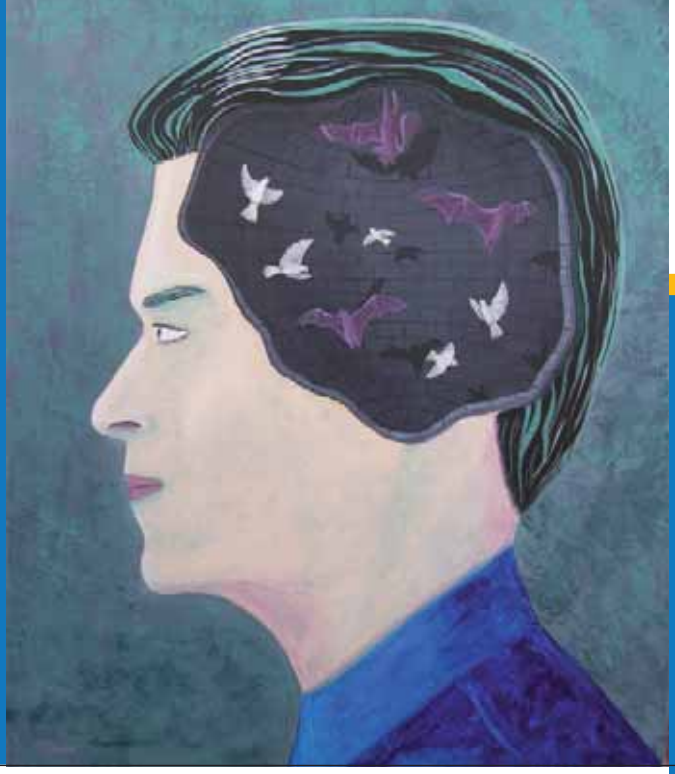

**Tero Ylisaukko-oja**

# Search for Susceptibility Genes in Autism Spectrum Disorders

Publications of the National Public Health Institute A 21/2005

Department of Molecular Medicine,
National Public Health Institute,
Helsinki, Finland
*and*
Department of Medical Genetics,
University of Helsinki, Finland

**Tero Ylisaukko-oja**

# SEARCH FOR SUSCEPTIBILITY GENES IN AUTISM SPECTRUM DISORDERS

A C A D E M I C  D I S S E R T A T I O N

*To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in the small lecture hall of the Haartman Institute, Haartmaninkatu 3, Helsinki, on November 11th, at 12 noon.*

Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

*and*

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2005

**Publications of the National Public Health Institute**
**KTL A21 / 2005**



**Julkaisija-Utgivare-Publisher**

**Kansanterveyslaitos (KTL)**
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

**Folkhälsoinstitutet**
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

**National Public Health Institute**
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-548-0
ISSN 0359-3584
ISBN 951-740-549-9 (pdf)
ISSN 1458-6290 (pdf)

**Kannen kuva - cover art:** Arto Ylisaukko-oja,"The Mystery of Consciousness", acrylic on canvas, 2005

Edita Prima Oy
Helsinki 2005

**Supervised by**

Academy Professor Leena Peltonen-Palotie
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Docent Irma Järvelä
Helsinki University Hospital,
Laboratory of Molecular Genetics and
University of Helsinki
Department of Medical Genetics
Helsinki, Finland

**Reviewed by**

Professor Peter Szatmari,
McMaster University,
Department of Psychiatry and
Behavioural Neuroscience,
Hamilton, Canada

Docent Tiinamaija Tuomi,
The Folkhälsan Research Center and
Helsinki University Hospital,
Division of Internal Medicine,
Helsinki Finland

**Opponent**

Professor Thomas Bourgeron,
Institut Pasteur,
Human Genetics and Cognitive Functions,
Paris, France

"*Kyllä te vietätte elämää! Korjaatte, puuhaatte ja hypitte aamusta iltaan. Mokoma hosuminen voi olla vaarallista. Johan sitä masentuu kun vain ajatteleekin kaikkia niitä, jotka tekevät työtä ja raatavat, ja mitä hyötyä siitä muka on. Eräs sukulaiseni luki trigonometriaa tuntokarvansa lerpalleen, ja kun hän oli oppinut kaiken, tuli Mörkö ja söi hänet suuhunsa. Joopa joo, Mörön vatsassa hän sitten lojui niin erinomaisen viisaana!*"

Tove Jansson, Muumipapan villi nuoruus, 1950

*"What a life! No end of changing and building up and pulling down again and jumping about. Such a lot of work may turn out to be really harmful. Oh, I'm dejected just to think of all the people who work and buzz and bumble about, and what it leads to. I had a cousin once who studied trigonometry until his whiskers drooped, and when he had learnt it all a Groke came and ate him up. Yes, so wise he was while lying in the Groke's stomach!"*

Tove Jansson, The Exploits of Moominpappa, Described by Himself, 1950

*To my family*

Tero Ylisaukko-oja, Search for Susceptibility Genes in Autism Spectrum Disorders
Publications of the National Public Health Institute, A21/2005, 133 Pages
ISBN 951-740-548-0; 951-740-549-9 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/suomi/julkaisut/julkaisusarjat/kansanterveyslaitoksen_julkaisuja_a/

# ABSTRACT

Autism (MIM 209850) is a severe neurodevelopmental disorder characterized by abnormalities in reciprocal social interaction and communication, restricted and stereotyped patterns of interests and activities, and the presence of developmental abnormalities, which are evident by the age of three years. Autism is the most severe form of a broad group of disorders known as pervasive developmental disorders (PDD) or autism spectrum disorders (ASD), which include also milder conditions such as Asperger syndrome (AS), disintegrative disorder, and atypical forms of autism. The population prevalence for autism is 10-20 per 10 000, whereas the total prevalence for all autism spectrum disorders is close to 10-60 per 10 000. Twin and family studies have repeatedly and unequivocally indicated a strong genetic component in ASDs, but the underlying genetic mechanisms are still largely unknown.

To analyze the genetic factors underlying autism, we have carried out two genome-wide scans in well characterized Finnish ASD families. First, we analyzed 38 families having diagnoses of autism, AS and developmental dysphasia, and identified a susceptibility locus at 3q25-27 with a maximum multipoint LOD score of 4.81. Also, other putative linkage findings were observed at 1q21-23 and Xq13. The second genome-wide scan was performed in a sample of 17 extended families with AS. The most solid linkage finding was observed at 3p14-24 with the highest multipoint $NPL_{all}$ score of 3.32. Other suggestive findings were identified at 1q21-23 and 13q31-33. Next, we analyzed the data from genome-wide scans performed in the US and Finnish autism samples combined, in order to reveal shared susceptibility loci for these two samples. The most promising shared locus was identified at 3p24-26 ($NPL_{all}$=2.20) in the vicinity of the earlier finding in the AS-sample.

Two members of the neuroligin gene family, *NLGN3* and *NLGN4*, have been shown to be mutated in individuals with autism, AS and mental retardation. Two neuroligin genes, *NLGN1* at 3q26 and *NLGN3* at Xq13, are located within the putative susceptibility loci identified in the first genome scan of this study. However, no evidence for functional mutations or significant allelic association was observed at *NLGN1*, *NLGN3* or *NLGN4* in the Finnish sample.

Recently, *DYX1C1* was identified as a candidate for developmental dyslexia in a Finnish study sample. It has been suggested that variants at this gene might confer susceptibility also for other disorders involving language development or acquiring specific competences. Therefore, we wanted to test whether the effect is present also in autism, which is an extreme example of such a disorder. No evidence for association or deviated allelic diversity in the haplotype analyses was observed, and thus, at least a presence of a major effect was excluded.

Taken together, the present study has involved the identification of potential susceptibility loci for ASDs at 1q21-23, 3p14-26, 3q25-27, 13q31-33, and Xq13. As the initial follow-up of the genome-wide scans, the most relevant candidates were analyzed but with no indication of etiological significance. The original articles included in this thesis represent the first step towards identification of susceptibility genes for ASDs in the Finnish population and this data provides a strong foundation for future fine-scale mapping of the identified loci.

Keywords: autism, Asperger syndrome, genetic mapping, linkage analysis, association analysis, neuroligin, oxytocin receptor, dyslexia, dysfasia

Tero Ylisaukko-oja, Search for Susceptibility Genes in Autism Spectrum Disorders
Publications of the National Public Health Institute, A21/2005, 133 Pages
ISBN 951-740-548-0; 951-740-549-9 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/suomi/julkaisut/julkaisusarjat/kansanterveyslaitoksen_julkaisuja_a/

# TIIVISTELMÄ

Autismikirjon sairaudet (MIM 209850) ovat vakavia lapsuusiällä alkavia sairauksia, joille on ominaista vastavuoroisen kommunikaation häiriöt, kielen kehityksen ongelmat, ja stereotyyppiset sekä toistuvat käyttäytymismallit. Autismikirjon sairauksiin kuuluvat lapsuusiän autismi, Aspergerin oireyhtymä ja epätyypilliset autismin muodot. Autismin yleisyys väestössä on 10-20 / 10 000, kun taas autismikirjon sairauksien kokonaisesiintyvyys on 10-60 / 10 000. Kaksos- ja perhetutkimukset ovat toistuvasti osoittaneet että autismikirjon sairaudet ovat vahvasti geneettisiä. Tästä huolimatta autismille altistavat geneettiset mekanismit ovat vielä epäselviä.

Tutkimuksen tarkoituksena oli paikantaa autismille altistavia geneettisiä tekijöitä genominlaajuisten kartoitustutkimusten avulla. Ensimmäisessä osatyössä analysoimme 38 hyvin kuvattua suomalaista perhettä, joissa esiintyi autismia, Aspergerin oireyhtymää ja kielenkehityksen viivettä (dysfasia). Lupaavin alttiusgeenialue löytyi kromosomista 3q25-27 (MLS=4.81). Muita mahdollisia alttiusgeenialueita paikansimme kromosomeista 1q21-23 sekä Xq13. Seuraavaksi suoritimme genominlaajuisen kartoituksen 17 laajassa suomalaisessa perheessä, jossa esiintyy Aspergerin oireyhtymää ainoana autismikirjon sairautena. Lupaavin kytkentälöydös oli kromosomissa 3p14-24 ($NPL_{all}$ = 3.32) ja muita mahdollisia alttiusgeenialueita paikansimme kromosomeista 1q21-23 sekä 13q31-33. Kolmannessa osatyössä analysoimme työn I geneettisen informaation yhdessä yhdysvaltalaisen Autism Genetic Resource Exchange (AGRE) – aineiston kanssa tarkoituksena etsiä näille aineistoille yhteisiä alttiusgeenialueita. Lupaavin yhteinen alttiusgeenialue löytyi kromosomista 3p24-26 ($NPL_{all}$ = 3.32). Tämä alttiusgeenialue on lähellä toisessa osatyössä paikannettua lupaavaa Asperger-lokusta.

Neuroligin-geeniperheeseen kuuluu viisi keskenään hyvin samankaltaista jäsentä, joista kahden (3 ja 4) on vastikään osoitettu mutatoituneen harvinaisissa tapauksissa autistisilla ja kehitysvammaisilla potilailla. Neuroligin-geeneistä kaksi, *NLGN1* kromosomissa 3q26 ja *NLGN3* kromosomissa Xq13, sijaitsevat ensimmäisessä osatyössä paikannetuilla alttiusgeenialueilla. *NLGN1, NLGN3 tai NLGN4* geenien toimintaan vaikuttavia mutaatioita tai alleeliassosiaatiota näillä alueilla ei kuitenkaan havaittu suomalaisissa autismiperheissä.

Suomalaisessa aineistossa on paikannettu lukihäiriölle mahdollisesti altistavia muutoksia kromosomissa 15 sijaitsevassa *DYX1C1*-geenissä. Kirjallisuudessa on esitetty, että tämän geenin muutokset saattavat altistaa myös laajemmalle kirjolle kehityksellisiä häiriöitä, etenkin kielenkehitykseen liittyen. Tutkimme tämän geenin alleeliassosiaatiota autismiaineistossamme, mutta emme löytäneet näyttöä *DYX1C1*-geenin osuudesta autismin syntyyn.

