combined analysis using all Swedish AD and control materials that we have used in this and previous association studies (Prince *et al*., 2003), a total of 537 AD cased and 499 controls. The result from this analysis was not significant.

### *9.3.4. Discussion*

The main finding in this paper was that significant effects upon plasma Aβ levels from IDE region markers were observed. Together with the observation that marker rs197822 was not associated with direct disease risk in AD case-control samples, these results suggest that IDE plays a role as a modifier of the severity of disease rather than risk, after the pathology has been initiated by other mechanisms. There is also a possibility that moderate deficits in enzyme function could have long term negative repercussions, especially since IDE has been shown to be rate-limiting for the degradation of cerebral Aβ (Farris *et al*., 2003; Qiu *et al*., 1998). We do acknowledge that the involvement of *IDE* in AD is still controversial, since the data suggest an association with disease in case-control models (Abraham *et al*., 2001; Boussaha *et al*., 2002). Still, a recent report has corroborated our original findings, which supports the existence of sequence variants in, or near IDE that contribute to AD (Ertekin-Taner *et al*., 2004).

## 9.4. Paper IV

### *9.4.1. Aim and background*

In this paper I addressed the problem that the vast majority of the association studies that are published are positive, when the reality should be the opposite. A large amount of the reported positive associations of genes with complex human phenotypes during the past decade have generated much skepticism (Weiss and Terwilliger, 2000). Given the often small gene effect sizes and the typically small amounts of test subjects employed in association studies, the vast proportion of tested polymorphic markers should produce negative results. Using published data in meta-analyses is one approach to identify disease genes (Lohmueller *et al*., 2003; Ioannidis *et al*., 2001). For this to be successful, the studies in the literature need to be valid and representative of a random sample of studies. If few negative studies are published, genes that have no impact on a phenotype may be implicated, or the predicted effect of a true risk factor could be inflated (Goring *et al*., 2001).

The aim of present paper was to test our hypothesis that there is an abundance of data that could help resolve the role of many gene(s) in complex phenotypes, but that numerous reasonably powered studies that produce negative results are not being published. The data in this study reflect both a lack of willingness on our part to submit small negative association studies based on a few markers as freestanding articles and also the refusal of some journal editors, based on journal policy to send out such studies for review (Esterbrook *et al*., 1991).

Another important issue is the nature of the phenotypes for which genetic association is reported. Modeling of genetic effects in a case-control study, and comparison of genotype/allele frequencies between people with a disease and healthy individuals is the standard approach. Additive studies of genetic effects on quantitative traits can provide complementary and independent evidence of association (Prince *et al*., 2003; Ertekin-Taner *et al*., 2003; Papassotiropoulos *et al*., 2003).

The biological pathways, Aβ-degradation and metabolism, and the formation of NFT and plaques in the aging brain, are important for the search for additional gene(s) that contribute to AD pathology. The difference in CSF $A\beta_{42}$ levels between AD cases and controls is an example of a

relevant quantitative trait for AD (Andreasen, Blennow, 2002). This phenotype is also strongly affected by *APOE* (Prince *et al*., 2004). In the present study, we modeled the relationships between multiple genes and phenotypes, including both disease risk and quantitative measures of AD severity, in an attempt to show that approaches involving multiple phenotypes will facilitate the research into the role of genes in complex phenotypes.

### *9.4.2. Materials and methods*

Scottish and Swedish AD materials were used for this study (table II). Of the 106 genotyped genetic markers, a total of 62 markers were included in analysis, tested for association with AD risk, and possible effects on quantitative indices of AD severity (i.e. MMSE, AAO, CSF Aβ and tau levels). Genes studied encompassed both novel candidates as well as several recently claimed to be associated with AD (e.g. urokinase plasminogen activator (PLAU) and acetyl-coenzyme A acetyltransferase 1 (ACAT1). Genotyping assays were performed using DASH.

The PubMed database was searched for all genetic association studies involving AD, and I read all such papers published between January 2004 and April 2005.

### *9.4.3. Results*

Within the tested 62 markers, only modest signals were present, which, in isolation, returned a broadly negative result.

The PubMed literature survey on AD identified 138 articles in which one or more markers in a candidate gene(s) were assessed for association with AD disease risk and/or an AD-related quantitative trait. 86 studies reported positive findings, which strongly indicate that there was a publication bias.

### *9.4.4. Discussion*

There is no established forum for the publication of negative genetic association studies. While candidate polymorphisms from negative association studies do not recommend further study by themselves, they could be interesting for future meta-analyses. By carrying out this study, we hoped to contribute to a more favourable climate for the publication of negative genetic association studies, so that a more balanced portrayal of the genetic basis of complex traits may emerge.