Tässä tutkimuksessa paikansimme mahdollisia autismikirjon sairauksien alttiusgeenialueita kromosomeista 1q21-23, 3p14-26, 3q25-27, 13q31-33 ja Xq13. Alustavana jatkotutkimuksena analysoimme selkeitä alttiusgeenikandidaatteja näiltä alueilta, mutta emme havainneet näyttöä näiden geenien osuudesta autismin syntyyn. Tässä tutkimuksessa esitetyt osatyöt ovat ensimmäinen askel kohti autismin alttiusgeenien paikantamista suomalaisessa väestössä ja ne luovat vahvan perustan kyseisten alueiden tihennyskartoitustutkimuksille.

Avainsanat: autismi, Aspergerin oireyhtymä, geenikartoitus, kytkentäanalyysi, assosiaatioanalyysi, neuroligin, oksitosiinireseptori, lukihäiriö, dysfasia

## CONTENTS

# LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, referred to in the text by roman numerals.

I. Auranen M, Vanhala R, Varilo T, Ayers K, Kempas E, Ylisaukko-oja T, Sinsheimer JS, Peltonen L, Järvelä I (2002) A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25-27. *American Journal of Human Genetics*, 71:777-790.

II. Ylisaukko-oja T, Nieminen-von Wendt T, Kempas E, Sarenius S, Varilo T, von Wendt L, Peltonen L, Järvelä I (2004) Genome-wide scan for loci of Asperger syndrome. *Molecular Psychiatry*, 9:161-168.

III. Ylisaukko-oja T, Alarcón M, Cantor RM, Auranen M, Vanhala R, Kempas E, von Wendt L, Järvelä I, Geschwind DH, Peltonen L. Search for autism loci by combined data analysis of AGRE and Finnish families. *Annals of Neurology,* in press.

IV. Ylisaukko-oja T, Rehnström K, Auranen M, Vanhala R, Alen R, Kempas E, Ellonen P, Turunen JA, Makkonen I, Riikonen R, Nieminen-von Wendt T, von Wendt L, Peltonen L, Järvelä I. Analysis of four neuroligin genes as candidates for autism. *European Journal of Human Genetics,* in press.

V. Ylisaukko-oja T, Peyrard-Janvid M, Lindgren CM, Rehnström K, Vanhala R, Peltonen L, Järvelä I, Kere J (2005) Family-based association study of *DYX1C1* variants in autism. *European Journal of Human Genetics*, 13:127-130.

Publication I has previously appeared in a thesis by Mari Auranen (2002) and publication II in a thesis by Taina Nieminen-von Wendt (2004). Some previously unpublished data are also presented.

# ABBREVIATIONS

| | |
|---|---|
| $\alpha$ | proportion of linked families / point-wise significance level |
| ADHD | attention deficit hyperactivity disorder |
| ADI-R | Autism Diagnostic Interview - Revised |
| AGRE | Autism Genetic Resource Exchange |
| AS | Asperger syndrome |
| ASD | autism spectrum disorder |
| ASP | affected sib-pair |
| *AUTS1* | autism susceptibility locus 1 |
| bp | base pair |
| CD/CV | common disease / common variant |
| cM | centiMorgan |
| CNS | central nervous system |
| DNA | deoxyribonucleic acid |
| DSM | Diagnostic and Statistical Manual of Mental Disorders |
| DZ | dizygotic |
| EST | expressed-sequence tag |
| FBAT | family-based association test |
| GSMA | Genome Search Meta-Analysis |
| $h^2$ | heritability |
| HGP | Human Genome Project |
| HRR | haplotype-relative risk |
| htSNP | haplotype-tagging SNP |
| IBD | identical by decent |
| ICD | International Classification of Diseases |
| IMGSAC | International Molecular Genetic Study of Autism Consortium |
| kb | kilobase |
| $\lambda_r$ | familial recurrence risk |
| LC | liability class |

| | |
|---|---|
| LD | linkage disequilibrium |
| LOD | logarithm of odds |
| Mb | megabase |
| MIM | Mendelian Inheritance in Man |
| miRNA | micro ribonucleic acid |
| MLS | maximum LOD score |
| MR | mental retardation |
| MZ | monozygotic |
| NPL | non-parametric linkage |
| NQA | not quite autism |
| PCR | polymerase chain reaction |
| PDD | pervasive developmental disorder |
| PDD-NOS | pervasive developmental disorder-not otherwise specified |
| PSD | phrase speech delay |
| QTL | quantitative trait locus |
| RNA | ribonucleic acid |
| RNAi | RNA interference |
| SNP | single nucleotide polymorphism |
| SSRI | selective serotonin reuptake-inhibitor |
| STR | short tandem repeat |
| TDT | transmission disequlibrium test |
| $\theta$ | recombination fraction |
| UTR | untranslated region |
| VNTR | variable number of tandem repeat |
| $Z_{max}$ | maximumLOD score |

# 1. INTRODUCTION

The genetic mapping approach offers tools for the identification of the genetic architecture of a disease, and thus, for understanding the biological mechanisms underlying the disease phenotype. Positional cloning has successfully been applied to detect causative variants for rare Mendelian disorders, which result from mutations of large functional effect. Inspired by this success, huge efforts are now directed to the genetic mapping of common phenotypes, such as asthma and allergy, cancer, cardiovascular disorders and mental disorders, which constitute a significant proportion of the public health burden. Common diseases are caused by a varying number of genetic, cultural, and environmental risk factors, each contributing a minor effect and probably having multiple interactions with each other. Identification of genetic variants for such phenotypes has turned out to be more difficult than initially thought, and it is now evident that revealing the genetic basis of these diseases requires substantial efforts in terms of sample collection and phenotyping as well as technological and statistical development.

Autism spectrum disorders, characterized by impairments in communication and social interaction as well as the presence of repetitive and restricted activities, are among the most strongly genetic neuropsychiatric disorders. Many rare causative genetic factors are known for autism, including chromosomal abnormalities, fragile X syndrome, and mutations in neuroligin genes. It is clear that functional characterisation of these pathways will provide novel information about the pathophysiology underlying autism. However, more prevalent causes are still unknown. Several genomewide linkage scans have been performed in order to reveal those chromosomal loci that might contain variants predisposing to autism spectrum disorders. These studies have resulted in some overlapping findings but, as in most other complex disorders, poor reproducibility has been the predominant result. Similarly, most of the candidate gene studies have resulted in conflicting results. However, this is expected given that the genetic background is likely to be a mixture of rare single-gene mutations and common variants contributing a

minor effect. It is therefore likely that most of the previous efforts have been underpowered to detect the predisposing genetic effects. Novel technological, statistical, and study design issues have been proposed to overcome these problems, as will be discussed in the following chapters.

# 2. REVIEW OF THE LITERATURE

## 2.1 Autism Spectrum Disorders

### *2.1.1 Historical background*

In 1943, Leo Kanner described 11 children, who had significant impairments in reciprocal social interactions and communication skills. Other clinical features included delayed early language development as well as repetitive and ritualistic interests and activities. This condition was named "infantile autism" (Kanner 1943). Asperger syndrome was first described in 1944 by Hans Asperger, who reported a group of boys with "autistic psychopathy". The clinical features of these boys included difficulties in social interaction and communication as well as circumscribed and restricted patterns of interest (Asperger 1944; Asperger 1991). Asperger was unaware of the previous work by Kanner (1943), in which these features were also included. The main differences between these early reports were that the patients described by Asperger were of higher intelligence, and the apparent onset of the condition seemed to be somewhat later than in autism. Furthermore, language skills were acquired on time in Asperger's patients and the motor deficits were more common. After the English review by Wing (Wing 1981), the combination of problems described by Asperger is generally referred to as Asperger syndrome (AS). It is worth noting that highly similar cases to those described by Asperger had been described already in 1926 by Eva Ssucharewa (translated in Ssucharewa and Wolff 1996). The patients in this early report were said to have a "schizoid personality disorder" and they were practically indistinguishable to those described later by Hans Asperger.

Nowadays, infantile autism is considered to be the prototype of disorders in a group called pervasive developmental disorders (PDD), which also include Asperger syndrome, childhood disintegrative disorder, Rett syndrome and pervasive developmental disorder-not otherwise specified (PDD-NOS) (World Health Organization 1993; American Psychiatric Association 1994). Autism, AS and PDD-NOS are also commonly referred to as autism spectrum disorders (ASD). While the criteria for

autism have been relatively well established for a long time, several alternative diagnostic concepts have been used to diagnose AS (Wing 1981; Tantam 1988; Gillberg and Gillberg 1989; Szatmari et al. 1989; World Health Organization 1993; American Psychiatric Association 1994). Therefore, the research findings concerning AS are sometimes difficult to interpret.

### *2.1.2 Diagnostic criteria and co-existing phenotypes*

The diagnoses of autism and AS are currently outlined in the International Classification of Diseases (ICD-10; World Health Organization 1993) and in the 4$^{th}$ edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association 1994). The diagnosis of autism is based on three core symptoms, which include (i) qualitative impairment in reciprocal social interaction, (ii) qualitative impairment in verbal and non-verbal communication and imaginative activity, and (iii) a markedly restricted repertoire of interests and activities (Table 1). These symptoms are present before the age of 3 years. The criteria for AS are the same as those for autism with three main exceptions: (i) the communication and imagination impairment criteria for autism are not listed for AS, (ii) individuals with AS do not have a clinically significant general delay in language and (iii) individuals with AS do not have a clinically significant delay in cognitive development (Table 2)(Nieminen-Von Wendt 2004). Furthermore, unlike autism, AS is usually recognized after the age of 3 years, or the problems in communication and social interactions before that age must not be of the type seen in autism (Volkmar and Klin 2000).

Numerous overlapping conditions commonly exist in patients with autism or AS. Some 75-80% of individuals with autism have mental retardation (MR) and 20-30% have epilepsy (Bailey et al. 1996; Gillberg & Billstedt 2000). Both autism and AS are commonly associated with sleeping problems (Gillberg and Billstedt 2000; Tani et al. 2003) and altered sensibility (Gillberg and Billstedt 2000; Nieminen-Von Wendt 2004). Motor clumsiness and prosopagnosia (face recognition difficulties) are commonly

observed in individuals with AS (Gillberg and Billstedt 2000; Smith 2000; Nieminen-Von Wendt 2004). In addition, several neuropsyciatric disorders, including attention deficit hyperactivity disorder (ADHD), tics and Tourette syndrome, depression, eating disorders, obsessive-compulsive behaviour, and schizophrenia spectrum disorders, may co-exist with the ASDs, especially with AS (Gillberg et al. 1996; Ghaziuddin et al. 1998; Gillberg and Billstedt 2000; Ringman and Jankovic 2000; Wolff 2000; Nieminen-Von Wendt 2004; Nieminen-von Wendt et al. 2005).

**Table 1. Diagnostic criteria for autism according to the ICD-10 classification** (World Health Organization 1993)**.**

---

**A.** **Presence of abnormal or impaired development before the age of three years, in at least one out of the following areas:**

1. receptive or expressive language as used in social communication
2. the development of selective social attachments or of reciprocal social interaction
3. functional or symbolic play

**B.** **Qualitative abnormalities in reciprocal social interaction, manifest in at least one of the following areas:**

1. failure adequately to use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction.
2. failure to develop (in a manner appropriate to mental age, and despite ample opportunities) peer relationships that involve mutual sharing of interests, activities and emotions.
3. A lack of socio-emotional reciprocity as shown by an impaired or deviant response to other people's emotions; or lack of modulation of behaviour according to social context, or a weak integration of social, emotional and communicative behaviours.

**C.** **Qualitative abnormalities on communication, manifest in at least two of the following areas:**

1. a delay in, or total lack of development of spoken language that is not accompanied by an attempt to compensate through the use of gesture or mime as alternative modes of communication (often preceded by a lack of communicative babbling)
2. relative failure to initiate or sustain conversational interchange (at whatever level of language skills are present) in which there is reciprocal to and from responsiveness to communications of the other person
3. Stereotyped and repetitive language or idiosyncratic use of words or phrases
4. abnormalities in pitch, stress, rate, rhythm and intonation of speech

**D.** **Restricted repetitive, and stereotyped patterns of behaviour, interests and activities, manifest in at least two of the following areas:**

1. an encompassing preoccupation with one or more stereotyped and restricted patterns of interest that are abnormal in content or focus; or one or more interests that are abnormal in their intensity and circumscribed nature although not abnormal in their content or focus.
2. apparently compulsive adherence to specific, non-functional, routines or rituals
3. stereotyped and repetitive motor mannerisms that involve either hand or finger flapping or twisting, or complex whole body movements
4. preoccupations with part-objects or non-functional elements of play materials (such as their odour, the feel of their surface, or the noise or vibration that they generate)
5. distress over changes in small non-functional, details of environment

**E.** **The clinical picture is not attributable to other varieties of pervasive developmental disorder; specific developmental disorder of receptive language (F80.2) with secondary socio-emotional problems; reactive attachment disorder (F94.1) or disinhibited attachment disorder (F94.2); mental retardation (F70-72) with some associated emotional or behavioural disorder; schizophrenia (F20) of unusually early onset; and Rett's syndrome (F84.2).**

---

**Table 2. Diagnostic criteria for Asperger syndrome according to the ICD-10 classification** (World Health Organization 1993)**.**

| | |
|---|---|
| **A.** | **A lack of any clinically significant general delay in spoken or receptive language or cognitive development. Diagnosis requires that single words should have developed by 2 years of age or earlier and that communicative phrases be used by 3 years of age or earlier. Self-help skills, adaptive behaviour, and curiosity about the environment during the first 3 years should be at a level consistent with normal intellectual development. However, motor milestones may be somewhat delayed and motor clumsiness is usual (although not a necessary diagnostic feature). Isolated special skills, often related to abnormal preoccupations, are common, but are not required for the diagnosis.** |
| **B.** | **Qualitative abnormalities in reciprocal social interaction (criteria as for autism). Diagnosis requires demonstrable abnormalities in at least two out of the following four areas:**<br>1. Failure adequately to use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction.<br>2. Failure to develop (in a manner appropriate to mental age, and despite ample opportunities) peer relationships that involve a mutual sharing of interests, activities and emotions.<br>3. Lack of socio-emotional reciprocity as shown by an impaired or deviant response to other people's emotions, and/or lack of modulation of behaviour according to social context, and/or a weak integration of social, emotional and communicative behaviours<br>4. Lack of spontaneous seeking to share enjoyment, interests or achievements with people (e.g. a lack of showing, bringing or pointing out to other people objects of interests to the individual) |
| **C.** | **An unusually intense circumscribed interests, or restricted, repetitive, and stereotyped patterns of behaviour, interests, and activities (criteria as for autism, however it would be less usual for these to include either motor mannerisms or preoccupations with part-objects or non-functional elements of play materials). Diagnosis requires demonstrable abnormalities in at least two out of the following four areas:**<br>1. An encompassing preoccupation with one or more stereotyped and restricted patterns of interests that is abnormal in context or focus; or one or more interests that are abnormal in their intensity and circumscribed nature though not in their content or focus.<br>2. Apparently compulsive adherence to specific, non-functional routines or rituals.<br>3. Stereotyped and repetitive motor mannerisms that involve either hand/finger flapping or twisting, or complex whole body movements.<br>4. Preoccupation with part-objects or non-functional elements of play materials. |
| **D.** | **The disorder is not attributable to other varieties of pervasive developmental disorder; schizotypal disorder (F21); simple schizophrenia (F20.6); reactive and disinhibited attachment disorders of childhood (F94.1 and .2); obsessional personality disorder (F60.5); obsessive-compulsive disorder (F42).** |

### *2.1.3 Prevalence of Autism Spectrum Disorders*

The prevalence for ASDs has traditionally been reported as 4-6 / 10 000 for strictly defined autism and 20 / 10 000 for the broad spectrum of ASDs (Charman 2002). It seems likely, however, that the current true prevalence for ASDs is considerably higher

than previously thought. For example, Chakrabarti and Fombonne (2001) reported a prevalence of 16.8 / 10 000 (95% CI, 11.0-24.6) for autism and a total prevalence of 62.6 / 10 000 (95% CI, 50.8-76.3) for all PDDs in an English study population of 15 500 children. An US study by Yeargin-Allsopp et al. (2003) reported a prevalence of 34 / 10 000 (95% CI, 32-36) for ASDs in a study population of 289 456 children in the Atlanta metropolitan area. Also, other recent epidemiological studies for ASDs, including one conducted in the Northern Finnish population (Kielinen et al. 2000), have yielded similar prevalence estimates (Charman 2002). The apparent increase in prevalence most probably arise from the increased recognition of ASDs, the broadening of the diagnostic concept over time, and methodological differences, e.g. novel screening methodology, which may powerfully affect the prevalence estimates. An alternative, but far more unlikely, explanation for the increased prevalence of ASDs is the recent emergence of some powerful but yet unidentified environmental factors (Chakrabarti and Fombonne 2001; Charman 2002; Yeargin-Allsopp et al. 2003). A highly consistent finding in the epidemiological studies is that there is a clear excess of males in both autism and AS with the male-female ratio being around three to one (Ehlers and Gillberg 1993; Chakrabarti and Fombonne 2001; Charman 2002; Yeargin-Allsopp et al. 2003).

So far, only one study has been conducted, in which the prevalence of AS has been investigated exclusively without taking other ASDs into account. A prevalence rate of 28.5 / 10 000 (95% CI, 0.6-56.5) was reported with ICD-10 criteria in a small study population of 1519 Swedish children. A slightly higher prevalence estimate of 36 / 10 000 was obtained when the criteria of Gillberg and Gillberg (1989) were used (Ehlers and Gillberg 1993). The prevalence estimates for AS in those studies that have focused on ASDs in general have ranged significantly from 0.3 to 48.4 / 10 000, indicating major diagnostic and methodological differences across the studies. Many of these studies have, for example, focused on such young age groups, that AS is usually not recognized. In addition, the diagnosis of AS is not well established in many countries potentially leading to too low prevalence estimates. It has been suggested that a conservative

estimate might be around 2 / 10 000, but it is probably fair to conclude that there are currently no reliable prevalence estimates for AS (Fombonne and Tidmarsh 2003).

## 2.2 Genetic mapping of complex diseases

Genetic mapping studies aim to detect correlation between the genotypes of marker loci and the phenotype of interest and to identify those genomic regions, which include the genetic variant(s) influencing the trait of interest. Ultimately, the goal is to be able to identify the specific predisposing genetic variant(s) among the ~$3x10^9$ nucleotides in the human genome. Conventional genetic mapping studies start by collecting family material of a sufficient size for linkage analysis. For this purpose, DNA from families with two or more affected individuals is needed. These samples are analysed in a genome-wide scan, which employs a genome-wide set of polymorphic genetic markers covering the entire human genome usually at 4-10 centiMorgan (cM) intervals (400-1000 genetic markers). Genetic mapping utilises a phenomenon called linkage, which means the tendency of two closely spaced loci in the genome to be inherited together rather than independently. In linkage analysis, the chromosomal localisation of a disease gene is statistically determined by comparing the inheritance of marker alleles with the inheritance of the disease within families. The obvious advantage of this approach is that it is based solely on the knowledge that the phenotype is inherited and no prior knowledge is needed about the disease pathogenesis. Therefore, identification of disease mutations is usually the primary route towards understanding the biological mechanisms underlying the disease phenotype (Botstein and Risch 2003).

Once a disease gene has been assigned to a certain chromosomal locus, fine-scale mapping by analysing more families and a denser set of genetic markers is usually performed. The region harbouring the disease-causing gene can be restricted by recombinations within the families (linkage) or, in the presence of linkage disequilibrium (LD), by the ancient recombinations (see, 2.2.6). In some cases, chromosomal abnormalities may provide a shortcut for disease gene identification (2.3.3). As a result

of the Human Genome Project (HGP) and the competitive genome project by Celera Genomics, the identification of human disease genes has changed dramatically (Lander et al. 2001; Venter et al. 2001). Bioinformatic approaches rather than laborious positional cloning process can be directly used to identify known genes, expressed-sequence tags (EST), or predicted genes from the region of interest. In addition, the availability of the complete sequences of several organisms enables comparative genomic approaches to identify new genes, regulatory elements, or other functionally important regions of the genome (Miller et al. 2004). The positional candidates are analysed in the patient samples in order to identify the disease-causing variant. Several methods are used to test the validity of identified variants as disease-causing mutations. These include e.g. testing association with the phenotype, comparative genomic approaches, *in silico* modelling of the protein structure and function, functional analyses *in vitro* and, ultimately, functional analyses in model organisms (Figure 1).

The most success in genetic mapping has been experienced in simple and rare Mendelian diseases (Botstein and Risch 2003). Mendelian diseases have clearly identifiable inheritance patterns, meaning that the disease must result from a single mutation of a large functional effect. By contrast, the success in the identification of genetic variants responsible for common complex phenotypes has not been impressive so far (Altmuller et al. 2001). Complex traits are much less clear in their inheritance patterns because these phenotypes result from the combined effects of many genetic variants, environmental factors and cultural factors (Risch 2000; Weiss and Terwilliger 2000; Glatt and Freimer 2002). Most problems encountered in disease mapping of complex diseases are caused by a lack of one-to-one correlation between genotype and the diagnostic disease phenotype (Figure 2). At present, several of the mapped complex disease genes account for an uncommon subset of generally more common disorders (Risch 2000).

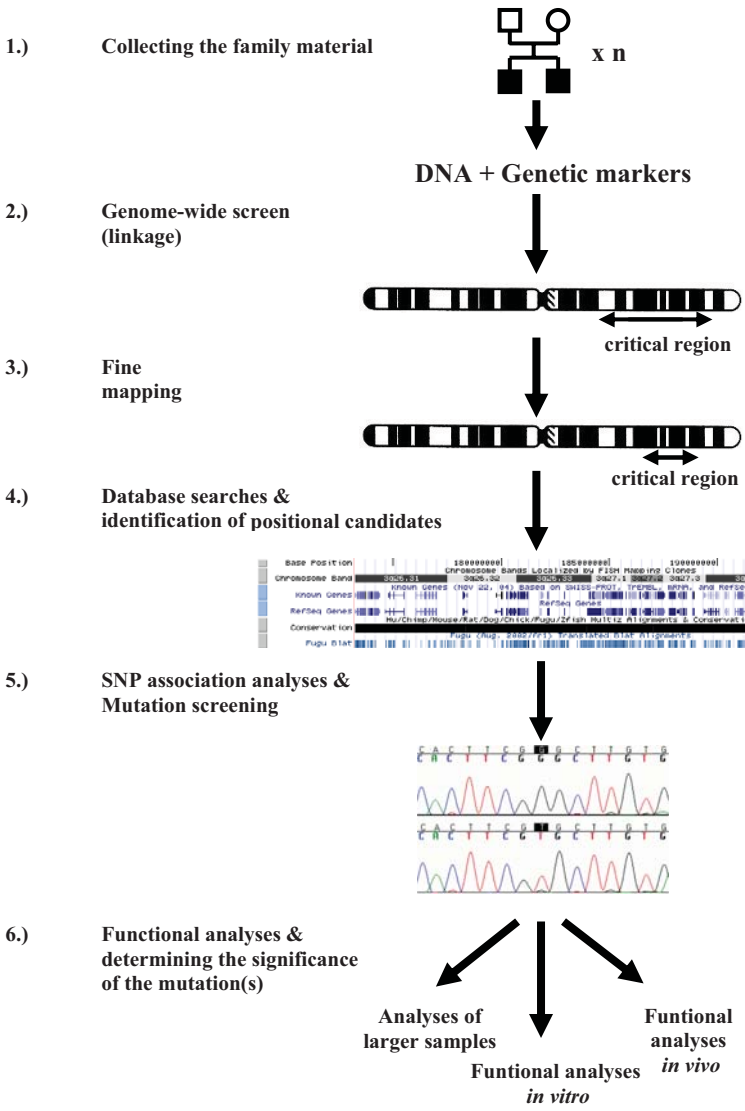


**Figure 1. Genetic mapping – an overview.** The gene mapping effort starts with a genomewide scan (1-2) and the positive regions are followed up by using more samples and a denser map of markers (3). Positional candidates are identified from the databases (4) and the genetic variants potentially conferring susceptibility are identified by SNP association and mutation analyses (5). The validity of identified variants is tested in large population samples and by functional analyses (6).