**Table II.** *Analytical strategies used in the four studies.*
*EO=early onset, LO=late onset,*
*HWE=Hardy-Weinberg equilibrium, PLSD=Fisher's protected least significant difference.*

| | Type of study | Classified syndrome of cognitive impairment | Statistic tests |
|---|---|---|---|
| **Study I** | Retrospective case-control | EOAD, LOAD/None | HWE |
| sample n=1006 | Prospective longitudinal | | $r^2$-metric |
| | Longitudinal geriatric population | | ANOVA |
| | | | Chi-square |
| **Study II** | Retrospective case-control | EOAD, LOAD, EOPD, | HWE |
| sample n=917 | Prospective longitudinal | LOPD/None | Chi-square |
| | Longitudinal geriatric population | | |
| **Study III** | Retrospective longitudinal | LOAD | Chi-square |
| sample n=462 | Retrospective geriatric population | | ANOVA |
| sample n=1036* | Longitudinal geriatric population | | Kolmogorov-Smirnov |
| | | | Linear regression |
| | | | $r^2$-metric |
| **Study IV** | Retrospective case-control | EOAD, LOAD/None | HWE |
| sample n=1414 | Prospective longitudinal | | Chi-square |
| | Longitudinal geriatric population | | Binary logistic regression |
| | | | ANOVA |
| | | | PLSD |

*Used in a combined analysis with all Swedish materials reported on in earlier association studies (Prince *et al*., 2003). Controls were used in this analysis (EOAD, LOAD/None).

# 10 PERSPECTIVES

Since I started my PhD, there has been a change in the size of association studies, from rather small towards large genome-wide scans. Even if the Human Genome project inspired a cascade of new and improved genotyping techniques and facilities, we are still a long way from uncovering the secrets of many complex genetic diseases. In this case, the size of the studies does not necessarily equal the right answers or good quality analysis of the results. Therefore, I believe it can be useful to perform association studies if one has the right knowledge of good laboratory practice and knows how to calculate the statistical outcomes, even if the facilities, cohort sizes, and the financials are limited.

Understanding the patterns of nucleotide variation can clear the way for a deeper understanding of the forces that underlie the polymorphisms. Through comparative re-sequencing, we can discover sequence variants between individuals with or without the disease in the studied population

Besides continuing my search for potential candidate genes involved in AD, my intention is to turn the focus onto potential connections between AD, stress, and/or major depression. The project could focus on glucocorticoids (GCs) and their role in stress-associated diseases. Stress induced release of glucocorticoids (GC) and the potential of this process as an additive risk factor for AD, either contributing to the development of the disease or to the maintenance of it, was highlighted in a recent study (Catania *et al*., 2009).

# 11 ACKNOWLEDGEMENTS

There are many people I owe my largest gratitude and all my humbleness after these years at KI. Here are some of them (please forgive me if I have forgotten anyone – my only excuse is all the stress connected to finishing up my thesis).

To the examination board and my opponent, **Teresa Frisan**, **Maria Ankarcrona**, **Marta Alarcón-Riquelme**, and **Anna Glaser**: many warm thanks for all the time and effort involved when reading my e-mails, my articles, my thesis, and visit my dissertation. Big hugs to all of you!

To **Björn Andersson**: what a great guy, in addition to at least as great scientist you are! It has been more than a pleasure to work with you. Thanks for 'picking me up when I fell', for our discussions, for your patience, and your knowledge.

To **Jonathan Prince**: thanks a million for all your support in the beginning of my time at 'old CGB', KI. Even if we had our disagreements, I am truly very grateful for your believe in me, your great knowledge in this area of science, and for sharing this with me.

To **Kjell Edman**: thank you for being one of my favorite teachers in Kalmar, and for all your help, input, knowledge, and patience that you have. I really appreciate it!

To **Christer Höög**: thank you for all your help, and patience with me.

To **Hamid**, **Shane**, **Marcela**, **Ellen**, **Oscar**, **Erik**, **Daniel**, **Stephen**, and **Bryndis**: thank you big times for your support, friendship, question-time and patience and for welcoming me into your group. You are the best☺. A special thank to **Shane** (and **Em**) that helped me out when times were rough, and for helping me with my research, and to **Hamid** for all our discussions, bad jokes, and for letting me have a chair in your room (big hugs to your family as well). Also a big thank you to **Bengt Persson** and his group, and to **Rickard Sandberg** and his group.

To **Hagit**: thanks for most of the things we shared…including getting surprises as having all the stuff in your (and mine) room covered up in metal

foil! You sure were a good student forcing you to speak Swedish once a day. It was a pleasure hanging with you.

To **Mun-Gwan**: thanks for teaching me when I tried to teach you in the beginning! Thanks for all the good food inspiration you brought, in addition to scientific inspiration, and for all funny jokes.

To **Sandra**, and all the lab technicians, PhD's, and post-doc's in **Brooke's** group, and all the people in 'our former corridor': thanks a bunch for your friendship, patience, guidance, and laughter. Sandra: I would have been lost without you!

To **Anna** and **Lisa**: you little applepies, you! That is all I have to say☺. To **Linda**: you are also an applepie. To **Fredrik**: I miss your food and your company. To **Denise**, **Benosh**, and **Bruna**: very nice to get to know you, and to have worked with you.

To **Maria Seddighzadeh**: thanks for your support and all the help with TaqMan.

To **Maja**, **Bea**, and **Matti**: I have no clue where I would have landed without your help, input, patience, encourage, 'self-esteem courses', and guidance. Probably not here. At all. I hope that all of you acknowledge how much you mean to me!

To **Gitt**, **Brittis**, **Bettan**, **Christine J**, **Ewa H**, **Irene R**, **Zdravko**, **Riita L**, **Kim A**, **Elisabeth V**, **Janet C**, **Veronica G**, **Elisabeth L**, **Rosa-Amanda V**, IT-pojkarna: you are all very special people to me. Thanks for all your help and input!