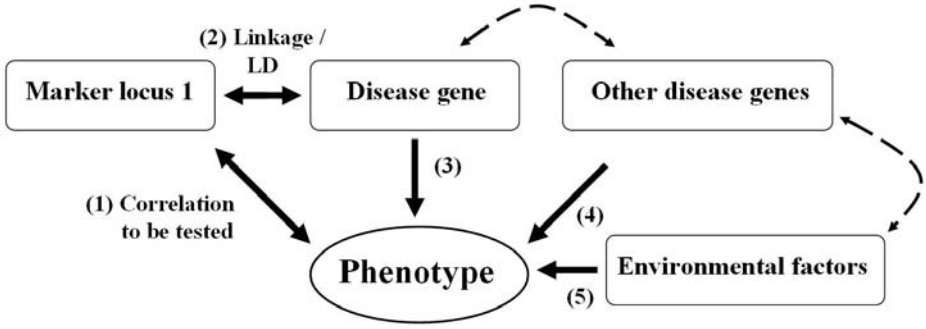


**Figure 2. A simplified model of the etiological factors predisposing to complex diseases.** In genetic mapping studies, the correlation between marker locus genotype and disease phenotype (1) is tested. Therefore, its success depends on both a correlation between the marker locus and the disease gene (linkage, 2) and also on a correlation between the disease gene and the phenotype (3). In Mendelian disorders, the correlation between the trait phenotype and disease gene (3) is strong, and the success depends primarily on how closely the tested marker is correlated with the disease gene. The existence of other predisposing factors, such as other susceptibility genes (4) and environmental and cultural factors (5), reduce the correlation between the disease gene and the phenotype (3) and, thus, the tested correlation (1). This is the situation usually observed in complex phenotypes. Modified from Weiss and Terwilliger (2000).

### *2.2.1 Determining the genetic component in complex disorders*

Many common disorders are caused by a varying number of susceptibility genes, environmental factors, or both combined. Only individuals, whose liability exceeds a certain threshold, manifest the disease (Figure 3). Because the affected individuals' relatives have a higher risk for predisposing factors than the general population, they also have, on average, a higher liability for the disease (Farrer and Cupples 1998).

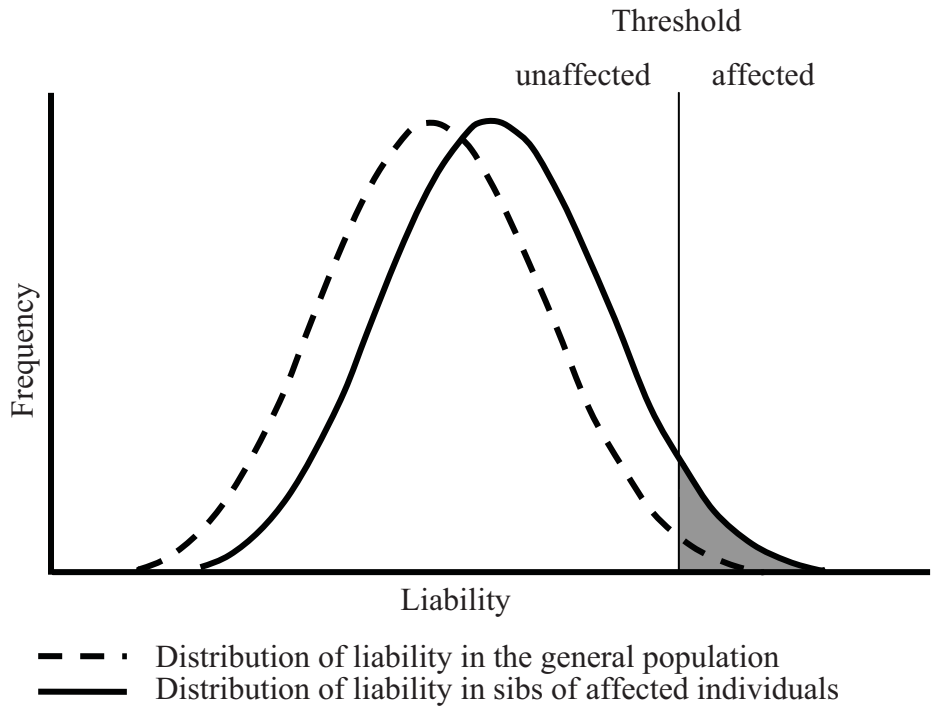


**Figure 3. The multifactorial threshold model.** The predisposing factors for the disease are normally distributed in the population. The individuals whose liability exceeds the threshold manifest the disease. The sibs of affected individuals have a higher average liability than the population mean and a greater proportion of them have liabilities exceeding the threshold (modified from Strachan and Read 2004).

The determination of a genetic component is a crucial step preceding genetic mapping studies because their success is highly dependent on the correlation between a genotype and a disease phenotype (Figure 2). The first step towards identification of the genetic component is the determination of familial aggregation. When a disease has genetic aetiology, the affected probands' first-degree relatives have an increased risk for the disease compared with the risk in more distant relatives or in the general population. Therefore, the degree of familial clustering can be measured as a familial recurrence

risk, $\lambda_r$, which is the ratio of the risk for the affected probands' relatives versus the disease prevalence in population ($\lambda_r=k_r/k$, where $\lambda_r$ is familial recurrence risk, $k_r$ is the risk of the affected proband's relatives, and $k$ is the prevalence in general population; Risch 1990). The most frequently used familial recurrence risk is sibling risk, $\lambda_s$, which is the recurrence risk in the probands' siblings versus the risk in the general population.

Familial aggregation of a trait is not a sufficient proof for genetic aetiology and more sophisticated methods have been developed to prove a genetic basis. Twin and adoption studies have been specially suited for diseases, where making a distinction between genetic and environmental factors is difficult. A classical twin study compares the concordance of monozygotic (MZ) and dizygotic (DZ) twins for a trait indicating the relative importance of genetic and environmental effects. The difference between MZ and DZ concordance indicates a strong genetic effect whereas equal concordance between MZ and DZ twins or low MZ concordance suggest strong environmental effects (Allen et al. 1967). Twin studies can be used to determine another important measure, heritability ($h^2$), which is defined as the proportion of the total variance of the phenotype that is genetic ($h^2=V_G/V_P$, where $V_G$ is genetic variance and $V_P$ is overall variance of the phenotype) (Strachan and Read 2004). If $r_{MZ}$ is a correlation in MZ twins and $r_{DZ}$ is a correlation in DZ twins, the heritability can be defined as $h^2=2(r_{MZ}-r_{DZ})$ (Boomsma et al. 2002).

Segregation analyses are used for determining whether a major genetic component is involved in the aetiology of a trait. A variety of both genetic and non-genetic models are tested, and the best model fitting the family data is estimated by likelihood ratio tests (Farrer and Cupples 1998; Strachan and Read 2004).

### *2.2.2 Genetic markers and maps*

Currently, the most common markers for genome scanning purposes are microsatellites (short tandem repeats, STR), which are di-, tri-, or tetranucleotide repeats flanked by

unique sequences on both sides (Litt and Luty 1989; Weber and May 1989). Utilisation single nucleotide polymorphisms (SNP) as a genetic mapping tool has gained much interest recently. SNPs are most densely distributed variations in the human genome occurring approximately once in every 300 bp (minor allele frequency >1%), which make them an optimal choice for fine-scale mapping (Kruglyak and Nickerson 2001). As a result from HGP, 1.4 million SNPs in the human genome were identified, but since then this number has been multiplied to nearly 9 million (Sachidanandam et al. 2001; The International HapMap Consortium 2003; Hirschhorn and Daly 2005). The mutation rates of SNPs are much lower than those of microsatellites, which makes them more optimal for association analyses where co-segregation of a marker allele and disease is monitored for many generations (Kruglyak 1997). Yet another advantage is that several highly automated high-throughput methods for SNP genotyping have been developed (Syvänen 2005).

The distances and marker ordering on genetic maps are based on meiotic recombination ($\theta$), whereas physical maps quantify the distances in number of base pairs (bp). Most of the currently available genetic maps are based solely on microsatellites (NIH/CEPH Collaborative Mapping Group 1992; Sheffield et al. 1995; Broman et al. 1998), but the most recently published genetic map has incorporated also SNP markers (Kong et al. 2002). The primary unit of distance along the genetic maps is centiMorgan (cM), which is equivalent to approximately 1% probability for recombination between two loci. As a rule of thumb, the equivalence between genetic and physical maps can be defined as one cM equalling one megabase (Mb)(Speer 1998). In reality, the recombination rate varies greatly along each chromosome from 0 to 9 cM per Mb (Yu et al. 2001).

The construction of a novel type of genetic map is currently underway. This so called HapMap involves characterization of haplotype structure and haplotype tagging SNPs of entire genomes in different populations from Africa, Asia and Europe (The International HapMap Consortium 2003). The rationale behind this effort is discussed more closely in 2.2.7.

### *2.2.3 Parametric linkage methods*

The aim of linkage analysis is to reveal those chromosomal loci, which harbour genetic variant(s) predisposing to a certain phenotype. Parametric linkage analysis (also referred to as *model-based* analysis) follows co-segregation of two genetic factors, the marker and the trait, at specific loci in pedigrees using the frequency of meiotic recombination as an estimate of genetic distance. The recombination fraction ranges from $\theta = 0$ for loci right next to each other through $\theta = ½$ for loci far apart or on different chromosomes. Two loci are said to be genetically linked when $\theta < ½$, in other words the two loci do not segregate independently. The object of linkage analysis is to test whether an observed deviation from 50% recombination between two loci is statistically significant (Terwilliger and Ott 1994; Ott 1999). The statistical measure for linkage analysis is the logarithm of odds (LOD) score (Morton 1955):

$$Z(\theta) = \log\left(\frac{L(\theta)}{L(0.5)}\right), \text{ where}$$

$Z(\theta)$ = LOD score

$L(\theta)$ = Likelihood function with given $\theta$

$L(0.5)$ = Likelihood function when $\theta = 0.5$

The likelihood function of the parametric linkage analysis requires specification of the disease parameters. These parameters include recombination fraction, marker allele-frequencies, penetrance and phenocopy rate (probability of disease when 2, 1, or 0 disease alleles are present), and the disease allele frequency. When complex diseases are considered, the true values of these parameters are unknown. The incorrect specification of the mode of inheritance may lead to loss of power and, consequently, type II errors (false negatives). Based on simulations, however, it has been proposed that parametric linkage analysis with both the recessive and dominant models is a robust method for detecting linkage in complex disorders, despite the fact that the model is not entirely correct (Greenberg et al. 1998; Xu et al. 1998; Abreu et al. 1999; Durner et al. 1999).

However, the number of statistical models tested should be modest, because a large number of tests increases the probability for type I errors (false-positives). Parametric LOD score analyses commonly also involve several analytical modifications when they are applied in the context of complex diseases including tests of heterogeneity, affected-only analyses, and utilisation of liability classes (Xu et al. 1998).

### *2.2.4 Non-parametric linkage methods*

Non-parametric methods (also referred to as *model-free* methods) are based on the allele-sharing of affected related individuals. The simplest study design involves a large sample of affected sib-pairs (ASP). ASP methods observe how frequently two affected offspring share copies of the same parental alleles, i.e. alleles that are shared identical by decent (IBD). In the absence of linkage each affected sib-pair is expected to share 0, 1, and 2 alleles IBD with respective probabilities of 0.25, 0.50 and 0.25. If the marker locus is linked to a disease locus, ASPs will share marker alleles more often than expected by chance. The deviation from the expected sharing is tested statistically by using e.g. $\chi^2$ or mean tests (Ott 1999). Currently, the most widely used non-parametric method is NPL-statistics implemented in the Genehunter-package, which overcomes several of the weaknesses evident in the earlier model-free methods. Like other non-parametric methods, NPL-statistics measures allele sharing among affected relatives within a pedigree. The main advantage is that NPL-statistics, specifically $NPL_{all}$-statistics, can consider all affected relatives simultaneously rather than as a combination of all possible comparisons of pairs. Another advantage of Genehunter is that it is inherently multipoint method enabling the use of a large number of markers in one analysis (Kruglyak et al. 1996). The main rationale for the wide use of non-parametric methods in the context of complex diseases is that there is no need for defining the mode of inheritance. Generally, NPL statistics tends to give overly conservative estimates of linkage when there is limited information available, but this problem is at least partly overcome by the multipoint approach (Kruglyak et al. 1996; Davis and Weeks 1997).

### *2.2.5 Determining the statistical significance in linkage analysis*

The point-wise significance level refers to a single test of the hypothesis of no linkage. This involves comparing the observed deviation at a specific locus with the probability that such a deviation is encountered by chance. The genome-wide significance level, by contrast, is the probability that one would encounter a deviation by change somewhere in a whole-genome scan (Lander and Kruglyak 1995). The traditional threshold for significance in two-point analyses of Mendelian disorders has been a LOD score of 3.0 (Ott 1999). This value corresponds to 1000:1 odds for linkage, which means that the observed data is 1000-fold more likely to arise under the alternative hypothesis of linkage than under the null hypothesis of no linkage. It has been shown elsewhere, that the LOD score of 3.0 corresponds to a point-wise p-value of 0.0001 (Xu et al. 1998; Ott 1999). Assuming that the genome-wide scan includes 400 markers and if these were treated as independent tests, the point-wise significance level of p = 0.0001 (LOD = 3.0) would correspond to a genome-wide significance level of $\alpha^* = 1-(1-0.0001)^{400} = 0.039$ (Bonferroni correction).

However, the tests in a genome-wide scan are not independent, but the tested markers are correlated with the nearby markers. According to Lander and Kruglyak (1995), the genome-wide significance level can be defined as follows:

$\alpha^* = 1-e^{-(C+2GX)\alpha}$, where

$\alpha^*$ = genome-wide significance level

$\alpha$ = point-wise significance level

C = number of chromosomes

G = the size of the genome in Morgans (M)

X = 4.61 x LOD score

If a genome-wide significance level of 0.05 is desired, the point-wise significance level should be $\alpha = 5 \times 10^{-5}$, which corresponds to a LOD score of 3.3 (Lander and Kruglyak

1995). According to standards proposed by Lander and Kruglyak (1995) this is considered as significant evidence for linkage in general pedigrees. Two additional thresholds were suggested: a point-wise p-value of $1.7 \times 10^{-3}$ corresponding to a LOD score of 1.9 for suggestive linkage, and a point-wise p-value of 0.05 corresponding to a LOD score of 0.59 for nominal evidence for linkage. For a sib-pair study, corresponding thresholds for nominal, suggestive, and significant linkage were suggested as 2.2, 3.6, and 5.4 (Lander and Kruglyak 1995). It is of interest that the approach by Lander & Kruglyak (1995) to determine significance levels has aroused criticism because they assume an infinitely dense set of markers and full marker information content, which is not the case in practice (Sawcer et al. 1997).

Permutation testing is highly recommended when evaluating genome-wide significance levels (Lander and Kruglyak 1995; Sawcer et al. 1997). In this approach, a large number of genome-wide scans are simulated under the null hypothesis of no linkage anywhere in the genome. The pedigree structures, marker allele frequencies and intermarker distances are derived from the actual study. Genome-wide significance level is determined by testing how many times the empirically observed LOD score is exceeded by change in a large number of genome-wide scans. In permutation testing, a larger number of iterations result in a more reliable estimate of genome-wide significance (Sawcer et al. 1997).

It is generally accepted that the replication of the initial significant linkage claim in an independent data set is required for the confirmed linkage (Lander and Kruglyak 1995). This has turned out to be a difficult task when dealing with complex diseases (Altmuller et al. 2001). Ascertainment protocols are used to increase the power of mapping, which means that the sample ascertained on a given phenotype is not representative of the population as a whole but, rather, some of the predisposing loci are likely to be over-represented in the study sample. In a genome-wide scan, LOD scores are maximised in the ascertained sample over several parameters and over numerous point-wise tests throughout the genome. Therefore, genome-wide studies necessarily overestimate the effect size of any locus identified, which leads to the conclusion that the power needed

for replication will presumably be underestimated. Likewise, it is unlikely that the same subset of predisposing factors would have enriched in the replication sample and, thus, significantly larger sample size is needed detect this signal in an independent replication sample than that used in the original study (Göring et al. 2001). In addition, many simulation studies have shown that chance variation in location estimate of genetic linkage studies is substantial. The displacement decreases as the sample size and/or marker information content increases. Such inaccuracy clearly has consequences on replication attempts. It also makes it difficult to estimate the true location of the predisposing variant and to determine whether the two linkage peaks on the same chromosome from separate study samples represent the same genetic variant (Hovatta et al. 1998; Roberts et al. 1999). Also methodological differences across the studies may affect the variability of the results. The factors potentially improving the probability of replication include increased sample size, uniform study designs and focus on samples drawn from only one ethnic group (Altmuller et al. 2001; Göring et al. 2001). Based on the above-mentioned provisions, it is also clear that linkage cannot be reliably disproved by failure to replicate the original finding in other samples (Hovatta et al. 1998; Roberts et al. 1999).

Even relatively large studies are likely to have a low power to map genes of modest effect by linkage analysis (Terwilliger and Göring 2000). It is commonly noted that the most powerful way to increase statistical power is to analyse original genotype data from several independent studies combined (Lander and Kruglyak 1995; Wise et al. 1999). Data sharing is not always possible in practice and that is why meta-analysis methodologies have been developed for genetic linkage analyses. Currently, the most widely used meta-analysis approach is the Genome Search Meta-Analysis (GSMA) method (Wise et al. 1999), but some other methods also exist (Dempfle and Loesgen 2004). The GSMA method assesses evidence for linkage within specified chromosomal regions by splitting the chromosomes into bins of approximately equal length. These bins are ranked based on linkage evidence within each scan, the ranks for each bin is summed across genome-wide scans, and finally the summed rank is compared to its

probability distribution under null hypothesis. The advantage of this approach is that it enables comparison of genome-wide scans, which have used different family structures, marker sets, and statistical analysis methods (Wise et al. 1999).

### *2.2.6 Association mapping*

Association mapping can be divided into direct and indirect methods. In direct association analysis, the variant itself is thought to predispose to the disease. In practice, this method is confined only to analyses of known functional variants in candidate genes. The indirect method, in contrast, employs linkage disequilibrium (LD), which is a phenomenon of non-random association of alleles at adjacent loci (Collins et al. 1997; Kruglyak 1999). The fundamental idea behind the LD mapping is that the affected individuals in the population have inherited the same ancestral predisposing allele identical-by-descent (IBD). When a new allele is incorporated into a population and is transmitted to the next generation, the haplotype of polymorphisms are brought along with it. LD around newly generated allele decays through recombination over time, but the ancient genetic architecture may still be detectable after several generations (Terwilliger and Göring 2000). Some of the early examples introducing LD mapping to medical genetics include studies of diastrophic dysplasia (DTD; MIM 222600) (Hastbacka et al. 1992) and infantile-onset spinocerebellar ataxia (IOSCA; MIM 271245), where only four affected individuals from two consanguineous pedigrees were sufficient for the initial disease gene localisation (Nikali et al. 1995).

The rationale for the use of association-based methods as a primary tool in mapping of complex diseases comes from the recognition that linkage-based methods have a low power to map susceptibility genes with low genotypic relative risk. In such cases, association analyses are more powerful than linkage-based methods, at least in theory (Risch and Merikangas 1996). In addition, linkage-based methods are not suitable for fine-scale mapping. Since the direct association method is confined only to some special cases, the crucial question is whether susceptibility loci could be detected through LD

between the disease variant and a nearby marker (Collins et al. 1997). Based on the common disease/common variant (CD/CV) hypothesis, genetic susceptibility for common diseases is often influenced by relatively common disease predisposing alleles. In cases consistent with CD/CV hypothesis, common alleles have a modest impact on individuals, but a strong effect on human population (Lander 1996; Chakravarti 1999; Reich and Lander 2001). This assumption is the prerequisite for success in LD-based association mapping of common disorders.

The currently available empirical evidence is not sufficient to estimate how commonly the CD/CV hypothesis holds true. There are now several examples of variants conferring susceptibility for a common disease, but most of these are rare subtypes, which have modest impact at the population level (Risch 2000). There are, however, also some known cases, in which common variant predisposes to a common phenotype (Lohmueller et al. 2003). Based on theoretical models, it seems that in general the loci with a strong effect would have higher mutation rates and lower population frequencies, whereas susceptibility genes with modest effect would be more common and have lower mutation rates (Pritchard 2001; Reich and Lander 2001; Pritchard and Cox 2002). So far, however, LD mapping has been most successful in identification and restriction of the loci for rare recessive diseases in isolated populations, which have small number of founders and have undergone a rapid expansion (Peltonen et al. 1999).

### *2.2.7 Structure of LD in the human genome*

Considerable attention has been focused on to what extent LD is observed in the human genome. The extent of LD relates to the question, how dense marker maps are needed for the association analysis. It is important to minimise the number of markers used while maximising information content due to limited genotyping capacity and the fact that multiple statistical tests need corrected p-values. Several measures exist for quantifying the LD, of which the most commonly used are D, absolute value of D', and $r^2$ (Ardlie et al. 2002; Weiss and Clark 2002).

The empirical evaluations of the extent of LD in the human genome have involved a wide range of strategies including different populations, chromosomal regions, marker types, and marker densities, as well as measures for reporting LD, which makes the comparison of the results from these efforts difficult. Generally, the empirical studies have concluded that LD is extremely variable in several respects, differing markedly across the genomic regions and populations (Ardlie et al. 2002). The average extent of useful LD seems to extend 10-30 kb in European populations and markedly less in the African populations (Reich et al. 2001; Ardlie et al. 2002; Shifman et al. 2003). In rapidly expanded isolated populations, e.g. in the Finnish population, haplotypes extending up to 13 cM have been reported around rare disease alleles (Peltonen et al. 1999). However, the alleles with high frequency (>5%) may not express significantly higher LD in recently expanded isolated populations than in more mixed populations (Kruglyak 1999; Eaves et al. 2000; Taillon-Miller et al. 2000). On the other hand, small constant sized populations (Laan and Paabo 1997; Kaessmann et al. 2002) and small sub-isolates (Varilo et al. 2000; Varilo et al. 2003) seem to express elevated levels of LD. Thus, they could be better suited for the initial coarse mapping of diseases caused by common variants than the traditional isolated populations, given that the population size allows for the collection of a sufficient number of samples for the statistical analyses (Kaessmann et al. 2002; Varilo et al. 2003). Consistently with this idea, Reich and colleagues (2001) suggested that fine-scale mapping should be performed in African populations, because they show significant LD only over very short intervals and therefore could better refine the critical region. Therefore, gaining access to multiple populations with divergent demographic histories would theoretically be an optimal setting for LD mapping studies, but such an approach obviously has some practical limitations.

There has recently been a shift from quantifying usable levels of LD to studying the structure of LD. Despite the apparent complexity of observed LD patterns, recent studies have proposed that the underlying structure of LD in the human genome can be modelled

as a series of discrete haplotype blocks, which are flanked by regions of recombination hot spots (Daly et al. 2001; Jeffreys et al. 2001; Patil et al. 2001; Dawson et al. 2002; Gabriel et al. 2002; Phillips et al. 2003). Gabriel and colleagues (2002) identified haplotype blocks with an average block size of 11 and 22 kb in African and non-African samples, respectively. It is noteworthy that only a few haplotypes per block were generally observed, and a large portion of the haplotypes seemed to be shared across the populations (Gabriel et al. 2002). Even if some regions in the genome fit well with the haplotype block model, LD seems to be more complicated in large portions of the genome. All currently available studies have reported some large blocks extending >100 kb, but also regions of no detectable LD over small distances (<1kb) (Daly et al. 2001; Patil et al. 2001; Dawson et al. 2002; Gabriel et al. 2002; Phillips et al. 2003).

Some of the basic assumptions in the haplotype block model are still unclear. Recent studies support the fact that recombination is often clustered on hot spots, but it is yet unknown to what extent recombination is concentrated in that way (Arnheim et al. 2003; Zhang et al. 2003). It is not clear to what extent other factors than recombination, such as gene conversion or genetic drift, affect the LD block structure (Ardlie et al. 2001; Weiss and Clark 2002; Zhang et al. 2003). If the recombination hot spot model is the predominant mechanism for block formation, the block boundaries should be strongly correlated across different populations. Otherwise, the block structure has to be empirically determined for each population (Wang et al. 2002; Zhang et al. 2003). Furthermore, the use of only common SNPs, sparse marker maps and small sample sizes may have led to overestimation of block lengths in the empirical haplotype block analyses and it has been argued that only a few of the currently available studies satisfy the criteria for the optimal study design (Wang et al. 2002; Zhang et al. 2003; van den Oord and Neale 2004).