To **Veronica Murray**: thanks for all your support, your knowledge, encourage, and believe in me. You are something else!

To **Aleksander Perski**: thanks for your encourage and believe in me. I am looking forward to hopefully work with you in the future!

To all the people involved in all the PhD courses I took: a special praise to LIME that taught me some really useful pedagogic, and to **Pernilla**: I miss our candy sessions!

To my family: I cannot express with words how much you all mean to me! I do love you very much. To **Robbis**: this thesis is for you, little one. You are the strongest shining star in space now. **Mormor**: du är en av mina stora inspirationskällor! **Pappa**: tack för stöd, uppmuntran och kärlek. Och för att du hjälpte oss med vårt badrum när vi behövde det mest! **Mamma**: bara du vet exakt hur mycket du betyder för mig. Kan inte önska mig mer! **Bosse**: thanks for inspiring my curiosity and know-how. Du är bra, du! With hugs to **Alfia** as well. **Tova** and **Carro**: you are my favourite Piff and Puff. Jag är mycket stolt över er! Thanks to **Lena**, **Gun**, **Mikael**, **Ebba**, **Pia** med familj, and of course to **Hanna** and **Göran**. And to all my relatives: **Anders** and **Birgitta**, **Lars** and **Irene**, **Jan** and **Britt**, and **Vingard** and **Annelie** with children.

To my friends: I would only be half without you. That wouldn't look good at all! Sorry for not mentioning all of you, but here are some special applepies: **Jenny**, **Micke**, **Freja**, **Matilda**, **Birgitta**, and **Hans**: Tynningö would be very quiet without you. Rock´n Roll! Many thanks to **Allan** and **Eva** for dinner, laughter, escape into the woods, and your knowledge. **Liina** (and **Elvira**) and **Anna** (and **Tojje** and **Vendela**): thanks for being an important part of my thesis. Pussgurkor☺. Then: **Mia**, **Johanna**, **Jonas**, **Helena**, **Rikkard**, **Erika**, **Jocke**, **Sus**, **Ragge**, **Juniper Nova**, **Christine** (for always having shelter for me), **Pelleplupp** with family, **Peter** and **Lotta**, and many more. **Fru Beckerell** (aka **Rebecca Wall**): you crazy one! I miss 'fläskkotleeett' and everything else. Thanks for being a true friend back in Kalmar.

To **Erik**: the love of my life, the best little man in the world. For your unconditional love, for your…hrmm…very great patience with me, for your intelligence, for being you. You are the greatest gift I ever got! Also a big bowl of tuna fish for **Cosmos** and **Aston**.

# 12 REFERENCES

Abraham R, Myers A, Wavrant-DeVrieze F, Hamshere ML, Thomas HV, Marshall H, Compton D, Spurlock G, Turic D, Hoogendoorn B, Kwon JM, Petersen RC, Tangalos E, Norton J, Morris JC, Bullock R, Liolitsa D, Lovestone S, Hardy J, Goate A, O'Donovan M, Williams J, Owen MJ, Jones L. Substantial linkage disequilibrium across the insulin-degrading enzyme locus but no association with late-onset Alzheimer's disease. Hum Genet. 2001; 109(6):646-52.

Alafuzoff I, Iqbal K, Friden H, Adolfsson R, Winblad B. Histopathological criteria for progressive dementia disorders: clinical-pathological correlation and classification by multivariate data analysis. Acta Neuropathol (Berl). 1987; 74(3):209-25.

Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res. 2000; 10(8):1249-58.

Altshuler D. The inherited basis of common diseases. In: Goldman L, Ausiello D, editors. Cecil textbook of medicine. 22rd Ed., Saunders; 2004, pp. 198–202. Philadelphia, USA.

Alzheimer, A. Ueber eine eigenartige Erkrankung der Hirnrinde. Allgemeine Zeitschrift für Psychiatrie und psychisch-gerichtliche Medizin, Berlin. 1907; 64:146-48.

Andreasen N, Minthon L, Davidsson P, Vanmechelen E, Vanderstichele H, Winblad B, Blennow K. Evaluation of CSF-tau and CSF-Abeta42 as diagnostic markers for Alzheimer disease in clinical practice. Arch Neurol. 2001; 58(3):373-9.

Andreasen NC, Blennow K. Beta-amyloid (Abeta) protein in cerebrospinal fluid as a biomarker for Alzheimer's disease. Peptides. 2002; 23:1205-14.

Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet. 2002; 3(4):299-309.

Authier F, Bergeron JJ, Ou WJ, Rachubinski RA, Posner BI, Walton PA. Degradation of the cleaved leader peptide of thiolase by a peroxisomal proteinase. Proc Natl Acad Sci U S A. 1995; 92(9):3859-63.

Authier F, Cameron PH, Taupin V. Association of insulin-degrading enzyme with a 70 kDa cytosolic protein in hepatoma cells. Biochem J. 1996; 319(Pt 1):149-58.

Authier F, Posner BI, Bergeron JJ. Insulin-degrading enzyme. Clin Invest Med. 1996; 19(3):149-60.

Bacanu SA, Devlin B, Roeder K. The power of genomic control. Am J Hum Genet. 2000; 66(6):1933-44.

Bertram L, Blacker D, Mullin K, Keeney D, Jones J, Basu S, Yhu S, McInnis MG, Go RC, Vekrellis K, Selkoe DJ, Saunders AJ, Tanzi RE. Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. Science. 2000; 290(5500):2302-3.

Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. J. Clin. Invest. 2005; 115(6):1449-57.

Blomqvist ME, Andreasen N, Bogdanovic N, Blennow K, Brookes AJ, Prince JA. Genetic variation in CTNNA3 encoding alpha-3 catenin and Alzheimer's disease. Neurosci Lett. 2004; 358(3):220-2.

Blomqvist ME, Silburn PA, Buchanan DD, Andreasen N, Blennow K, Pedersen NL, Brookes AJ, Mellick GD, Prince JA. Sequence variation in the proximity of IDE may impact age at onset of both Parkinson disease and Alzheimer disease. Neurogenetics. 2004; 5(2):115-9.

Blomqvist ME, Chalmers K, Andreasen N, Bogdanovic N, Wilcock GK, Cairns NJ, Feuk L, Brookes AJ, Love S, Blennow K, Kehoe PG, Prince JA. Sequence variants of IDE are associated with the extent of beta-amyloid deposition in the Alzheimer's disease brain. Neurobiol Aging. 2005; 26(6):795-802.

Blomqvist ME, Reynolds C, Katzov H, Feuk L, Andreasen N, Bogdanovic N, Blennow K, Brookes AJ, Prince JA. Towards compendia of negative

genetic association studies: an example for Alzheimer disease. Hum Genet. 2006; 119(1-2):29-37.

Boussaha M, Hannequin D, Verpillat P, Brice A, Frebourg T, Campion D. Polymorphisms of Insulin degrading enzyme gene are not associated with Alzheimer's disease. Neurosci Lett. 2002; 329(1):121-3.

Brookes AJ. The essence of SNPs. Gene. 1999; 234(2):177-86.

Brookes AJ, Prince JA. Genetic association analysis: lessons from the study of Alzheimers disease. Mutat Res. 2005; 573(1-2):152-9.

Brown TA. Genomes. John Wiley & Sons Inc. 1999. Bios Scientific Publishers Ltd, Oxford, U.K.

Bulmer MG. The effect of selection on genetic variability. Am Nat. 1971; 105(943):201-11.

Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet. 2005; 366(9489):941-51.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. Natural selection on protein-coding genes in the human genome. Nature. 2005; 437(7062):1153-7.

Butler JM. Forensic DNA Typing, 2nd Ed. Academic Press, London, GB. 2005; pp 181-4.

Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, Thomas-Anterion C, Michon A, Martin C, Charbonnier F, Raux G, Camuzat A, Penet C, Mesnage V, Martinez M, Clerget-Darpoux F, Brice A, Frebourg T. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. Am J Hum Genet. 1999; 65(3):664-70.

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

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature. 2004; 429(6990):446-52.

Catania C, Sotiropoulos I, Silva R, Onofri C, Breen KC, Sousa N, Almeida OF. The amyloidogenic potential and behavioral correlates of stress. 2009; 14(1):95-105.

Chalmers K, Wilcock GK, Love S. APOE epsilon 4 influences the pathological phenotype of Alzheimer's disease by favouring cerebrovascular over parenchymal accumulation of A beta protein. Neuropathol Appl Neurobiol. 2003; 29(3):231-8.

Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. Nature. 2008; 455(7215):919-23. Review.

Corbo RM, Scacchi R. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet. 1999; 63(Pt 4):301-10.

Cordell HJ, Clayton DG. Genetic association studies. Lancet. 2005; 366(9491):1121-31.

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993; 261(5123):921-3.

Crapper DR, Krishnan SS, Dalton AJ. Brain aluminum distribution in Alzheimer's disease and experimental neurofibrillary degeneration. Science. 1973; 180(85):511-3.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet. 2001; 29(2):229-32.

Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995; 29(2):311-22.

Duckworth WC, Bennett RG, Hamel FG. Insulin degradation: progress and potential. Endocr Rev. 1998; 19(5):608-24.

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet. 1991; 337(8746):867-72.

Eckman EA, Reed DK, Eckman CB. Degradation of the Alzheimer's amyloid beta peptide by endothelin-converting enzyme. J Biol Chem. 2001; 276(27):24540-8.

Edbauer D, Willem M, Lammich S, Steiner H, Haass C. Insulin-degrading enzyme rapidly removes the beta-amyloid precursor protein intracellular domain (AICD). J Biol Chem. 2002; 277(16):13389-93.

Ertekin-Taner N, Graff-Radford N, Younkin LH, Eckman C, Baker M, Adamson J, Ronald J, Blangero J, Hutton M, Younkin SG. Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. Science. 2000; 290(5500):2303-4.

Ertekin-Taner N, Ronald J, Asahara H, Younkin L, Hella M, Jain S, Gnida E, Younkin S, Fadale D, Ohyagi Y, Singleton A, Scanlin L, de Andrade M, Petersen R, Graff-Radford N, Hutton M, Younkin S. Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. Hum Mol Genet. 2003; 12(23):3133-43.

Ertekin-Taner N, Allen M, Fadale D, Scanlin L, Younkin L, Petersen RC, Graff-Radford N, Younkin SG. Genetic variants in a haplotype block spanning IDE are significantly associated with plasma Abeta42 levels and risk for Alzheimer disease. Hum Mutat. 2004; 23(4):334-42.

Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. PLoS Genet. 2007; 3(10):1787-99.

Farrer LA, Myers RH, Cupples LA, St George-Hyslop PH, Bird TD, Rossor MN, Mullan MJ, Polinsky R, Nee L, Heston L, et al. Transmission and age-at-onset patterns in familial Alzheimer's disease: evidence for heterogeneity. Neurology. 1990; 40(3 Pt 1):395-403.

Farris W, Mansourian S, Chang Y, Lindsley L, Eckman EA, Frosch MP, Eckman CB, Tanzi RE, Selkoe DJ, Guenette S. 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. 2003; 100(7):4162-7.

Ferreon AC, Gambin Y, Lemke EA, Deniz AA. Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. Proc Natl Acad Sci U S A. 2009; 106(14):5645-50.

Feuk L. *SNP based strategies to study candidate genes for Alzheimer disease*. 2002. Thesis. Center for genomics and bioinformatics, Karolinska Institute, Sthlm, Sweden.

Finckh U, von der Kammer H, Velden J, Michel T, Andresen B, Deng A, Zhang J, Muller-Thomsen T, Zuchowski K, Menzer G, Mann U, Papassotiropoulos A, Heun R, Zurdel J, Holst F, Benussi L, Stoppe G, Reiss J, Miserez AR, Staehelin HB, Rebeck GW, Hyman BT, Binetti G, Hock C, Growdon JH, Nitsch RM. Genetic association of a cystatin C gene polymorphism with late-onset Alzheimer disease. Arch Neurol. 2000; 57(11):1579-83.

Flaten TP. Aluminium as a risk factor in Alzheimer's disease, with emphasis on drinking water. Brain Res Bull. 2001; 55(2):187-96.

Fratiglioni L, Grut M, Forsell Y, Viitanen M, Grafstrom M, Holmen K, Ericsson K, Backman L, Ahlbom A, Winblad B. Prevalence of Alzheimer's disease and other dementias in an elderly urban population: relationship with age, sex, and education. Neurology. 1991; 41(12):1886-92.

Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, Fu G, Saitou N, Osoegawa K, de Jong PJ, Suto Y, Hattori M, Sakaki Y. Construction and analysis of a human-chimpanzee comparative clone map. Science. 2002; 295(5552):131-4.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The Structure of Haplotype Blocks in the Human Genome. Science. 2002; 296(5576):2225-9.

Galimberti D, Venturelli E, Gatti A, Lovati C, Fenoglio C, Mariani C, Forloni G, Bresolin N, Scarpini E. Association of neuronal nitric oxide

synthase C276T polymorphism with Alzheimer's disease. J Neurol. 2005; 252(8):985-6.

Geerlings MI, den Heijer T, Koudstaal PJ, Hofman A, Breteler MM. History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease. Neurol. 2008; 70(15):1258-64.

Glenner GG, Wong CW. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. Biochem Biophys Res Commun. 1984; 120(3):885-90.

Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991; 349(6311):704-6.

Goldstein DB. Islands of linkage disequilibrium. Nat Genet. 2001; 29(2):109-11.

Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet. 2001; 69(6):1357-69.

Grant WB, Campbell A, Itzhaki RF, Savory J. The significance of environmental factors in the etiology of Alzheimer's disease. J Alzheimers Dis. 2002; 4(3):179-89.

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. An introduction to genetic analysis. 7th Ed. W.H. Freeman and company. 2000. NY, USA.

Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S, Iovannisci DM, Woo S, Winn-Deen ES. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res. 1994; 22(21):4527-34.

Hardy GH. Mendelian proportions in a mixed population. Science. 1908; 28(706): 49–50.

Hardy J, Allsop D. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. Trends Pharmacol Sci. 1991; 12(10):383-8.

Hardy J. Impact of genetic analysis on Parkinson's disease research. Mov Disord. 2003; 18 Suppl 6:S96-8.

Hartl DL, Clark AG. Principles of population genetics. 3rd Ed, Chap. 3. Sinauer Assoc. Inc. 1997. Sunderland, USA.

Hattersley AT, McCarthy MI. What makes a good genetic association study? Lancet. 2005; 366(9493):1315-23.

Heinzen EL, Need AC, Hayden KM, Chiba-Falek O, Roses AD, Strittmatter WJ, Burke JR, Hulette CM, Welsh-Bohmer KA, Goldstein DB. Genome-Wide Scan of Copy Number Variation in Late-Onset Alzheimer's Disease. J Alzheimers Dis. 2009 PMID: 19749422. [Epub ahead of print].

Helisalmi S, Hiltunen M, Vepsäläinen S, Iivonen S, Mannermaa A, Lehtovirta M, Koivisto AM, Alafuzoff I, Soininen H. Polymorphisms in neprilysin gene affect the risk of Alzheimer's disease in Finnish patients. J Neurol Neurosurg Psychiatry. 2004; 75(12):1746-8.

Higuchi S, Arai H, Matsushita S, Matsui T, Kimpara T, Takeda A, Shirakura K. Mutation in the alpha-synuclein gene and sporadic Parkinson's disease, Alzheimer's disease, and dementia with lewy bodies. Exp Neurol. 1998; 153(1):164-6.

Hill WG. Estimation of linkage disequilibrium in randomly mating populations. Heredity. 1974; 33(2):229-39.