The haplotype-block model may have important implications for large-scale association studies, since it indicates a simple rationale for how to select SNP markers. The main haplotypes could be labelled with a small number of haplotype-tagging SNPs (htSNPs)

sufficient to distinguish between the main haplotypes within a block. Thus, most of the variation in the genome could be traced with a limited number of markers (Daly et al. 2001; Patil et al. 2001; Dawson et al. 2002; Gabriel et al. 2002; Phillips et al. 2003). The United States National Human Genome Research Institute (NHGRI) and Wellcome Trust have initiated an effort, called the International HapMap project, which aims to create a genome-wide map of haplotype-blocks in several populations. This effort is hoped to facilitate the positional cloning of common variants by genome-wide association studies (The International HapMap Consortium 2003; Hirschhorn and Daly 2005). The success of the HapMap tool is highly dependent on the CD/CV hypothesis, and it is probably not a powerful tool for mapping rare and geographically localised variants (Weiss and Clark 2002; van den Oord and Neale 2004). Yet, even if novel high-throughput technologies allow genotyping of massive numbers of SNP markers per sample (Syvänen 2005; Hirschhorn and Daly 2005), huge sample sizes would be needed to reach statistical significance in a genome-wide association study due to multiple statistical testing. Various estimates of the amount of SNPs needed to extract information of genome have been put forward and they reach from 300 000 to 3 million or more. The HapMap concept involves several unanswered questions, and its relevance for mapping of common disease genes, at least at the genome-wide scale, remains elusive.

### *2.2.8 Study designs in association mapping*

The case-control strategy, which compares the allele-frequencies between the cases and controls, has traditionally been the most widely applied approach in association analyses (Figure 4). The advantage of this strategy is that such studies are relatively easy to set up. On the other hand, case-control studies are prone to false-positive associations caused by population stratification and systematic errors in control sampling. The careful matching of cases and controls is crucial, since any differences in allele-frequencies lead to positive associations. Control ascertainment can be improved by using a prospective cohort study, where a large number of individuals are collected prior onset of the

disease, and the association between cases and controls is tested after the age of disease onset. In this way, bias in control population selection is avoided. However, such study design requires much more resources, larger sample sizes, and longer time than the traditional case-control design (Cardon and Bell 2001; Ott 2004). Alternatively, population stratification may be assessed by using data from a series of unlinked markers. A method entitled genomic control compares the chi-square statistics for the test data with the chi-square obtained on unlinked markers, which results in corrected chi-square statistics (Pritchard et al. 2000).

Methods using family-based controls have been developed to avoid bias in control ascertainment. In haplotype-relative risk (HRR) method, parental alleles are classified into those transmitted to affected children and those not transmitted. HRR compares allele and genotype frequency between the proband group and a control group, which is constructed from non-transmitted alleles. The relative risk is assessed by chi-square statistics (Falk and Rubinstein 1987; Ott 1999). The transmission disequilibrium test (TDT) is a modification of HRR, which tests linkage in the presence of LD using data consisting of parents and the affected offspring. The frequency of transmitted alleles is compared with non-transmitted parental control alleles, and the statistical significance of the deviation is tested (Spielman et al. 1993; Spielman et al. 1994; Spielman and Ewens 1996). Various extensions of TDT are currently available to overcome the problems caused by missing or homozygous parental data, eliminate the effect of linkage when multiple siblings are present in a family, and to include quantitative traits in the analyses (Clayton 1999; Abecasis et al. 2000; Göring and Terwilliger 2000; Horvath et al. 2001). There are also methods to combine information from several adjacent SNP markers to construct haplotypes, which can be used as multiallelic markers in the family-based association analyses (Clayton 1999; Horvath et al. 2001). The major drawback in all the family-based association methods is clearly that parental samples are not always available, especially in late-onset diseases. It has also been suggested that TDT-based tests are more susceptible to false-positive results due to technical artefacts than case-control studies (Hirschhorn and Daly 2005).

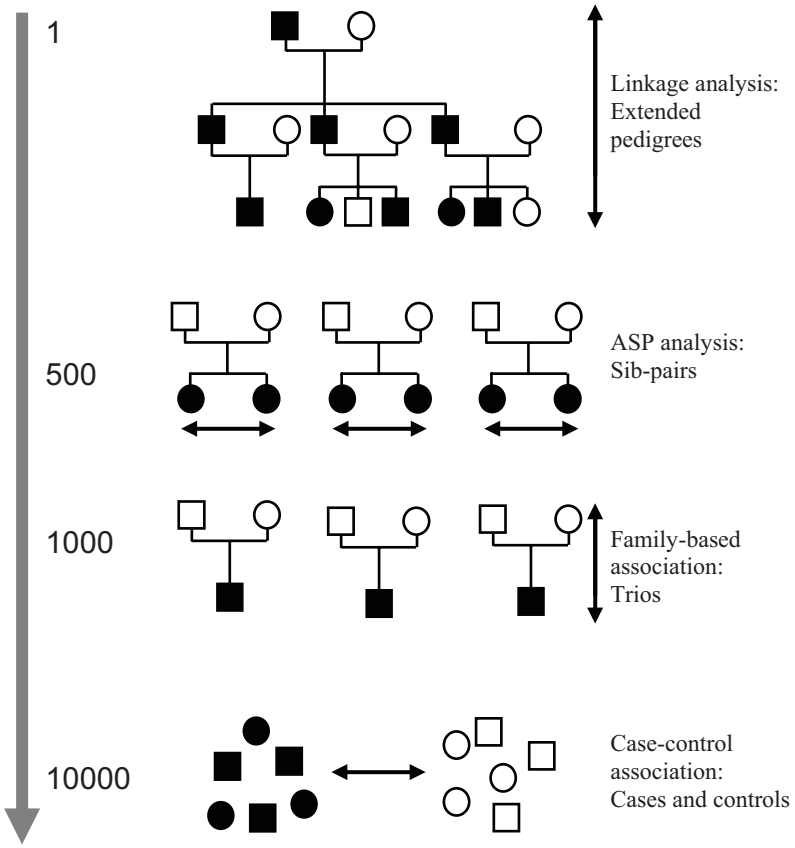


**Figure 4. Comparison of different study designs in genetic mapping.** Linkage analysis tests correlation of marker alleles and phenotype within families. Association analyses, by contrast, test correlation of marker alleles and phenotype between the families.

### *2.2.9 Validity of genetic associations*

Promising associations have recently emerged for common complex diseases, but the inability to replicate the original findings is still the predominant phenomenon in the genetic association studies. In principle, the inconsistency may be due to false negative studies, false positive studies, or true variability in association among different

populations (Lohmueller et al. 2003). Some of the difficulties in the interpretation of the reported associations are listed below.

Association analyses commonly consider candidate genes, which can be divided into positional and hypothesis-driven candidates. Positional candidates are selected because they are located in a genomic region that has been identified by mapping studies. Hypothesis-driven candidates, by contrast, are selected because they fit into a proposed model of the physiology of the phenotype (Glatt and Freimer 2002). The findings from these two types of studies should be considered somewhat differently. The hypothesis-driven candidates have been selected based on a limited knowledge of relevant pathways for the disease pathogenesis, and thus the vast majority of functional candidates are excluded from these analyses. Therefore, the prior probability is in many cases extremely low, which implies that highly convincing evidence is needed to reject the null hypothesis. For positional candidates, prior probability is higher due to advance evidence from mapping studies (Owen et al. 1997; Glatt and Freimer 2002; Ott 2004).

Many of the problems in association studies relate to the sample-size. Studies with small or moderate samples are prone to random error and often provide wide confidence intervals for any significant conclusions. Generally, the genetic determinants are thought to be of modest effect, which means that large samples are required for convincing associations (Owen et al. 1997; Cardon and Bell 2001; Glatt and Freimer 2002). Furthermore, as in linkage analysis, it has been shown that the estimate of the genetic effect in the first positive report is usually biased upward complicating the replication attempts in the samples that are of similar size as in the original study (Göring et al. 2001; Lohmueller et al. 2003). Meta-analyses have become popular methods also in genetic association studies to increase statistical power and estimate significance levels across published studies (Munafo and Flint 2004).

The significance level is often difficult to specify in the presence of multiple testing and low prior probability. It is obvious, however, that more stringent significance level criteria than the conventional $p<0.05$ are needed for convincing associations (Ott 2004).

Typically, all the statistical tests performed are not reported or the reported associations are not corrected for multiple testing. This inevitably leads to overestimation of the significance (Owen et al. 1997; Ott 2004). The problem is multiplied by the recent emergence of haplotype-based association methods. Usually numerous combinations of tightly linked SNPs are tested, each with a possibility of observing type I error. On the other hand, traditional statistical corrections, such as Bonferroni correction, may be overly conservative when highly correlated markers are tested (Cardon and Bell 2001; Hennah et al. 2004). Production of empirical p-values overcomes some of the multiple testing problems, but the results are dependent upon the arbitrarily chosen number of permutations to be carried out (Hennah et al. 2004). It is also evident that negative results from the replication studies are not published as frequently as positive findings. This leads to publication bias, meaning that significant p-values tend to be over-represented in the literature (Terwilliger and Weiss 1998; Ott 2004).

Consideration of the haplotypes or LD across a gene of interest has started only recently. Failing to do so will likely produce type II error or result in inconsistent results across the studies (Ott 2004). Effect of LD on the sample size required to detect association is exemplified in Figure 5.

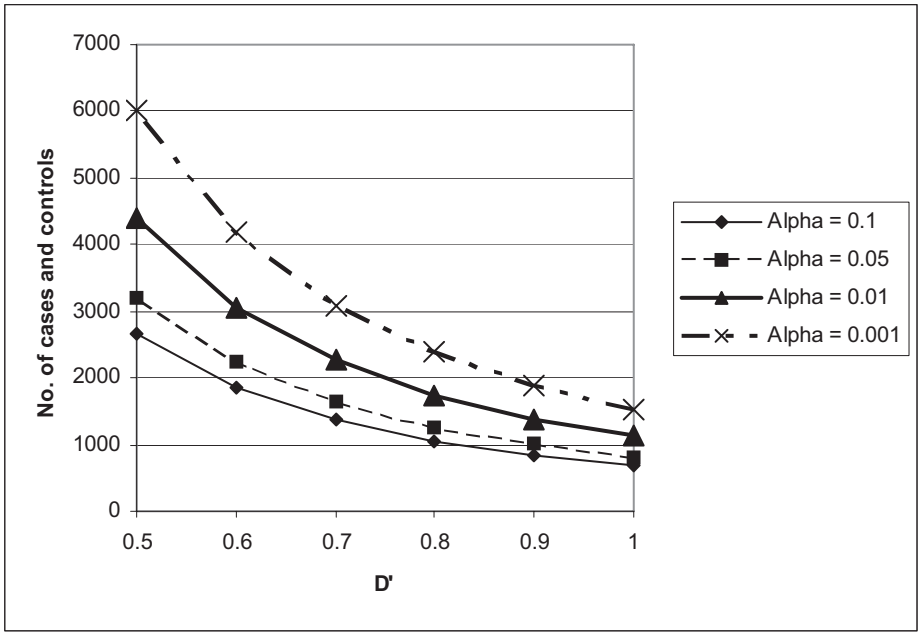


**Figure 5. Effect of LD on the sample size required in association mapping.** D' values ranging from 0.5 to 1.0 are show on the X-axis and number of cases and controls required to detect association at significance levels of 0.1, 0.05, 0.01 and 0.001 (95% power) are shown on the Y-axis. The arbitrarily chosen parameters are as follows: disease-allele frequency = 0.3, minor allele frequency of the marker = 0.3, population prevalence = 0.01, genotype relative risk = 1.5. Simulations were performed using the Genetic Power Calculator (Purcell et al. 2003).

Probably the only way either to confirm or disprove the reported associations is to perform several independent replication studies, each analysing a full spectrum of allelic variants within the associated region. The accumulated information, along with functional information, should eventually reveal whether the associations are valid or not (Hennah et al. 2004; Ott 2004).

### *2.2.10 Value of population isolates in the complex disease mapping*

Isolated populations, such as the Finnish population, have been extremely useful for mapping rare Mendelian diseases. By definition, these populations are founded by a relatively small number of individuals and remained isolated due to e.g. cultural and/or geographical factors. Several factors, such as a high degree of inbreeding, rapid expansion, population bottlenecks and genetic drift, have modified the genetic background leading to the reduced genetic diversity and enrichment of some alleles (Peltonen et al. 2000). In the Finnish population this is exemplified by enrichment of a special set of monogenic disorders caused by one or few ancestral mutations (the so called Finnish disease heritage) and absence of some diseases, which are common elsewhere. When the vast majority of cases are caused by the same ancestral mutation, LD is observed in significant intervals, thereby greatly facilitating the disease gene identification by haplotype-based methods (Peltonen et al. 1999; Norio 2003a; Norio 2003b).