Hill WG, Robertson A. The effects of inbreeding at loci with heterozygote advantage. Genetics. 1968; 60(3):615-28.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM. Control of confounding of genetic associations in stratified populations. Am J Hum Genet. 2003; 72(6):1492-1504.

Howell WM, Jobs M, Gyllensten U, Brookes AJ. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. Nat Biotechnol. 1999; 17(1):87-8.

Howell WM, Jobs M, Brookes AJ. iFRET: an improved fluorescence system for DNA-melting analysis. Genome Res. 2002; 12(9):1401-7.

Howell WM. *SNP technology and Alzheimer disease.* 2003. Thesis. Center for genomics and bioinformatics, Karolinska Institute, Sthlm, Sweden.

Hu J, Igarashi A, Kamata M, Nakagawa H. Angiotensin-converting enzyme degrades Alzheimer amyloid beta-peptide (A beta ); retards A beta aggregation, deposition, fibril formation; and inhibits cytotoxicity. J Biol Chem. 2001; 276(51):47863-8.

Hulse RE, Ralat LA, Wei-Jen T. Structure, function, and regulation of insulin-degrading enzyme. Vitam Horm. 2009; 80:635-48.

Huss A, Spoerri A, Egger M, Röösli M; Swiss National Cohort Study. Residence near power lines and mortality from neurodegenerative diseases: longitudinal study of the Swiss population. Am J Epidemiol. 2009; 169(2):167-75.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36(9):949-51.

Im H, Manolopoulou M, Malito E, Shen Y, Zhao J, Neant-Fery M, Sun CY, Meredith SC, Sisodia SS, Leissring MA, Tang WJ. Structure of substrate-free human insulin-degrading enzyme (IDE) and biophysical analysis of ATP-induced conformational switch of IDE. J Biol Chem. 2007; 282(35):25453-63.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409, 860 - 921.

International Human Genome Sequencing Consortium. 2004. Finishing the Euchromatic Sequence of the Human Genome. Nature 431, 931-45.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001; 29(3):306-9.

Iwata N, Tsubuki S, Takaki Y, Watanabe K, Sekiguchi M, Hosoki E, Kawashima-Morishima M, Lee HJ, Hama E, Sekine-Aizawa Y, Saido TC. Identification of the major Abeta1-42-degrading catabolic pathway in brain parenchyma: suppression leads to biochemical and pathological deposition. Nat Med. 2000; 6(2):143-50.

Jarvik A, Greenson H. Translation of 'About a peculiar disease of the cerebral cortex. By Alois Alzheimer, 1907'. Alzheimer Disease and Associated Disorders. 1987; 1(1):3-8.

Jobs M, Howell WM, Stromqvist L, Mayr T, Brookes AJ. DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. Genome Res. 2003; 13(5):916-24.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet. 2001; 29(2):233-7.

Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res. 2000; 10(10):1435-44.

Katzov H, Chalmers K, Palmgren J, Andreasen N, Johansson B, Cairns NJ, Gatz M, Wilcock GK, Love S, Pedersen NL, Brookes AJ, Blennow K, Kehoe PG, Prince JA. Genetic variants of ABCA1 modify Alzheimer disease risk and quantitative traits related to beta-amyloid metabolism. Hum Mutat. 2004; 23(4):358-67.

Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, Lovestone S, Perez-Tur J, Hutton M, Chartier-Harlin MC, Shears S, Roehl K, Booth J, Van Voorst W, Ramic D, Williams J, Goate A, Hardy J, Owen MJ. A full genome scan for late onset Alzheimer's disease. Hum Mol Genet. 1999; 8(2):237-45.

Koo EH, Sisodia SS, Archer DR, Martin LJ, Weidemann A, Beyreuther K, Fischer P, Masters CL, Price DL. Precursor of amyloid protein in Alzheimer disease undergoes fast anterograde axonal transport. Proc Natl Acad Sci U S A. 1990; 87(4):1561-5.

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet. 1999; 22(2):139-44.

Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet. 2001; 27(3):234-6.

Kurochkin IV, Goto S. Alzheimer's beta-amyloid peptide specifically interacts with and is degraded by Insulin degrading enzyme. FEBS Lett. 1994; 345(1):33-7.

Kivipelto M, Helkala EL, Laakso MP, Hanninen T, Hallikainen M, Alhainen K, Soininen H, Tuomilehto J, Nissinen A. Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study. BMJ. 2001; 322(7300):1447-51.

Leder P. Moving genes. Prog Clin Biol Res. 1982; 85(Pt A):41-50.

Levin-Allerhand JA, Lominska CE, Smith JD. Increased amyloid- levels in APPSWE transgenic mice treated chronically with a physiological high-fat high-cholesterol diet. J Nutr Health Aging. 2002; 6(5):315-9.

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, van Duinen SG, Bots GT, Luyendijk W, Frangione B. Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science. 1990; 248(4959):1124-6.

Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science. 1995; 269(5226):973-7.

Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 1964; 49: 49-67.

Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Allen FA Jr, Goetz CG, Mastaglia F, Stajich JM, Gibson RA, Middleton LT, Saunders AM, Scott BL, Small GW, Nicodemus KK, Reed AD, Schmechel DE, Welsh-Bohmer KA, Conneally PM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. Age at onset in two common

neurodegenerative diseases is genetically controlled. Am J Hum Genet. 2002; 70(4):985-93.

Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl. 1995; 4(6):357-62.

Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal. 1999; 14(5-6):143-9.

Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet. 1998; 19(3):225-32.

Llovera RE, de Tullio M, Alonso LG, Leissring MA, Kaufman SB, Roher AE, de Prat Gay G, Morelli L, Castaño EM. The catalytic domain of insulin-degrading enzyme forms a denaturant-resistant complex with amyloid beta peptide: implications for Alzheimer disease pathogenesis. J Biol Chem. 2008; 283(25):17039-48.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003; 33(2):177-82.

Malito E, Hulse RE, Tang WJ. Amyloid beta-degrading cryptidases: Insulin degrading enzyme, presequence peptidase, and neprilysin. Cell Mol Life Sci. 2008; 65(16):2574-85.

Maraganore DM, Lesnick TG, Elbaz A, Chartier-Harlin MC, Gasser T, Kruger R, Hattori N, Mellick GD, Quattrone A, Satoh J, Toda T, Wang J, Ioannidis JP, de Andrade M, Rocca WA; UCHL1 Global Genetics Consortium. UCHL1 is a Parkinson's disease susceptibility gene. Ann Neurol. 2004; 55(4):512-21.

Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, Pahwa R, Stern MB, Colcher A, Hiner BC, Jankovic J, Ondo WG, Allen FH Jr, Goetz CG, Small GW, Masterman D, Mastaglia F, Laing NG, Stajich JM, Ribble RC, Booze MW, Rogala A, Hauser MA, Zhang F, Gibson RA, Middleton LT, Roses AD, Haines JL, Scott BL, Pericak-Vance MA, Vance

JM. Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. JAMA. 2001; 286(18):2245-50.

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet. 2000; 67(2):383-94.

Masliah E, Rockenstein E, Veinbergs I, Sagara Y, Mallory M, Hashimoto M, Mucke L. Beta-amyloid peptides enhance alpha-synuclein accumulation and neuronal deficits in a transgenic mouse model linking Alzheimer's disease and Parkinson's disease. Proc Natl Acad Sci U S A. 2001; 98(21):12245-50.

Masters CL, Simms G, Weinman NA, Multhaup G, McDonald BL, Beyreuther K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. Proc Natl Acad Sci U S A. 1985; 82(12):4245-9.

Mendel G. Versuche über Plflanzen-hybriden.Verhandlungen des naturforschenden Ver-eines in Brünn, Bd. IV für das Jahr 1865, Abhand-lungen, 3–47 (1866) (Bateson translation).

Michikawa M. Cholesterol paradox: is high total or low HDL cholesterol level a risk for Alzheimer's disease? J Neurosci Res. 2003; 72(2):141-6.

Morgan TH. Random segregation versus coupling in Mendelian inheritance. Science. 1911; 34(873):384.

Myers A, Holmans P, Marshall H, Kwon J, Meyer D, Ramic D, Shears S, Booth J, DeVrieze FW, Crook R, Hamshere M, Abraham R, Tunstall N, Rice F, Carty S, Lillystone S, Kehoe P, Rudrasingham V, Jones L, Lovestone S, Perez-Tur J, Williams J, Owen MJ, Hardy J, Goate AM. Susceptibility locus for Alzheimer's disease on chromosome 10. Science. 2000; 290(5500):2304-5.

Nickerson DA, Kaiser R, Lappin S, Stewart J, Hood L, Landegren U. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. Proc Natl Acad Sci U S A. 1990; 87(22):8923-7.

Niu, T, Qin, ZS, Xu, X, Liu, JS. Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms. Am J Hum Genet. 2002; 70(1):157–69.

Notkola IL, Sulkava R, Pekkanen J, Erkinjuntti T, Ehnholm C, Kivinen P, Tuomilehto J, Nissinen A. Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease. Neuroepidemiology. 1998; 17(1):14-20.

Nussbaum RL, Ellis CE. Alzheimer's Disease and Parkinson's Disease. N Engl J Med. 2003; 348(14):1356-64.

Nyren P, Pettersson B, Uhlen M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem. 1993; 208(1):171-5.

Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet. 2005; 366(9492):1223-34.

Papassotiropoulos A, Streffer JR, Tsolaki M, Schmid S, Thal D, Nicosia F, Iakovidou V, Maddalena A, Lutjohann D, Ghebremedhin E, Hegi T, Pasch T, Traxler M, Bruhl A, Benussi L, Binetti G, Braak H, Nitsch RM, Hock C. Increased brain beta-amyloid load, phosphorylated tau, and risk of Alzheimer disease associated with an intronic CYP46 polymorphism. Arch Neurol. 2003; 60(1):29-35.

Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res. 2000; 10(7):1031-42.

Poirier J. Apolipoprotein E in animal models of CNS injury and in Alzheimer's disease. Trends Neurosci. 1994; 17(12):525-30.

Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. Genome Res. 2001; 11(1):152-62.

Prince JA, Feuk L, Gu HF, Johansson B, Gatz M, Blennow K, Brookes AJ. Genetic variation in a haplotype block spanning IDE influences Alzheimer disease. Hum Mutat. 2003; 22(5):363-71.

Prince JA, Zetterberg H, Andreasen N, Marcusson J, and Blennow K. APOE 4 allele is associated with reduced cerebrospinal fluid levels of Aß42. Neurology. 2004; 62(11): 2116-8.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000; 67(1):170-81.