For common alleles, the situation is less straightforward. Even if the number of founders is relatively small, common alleles will enter the founder population so many times that the length of shared haplotypes around these alleles will be indistinguishable from that of the much larger ancestral population. This is why significantly elevated levels of LD have not generally been detected around common alleles in the isolated populations (Kruglyak 1999; Eaves et al. 2000; Taillon-Miller et al. 2000; Hirschhorn and Daly 2005). However, population isolates offer some other advantages, which might prove important in the genetic mapping of many common diseases. First, the people in many isolates share a common environment and culture, which are potentially important factors in the development of many complex diseases. Second, it is easier to standardize diagnostic criteria and thus markedly decrease the phenotypic heterogeneity. In some isolated populations, clinical information is systematically collected into the health-care records, which are usually available for research purposes. Third, it is in some cases possible to identify small subisolates or reconstruct large pedigrees based on the

population registries. This approach should greatly increase the genetic homogeneity of the sample. Finally, the benefits of the reduced genetic heterogeneity in isolates will certainly hold in the cases where rare alleles confer significant increase in susceptibility to complex diseases (Peltonen et al. 2000).

## 2.3 Genetics of autism spectrum disorders

### *2.3.1 Establishing the genetic component in Autism Spectrum Disorders*

In a family study by Bolton and colleagues (1994) 2.9% of autistic probands' siblings had autism and a further 2.9% had features of the more broadly defined PDD (atypical autism or AS). Therefore, assuming a prevalence rate of 2/1000 the sibling risk compared with the risk in the general population ($\lambda_s$) would be around 15-30 for autism indicating strong familial clustering of the disorder. Twin studies have suggested that familial clustering is mainly due to genetic factors. Bailey and colleagues (1995) re-analysed the twin pairs originally included in the two studies by Folstein and Rutter (Folstein and Rutter 1977b; Folstein and Rutter 1977a) and expanded the sample to include 25 MZ and 20 DZ twins. A strong genetic component in autism was established with 60% MZ concordance and 0% DZ concordance for narrow phenotypic definition for autism. The concordance rate of 0% for DZ-twins is likely to be an underestimate due to the small sample size used, and the true DZ-twin concordance is probably similar to the nontwin-sib rate. The corresponding figures for MZ and DZ twins with broad spectrum of cognitive or social abnormalities was 90% and 10%, respectively. (Bailey et al. 1995). Even higher MZ concordance rates were reported in a study by Steffenburg and colleagues (1989). In their sample collected from Nordic countries, 91% (10/11) of MZ twins and 0% (0/10) of DZ twins were concordant for autism. Also, several other twin studies have been published in autism but they have generally failed to include sufficient number of cases or they have suffered from inappropriate methodological procedures (see, e.g. Ritvo et al. 1985).

Several studies have provided evidence that a complex mode of inheritance with multiple interacting genes and environmental factors is most likely for autism. The first evidence for this is a rapid cut-off in MZ and DZ twin concordance rates (Bailey et al. 1995) combined with further cut-off in both autism and broader phenotype going from first-degree relatives to second-degree relatives (Jorde et al. 1990). Furthermore, the less than 100% risk for MZ twins indicates that also other than genetic factors play a role (Bailey et al. 1995). The segregation analyses performed in autism have rejected a single locus model and supported the fact that a polygenic model is most likely (Jorde et al. 1991; Pickles et al. 1995). Pickles and associates (1995) concluded that a three-gene epistatic model is most likely, even if the range could be between two and ten. In their report of a genome-wide scan in autism Risch and colleagues (1999) proposed that more than 15 loci might contribute to the aetiology of autism.

Evidence concerning familial aggregation and genetic contribution to AS is still scarce. However, several case reports, including the original report by Asperger (1944), have noted that AS-like problems in social interactions are frequently present in AS individuals' family members (Asperger 1944; Burgoine and Wing 1983; Bowman 1988; Gillberg 1989; Volkmar et al. 1996; Volkmar et al. 1998; Folstein and Santangelo 2000). The preliminary results from the Yale Asperger Project with 99 families indicate that there was a positive family history of AS or similar condition in 46% of the families (Volkmar et al. 1998). Gillberg (1989) described 23 AS families, in which 28% of the parents fulfilled the criteria for a broad autism phenotype or AS. Thus, the preliminary evidence available suggests that familial aggregation of AS might be even stronger than that of autism (Gillberg 1989; Volkmar et al. 1998). Furthermore, autism and AS commonly exist in the same families suggesting that at least in some cases common aetiological factors might predispose to these syndromes (Gillberg 1989; Volkmar et al. 1998; Folstein and Santangelo 2000).

### *2.3.2 Known genetic aetiologies associated with autism*

Autism is over-represented as part of the behavioural phenotype in several syndromes of genetic origin (Table 3). Some of these monogenic causes may lead to a phenotype, which is indistinguishable from the idiopathic cases, whereas in some other cases the autistic-like features are only a one component of the phenotype. Known genetic syndromes together with chromosomal abnormalities constitute approximately 10-15% of autism cases (Bailey et al. 1996; Chakrabarti and Fombonne 2001; Folstein and Rosen-Sheidley 2001). However, the relevance of these genes to the understanding of the pathogenesis of idiopathic autism is yet uncertain.

*Fragile X syndrome (FRAXA).* The phenotype of fragile X syndrome (MIM 309550) includes moderate to severe mental retardation, macroorchidism, large ears, a prominent jaw, and high-pitched jocular speech. Males with FRAXA have also numerous autistic-like features (Feinstein and Reiss 1998; Sabaratnam et al. 2003). Fragile X syndrome is caused by mutations in fragile X mental retardation 1 gene (*FMR1*) that result from a CGG trinucleotide expansion in the 5' untranslated region and lead to DNA methylation and lack of transcription (Verkerk et al. 1991; Feinstein and Reiss 1998). *FMR1* gene encodes an RNA-binding protein FMRP (fragile X mental retardation protein), which influences synaptic plasticity through regulation of mRNA transport and local protein synthesis at synapses using micro RNA (miRNA) pathway as the main mechanism (Jin et al. 2004). Varying frequencies of FRAXA have been reported in autism. Earlier studies reported only a little association, whereas more recent studies have shown frequencies ranging from 2 to 8% (Bailey et al. 1996; Feinstein and Reiss 1998; Wassink et al. 2001; Muhle et al. 2004). ASDs are relatively common in the group of fragile X patients with up to 30% of the patients with FRAXA fulfilling the criteria for ASD (Feinstein and Reiss 1998; Rogers et al. 2001; Muhle et al. 2004).

*Tuberous sclerosis complex (TSC).* The tuberous sclerosis complex (MIM 191100; 191092; 605284) is a dominantly inherited disease characterized by the presence of

hamartomata in multiple organ systems. The clinical features include epilepsy, learning difficulties, behavioural problems, as well as skin and renal lesions. It has been estimated that as many as from 25 to 50% of patients with TSC fulfil the criteria for autism, whereas the prevalence of TSC in autism has been reported as 1 – 4% (Smalley 1998; Curatolo et al. 2004; Wiznitzer 2004). Mutations in two different tumour suppressor genes, *TSC1* encoding hamartin protein and *TSC2* encoding tuberin protein, may lead to TSC (The European Chromosome 16 Tuberous Sclerosis Consortium 1993; van Slegtenhorst et al. 1997). These proteins form a TSC1-TSC2 complex, which is involved in regulation of cell growth and proliferation (Pan et al. 2004). It is yet unclear whether the autistic features in TSC are caused by the presence of a non-specific disruption of brain function related to complications of TSC, such as seizures, MR, and presence of tubers in the temporal lobes, or whether the autism is caused by a distinct molecular mechanism involving TSC1-TSC2 complex (Wiznitzer 2004).

*Neurofibromatosis (NF1).* Type I neurofibromatosis (MIM 162200) is an autosomal dominant disorder characterized by cafe-au-lait spots and fibromatous tumours of the skin. *NF1* gene mutated in type 1 neurofibromatosis encodes a tumour suppressor called neurofibrin, which functions as a negative regulator of Ras, a fundamental protein controlling cell proliferation. Impaired neurofibrin function leads to elevated Ras activity and, consequently, increased cell proliferation (Cawthon et al. 1990; Dasgupta and Gutmann 2003). Neurofibromatosis appears to overlap with small fraction of autism cases, but greatly varying frequencies have been reported due to small sample sizes in the studies. However, the frequency of neurofibromatosis in autism seems to be much lower than those of FRAXA or tuberous sclerosis (Williams and Hersh 1998).

*NLGN4 and NLGN3 mutations.* Two X-chromosomal members of the neuroligin gene family, *NLGN3* and *NLGN4*, have been shown to be mutated in some patients with ASDs. Two affected males with autism and AS in a Swedish family had an insertion in *NLGN4*, which led to premature truncation of the protein (D396X). In another Swedish family with autism and AS, an R451C substitution in *NLGN3* was identified (Jamain et

al. 2003). In addition, a truncating *NLGN4* mutation (D429X) was later reported in a large family, where mental retardation with or without ASDs segregated X-chromosomally (Laumonnier et al. 2004).

Neuroligins were originally isolated as splice site-specific ligands of β-neurexins in the rat brain (Ichtchenko et al. 1995). The neuroligin gene family consists of five genes in humans: neuroligin-1 (*NLGN1*) located at 3q26, neuroligin-2 (*NLGN2*) located at 17p13, neuroligin-3 (*NLGN3*) located at Xq13, neuroligin-4 (*NLGN4*) located at Xp22, and neuroligin-4Y (*NLGN4Y*) located at Yq11. Neuroligin genes encode 816–863 amino acid transmembrane proteins which are composed of five distinct domains: an N-terminal cleaved signal peptide, a large extracellular domain homologous to acetylcholinesterases (AchE), a short linker domain, transmembrane domain, and a cytosolic region (Ichtchenko et al. 1995; Ichtchenko et al. 1996; Bolliger et al. 2001). The extracellular esterase-domain lacks an active site serine suggesting that it is not catalytically active (Ichtchenko et al. 1995).

The rodent neuroligins 1, 2, and 3 have been shown to be expressed at high levels only in the brain (Ichtchenko et al. 1995; Ichtchenko et al. 1996; Song et al. 1999). Similarly, human neuroligins are expressed in the brain with no considerable differences in the regional distribution (Song et al. 1999; Jamain et al. 2003). However, at least human *NLGN4* has been shown to be expressed also in heart, liver, skeletal muscle, and pancreas (Bolliger et al. 2001). Neuroligin-1 expression has been shown to be low in embryonic brains but increase dramatically after birth. The expression was again reduced during the period when most synapses are already formed, but some level of expression seems to persist throughout adulthood (Ichtchenko et al. 1996; Song et al. 1999; Scheiffele et al. 2000). It has been suggested that such an expression pattern might reflect the involvement of neuroligins in the synaptic remodelling processes or in the regulation of the size and strength of synaptic connections by inducing new presynaptic specializations or by stabilizing pre-existing structures (Scheiffele et al. 2000).

Neuroligins have several well-characterized interactions. As noted above, at the presynaptic side neuroligins bind to β-neurexins in a $Ca^{2+}$-dependent manner (Ichtchenko et al. 1995; Comoletti et al. 2003). At the postsynaptic density, neuroligins interact with PDZ-domains (PSD-95-*Drosophila* discs-large tumour suppressor protein-zona occludens-1) of several proteins, including PSD-95 (post synaptic density 95) and S-SCAM (synaptic scaffolding molecule) (Irie et al. 1997; Hirao et al. 1998; Meyer et al. 2004). However, the PSD-95 or β-neurexin interactions are not needed for the targeting of the neuroligins to synapses, but instead, neuroligins have an intracellular domain that is essential for postsynaptic targeting (Dresbach et al. 2004).