Przeworski M. The signature of positive selection at randomly chosen loci. Genetics. 2002; 160(3):1179-89.

Qiu WQ, Walsh DM, Ye Z, Vekrellis K, Zhang J, Podlisny MB, Rosner MR, Safavi A, Hersh LB, Selkoe DJ. Insulin-degrading enzyme regulates extracellular levels of amyloid beta-protein by degradation. J Biol Chem. 1998.; 273(49):32730-8.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature. 2006; 444(7118):444-54.

Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol. 2001; 20(1):4-16.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. Nature. 2001; 411(6834):199-204.

Rice JP, Saccone NL, Corbett J. The lod score method. Adv Genet. 2001; 42:99-113.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273(5281):1516-7.

Roberts GW, Gentleman SM, Lynch A, *et al*. Beta amyloid protein deposition in the brain after severe head injury: implications for the

pathogenesis of Alzheimer's disease. J Neurol Neurosurg Psychiatry 1994; 57(4):419-25.

Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T, et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature. 1995; 376(6543):775-8.

Sackett DL. Bias in analytic research. J Chronic Dis. 1979; 32(1-2):51-63.

Samiotaki M, Kwiatkowski M, Parik J, Landegren U. Dual-color detection of DNA sequence variants by ligase-mediated analysis. Genomics. 1994; 20(2):238-42.

Schork NJ, Fallin D, Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet. 2000; 58(4):250-64.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science. 2004; 305(5683):525-8.

Selkoe DJ. Clearing the brain's amyloid cobwebs. Neuron. 2001; 32(2):177-80.

Shen Y, Joachimiak A, Rosner MR, Tang WJ. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. Nature. 2006; 443(7113):870-4.

Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature. 1995; 375(6534):754-60.

Shi J, Zhang S, Tang M, Ma C, Zhao J, Li T, Liu X, Sun Y, Guo Y, Han H, Ma Y, Zhao Z. Mutation screening and association study of the neprilysin gene in sporadic Alzheimer's disease in Chinese persons. J Gerontol A Biol Sci Med Sci. 2005; 60(3):301-6.

Simchen S, Stamberg J. Fine and Coarse Controls of Genetic Recombination. Nature. 1969; 222: 329-32.

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. Nature. 1997; 388(6645):839-40.

Strachan T, Read AP. Human Molecular Genetics 3. Garland Science Publishers, 2003. Oxford, GB.

Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A. 1993; 90(5):1977-81.

Syvänen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet. 2001; 2(12):930-42.

Syvänen AC. Toward genome-wide SNP genotyping. Nat Genet Suppl. 2005; 37 Suppl:S5-10.

Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, Mayeux R, Langston JW. Parkinson disease in twins: an etiologic study. JAMA. 1999; 281(4):341-6.

Tanzi RE, Bertram L. New frontiers in Alzheimer's disease genetics. Neuron. 2001; 32(2):181-4.

Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 1984; 12(10):4127-38.

Terry RD. The fine structure of neurofibrillary tangles in Alzheimer disease. J Neuropathol Exp Neurol. 1963; 22(4):629-42.

The International HapMap Consortium. The International HapMap Project. Nature 426: 789-96. 2003.

The International HapMap Consortium. A haplotype map of the human genome. Nature, 437:1299-1320. 2005.

Tiraboschi P, Hansen LA, Thal LJ, Corey-Bloom J. The importance of neuritic plaques and tangles to the development and evolution of AD. Neurology. 2004; 62(11):1984-9.

Venter JC, Adams MD, Myers EW, Li PW, et al. The sequence of the human genome. Science. 2001; 291(5507):1304-51.

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953; 171(4356):737-8.

Weinberg W. Über den Nachweis der Verebung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg. 1908; 64:368–82.

Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? Nat Genet. 2000; 26(2):151-7.

Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. Trends Genet. 2002; 18(1):19-24.

Weir BS. Linkage disequilibrium and association mapping. Review. Annu Rev Genomics Hum Genet. 2008; 9:129-42.

Wilson RS, Arnold SE, Beck TL, Bienias JL, Bennett DA. Change in depressive symptoms during the prodromal phase of Alzheimer disease. Arch Gen Psychiatry. 2008; 65(4):439-45.

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. Am J Hum Genet. 2007; 80(1):91-104.

Yoshimoto M, Iwai A, Kang D, Otero DA, Xia Y, Saitoh T. NACP, the precursor protein of the non-amyloid beta/A4 protein (A beta) component of Alzheimer disease amyloid, binds A beta and stimulates A beta aggregation. Proc Natl Acad Sci U S A. 1995; 92(20):9141-5.

**Web site reference**

http://genomics.phrma.org/lexicon/index.2.html (accessed 21/05/03 at 11:29)

http://www.fao.org/biotech/index_glossary.asp (assessed 09/23/09 at 12:22)
http://www.genome.gov/ (assessed 06/03/09 at 15:42)

http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_040198.pdf (assessed 01/06/06 at 16.44)

http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=coffeebrk (assessed 09/10/10 at 09:53)

http://www.geneticalliance.org (assessed 09/07/11 at 13.24)

http://www.sickkids.ca/pdfs/4111-CNV_FAQ.pdf (assessed 09/14/11 at 16.02)