Neuroligins seem to be capable of triggering the *de novo* formation of presynaptic structures at glutamatergic synapses. Scheiffele and colleagues (2000) observed development of several presynaptic structures when CNS neurons were co-cultured with non-neuronal cells expressing neuroligins *in vitro*. These changes involved clustering of synaptic vesicles within axons and several functional and morphological changes typical for neuron-neuron synapses. Such changes are not typically observed when neurons are contacted with non-neuronal cells. The neuroligin activity seemed to require specific sequences in the extracellular, acetylcholinesterase-homologous domain of the protein. In addition, the neuroligin-activity was inhibited by adding soluble β-neurexin indicating that β-neurexins were primary candidates for mediating the neuroligin functions (Scheiffele et al. 2000). It was later shown that neuroligins form oligomers and these oligomers recruit neurexins to newly forming synaptic sites. A two-step model of neuroligin function was proposed, in which postsynaptic neuroligin oligomers initially cluster axonal neurexins. In response to this clustering, neurexins nucleate the assembly of a cytoplasmic scaffold to which the exocytotic apparatus is recruited (Dean et al. 2003). Neurexins are directly coupled to synaptotagmins, which are core molecules of synaptic vesicle machinery that regulate neurotransmitter exocytosis (Hata et al. 1993). Therefore, neuroligin - β-neurexin interaction may activate an array of presynaptic molecular responses, leading to structural re-organisation of the presynaptic compartment (Prange et al. 2004).

There is now evidence that neuroligins might also be involved in mediating the synapse formation at the inhibitory synapses. Neuroligins-1, -3, and –4 are localized exclusively at the postsynaptic side of the excitatory synapses (Song et al. 1999; Graf et al. 2004). Neuroligin-2, by contrast, has been shown to localize also at the inhibitory synapses (Graf et al. 2004; Varoqueaux et al. 2004; Chih et al. 2005). It has been shown that mislocalized expression of neuroligin-2 disperses the postsynaptic proteins and disrupts synaptic transmission at both inhibitory and excitatory synapses (Graf et al. 2004). Furthermore, neuroligin over-expression seems to increase both excitatory and inhibitory synaptic contacts (Prange et al. 2004; Chih et al. 2005). Finally, knockdown of neuroligin function by the RNA interference (RNAi) method reduces the number of both excitatory and inhibitory synapses. The effect concerning inhibitory synapses are more pronounced when the function of neuroligin-2 is suppressed (Chih et al. 2005). Thus, it seems that neuroligin - β-neurexin interaction is a major component in synapse formation mediating both GABAergic and glutamatergic synaptogenesis (Graf et al. 2004; Prange et al. 2004). Furthermore, based on RNAi studies and the neuroligin knock-out mouse (for which the phenotype has not been published), it seems that different neuroligin isoforms are able to compensate for each other's functions at least to some extent (Song et al. 1999; Chih et al. 2005).

Multiple lines of evidence indicate that the reported *NLGN3* and *NLGN4* mutations have deleterious effects at the molecular level. First, both of the truncating mutations reported by Jamain and co-workers (Jamain et al. 2003) and Laumonnier and co-workers (Laumonnier et al. 2004) involve deletion of the AchE-homologous domain, which is required for oligomerization and the synapse promoting activity of neuroligins (Dean et al. 2003). Second, Comoletti and colleagues (2004) reported markedly diminished β-neurexin binding activity of the R451C mutation in *NLGN3* compared to the wild-type protein. Third, both R451C mutation in *NLGN3* and D396X mutation in *NLGN4* lead to retention of the protein in the endoplasmic reticulum, and cell surface levels of the mutants are significantly lower than the levels of the wild-type protein. Finally, neither

of these mutants seem to promote presynaptic differentiation (Chih et al. 2004; Comoletti et al. 2004). Although the functional studies support that *NLGN3* and *NLGN4* mutations may be causative for autism, they occur at a low frequency. The two original mutations were identified among 158 samples from independent families with autism or AS (Jamain et al. 2003). Subsequently, Vincent and colleagues (2004) and Gauthier and colleagues (2004) performed mutation analysis of *NLGN3/NLGN4* genes by using samples of 196 and 96 autistic probands, respectively. No causative mutations were identified in either of the samples. Thus, the currently reported frequency of either *NLGN3* or *NLGN4* mutations is around 0.4%.

*Other syndromes associated with autism.* Several other rare syndromes having autistic behavioural pattern have been reported (Table 3). Phenylketonuria (PKU; MIM 261600) is the earliest reported association between a Mendelian condition and autism but it seems to occur in a very low frequency (Folstein and Rosen-Sheidley 2001; Baieli et al. 2003). Currently, the association is probably non-existent due to effective treatment of the PKU (Folstein and Rosen-Sheidley 2001). In addition, the phenotypes of e.g. Smith-Lemli-Opitz syndrome (SLOS; MIM 270400), Cornelia de Lange syndrome (CDLS; MIM 122470), Angelman syndrome (MIM 105830), Prader-Willi syndrome (PWS; MIM 176270), and Timothy syndrome (TS; MIM 601005) are associated with autistic behaviour (Steffenburg et al. 1996; Berney et al. 1999; Tierney et al. 2000; Splawski et al. 2004; Veltman et al. 2004). Furthermore, some female cases with atypical autism phenotype have been reported to have *MECP2* or *CDKL5/STK9* mutations originally associated with Rett syndrome (Carney et al. 2003; Weaving et al. 2004).

**Table 3. Syndromes and single gene mutations associated with autistic behavior.** Note that for most aetiologies it is difficult to give reliable frequency estimates due to the rarity of the overlap.

| Etiology | MIM | Characteristic phenotype | Defective gene | Chromosomal location | Mode of inheritance | Frequency in autism |
|---|---|---|---|---|---|---|
| Angelman syndrome | 105830 | MR, ataxia, hypotonia, characteristic facial features, Epilepsy, absence of speech, predominant smiling and laughter | *UBE3A* | 15q11.2 | Mostly sporadic | - |
| Cornelia de Lange Syndrome | 122470 | Characteristic facies in association with prenatal and postnatal growth retardation, MR, upper limb anomalies | *NIPBL* | 5p13.2 | Mostly sporadic | - |
| Fragile X syndrome | 30950 | Moderate to severe MR, macroorchidism, large ears, prominent jaw, high-pitched jocular speech | *FMR1* | Xq27.3 | X-chromosomal | 2-8% |
| Neurofibromatosis (Type I) | 162200 | Cafe-au-lait spots and fibromatous tumours of the skin, abnormalities of the skeleton and CNS, learning difficulties | *NF1* | 17q11.2 | Autosomal dominant | 0.3-6% |
| *NLGN3*/*NLGN4* mutations | 300425<br>300495 | Autism, AS, non-specific MR | *NLGN3*<br>*NLGN4* | Xq13.1<br>Xp22.3 | X-chromosomal | Few cases reported |
| Phenylketonuria | 261600 | MR, characteristic odour, light pigmentation, peculiarities of gait, stance, and sitting posture, eczema, epilepsy | *PAH* | 12q23.2 | Autosomal recessive | - |
| Prader-Willi syndrome | 176270 | MR, obesity, hypogonadism, hypotonia, short stature, small hands and feet, hypopigmentation, characteristic facial features | unknown | 15q11-13 | Mostly sporadic | - |

*Table 3 continued.*

| Etiology | MIM | Characteristic phenotype | Defective gene | Chromosomal location | Mode of inheritance | Frequency in autism |
|---|---|---|---|---|---|---|
| Rett syndrome/ Rett variants | 312750 | Loss of acquired skills between ages 6 and 30 months, communication dysfunction, social withdrawal, deceleration of head growth, stereotypic hand movements, gait appraxia/ ataxia | *MECP2* *CDKL5* | Xq28 Xp22.13 | X-chromosomal | Few atypical cases reported |
| Smith-Lemli-Opitz syndrome | 270400 | MR; dysmorphic features, occasionally malformations in the brain, lung, heart, and gastrointestinal tract | *DHCR7* | 11q13.4 | Autosomal recessive | - |
| Timothy syndrome | 601005 | Arrhythmia, congenital heart disease, syndactyly, immune deficiency, hypoglycemia, cognitive abnormalities, autism | *CACNA1C* | 12p13.3 | Sporadic | Few cases reported |
| Tuberous Sclerosis | 191100 191092 605284 | Presence of hamartomata, epilepsy, learning difficulties, behavioural problems, and skin and renal lesions. | *TSC1* *TSC2* | 9p34.13 16p13.3 | Autosomal dominant | 1-4% |

### *2.3.3 Cytogenetic findings*

The reported estimates of the rate of chromosomal aberrations have ranged from less than 5% to 48%, depending mainly on whether subjects with very low IQ or physical abnormalities are included. Chromosomal aberrations have been reported for most of the chromosomes, the most frequent findings being for the long arm of chromosome 15 and sex chromosomes (Gillberg 1998; Folstein and Rosen-Sheidley 2001).

*Chromosome 15q11-13 aberrations.* Most of the reports of chromosomal abnormalities reported for chromosome 15 have been maternal duplications involving 15q11-q13 -region, which account for 1-3% of patients with autism (Gillberg 1998; Wassink et al. 2001; Veenstra-VanderWeele and Cook 2004). Region 15q11-13 is a critical region for the Prader-Willi and Angelman syndromes, which are both associated with the autistic behavioural pattern (Steffenburg et al. 1996; Veltman et al. 2004). Other characteristic features for these syndromes have been listed in Table 3. Angelman syndrome results most commonly from a *de novo* maternal microdeletion of the chromosome 15q11-13, but a low frequency of Angelman syndrome cases is also caused by paternal uniparental disomy, *UBE3A* mutations, or imprinting mutations, in which the maternal chromosome has the paternal pattern of methylation (Veenstra-VanderWeele and Cook 2004). Prader-Willi syndrome, by contrast, results from a de novo paternal microdeletion of the 15q11-13 region or uniparental disomy for the maternal chromosome 15 (Veltman et al. 2004). It is probable that most of the structural abnormalities observed at 15q11-13 arise from recombinations involving highly homologous regions on both 15q11 (break-points BP1 and BP2, see Figure 6) and 15q13 (BP3), which derive largely from duplications of a *HERC2* (hect domain and RLD 2) gene (Nicholls and Knepper 2001).

Consistently with the parental effect in the inheritance, the 15q11-13 region contains a cluster of genes which are subject to genomic imprinting. The promoter of the small nuclear ribonucleoprotein polypeptide N (*SNRPN*) gene is located within the CpG island that is completely methylated on the maternal chromosome and completely

unmethylated on the paternal chromosome. This region is denoted as the imprinting centre of the 15q11-13 locus (Shemer et al. 2000). The genes located proximally to the imprinting centre are paternally expressed, whereas the genes distally to imprinting centre are maternally expressed. A cluster of non-imprinted $GABA_A$ receptor genes is located further distally to the maternal expression domain, but yet within the autism critical region (Jiang et al. 1999). Angelman syndrome is thought to result from the lack of expression of the maternally expressed *UBE3A* gene in the brain, whereas other deleted genes are considered as having a modifying effect. *UBE3A* encodes a ubiquitin protein ligase, E6-AP, which is involved in the proteasome-mediated protein degradation (Jiang et al. 1999; Nicholls and Knepper 2001). The PWS critical region includes several paternally expressed genes but no single gene mutations causing PWS or a PWS-like disorder have been reported. Thus, there is no single gene, which would be unequivocally linked to the PSW phenotype (Nicholls and Knepper 2001). An overview of the genomic organization of the 15q11-13 locus is shown in Figure 6. Analyses of the 15q11-13 candidates in autism have been reviewed in 2.3.5.

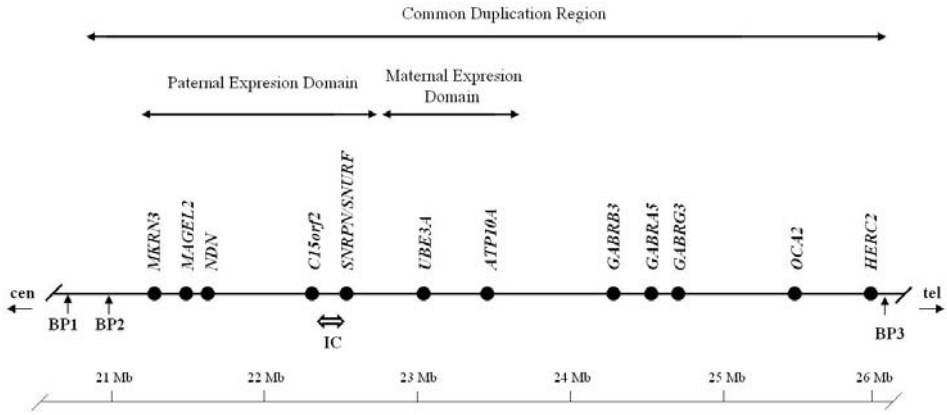


**Figure 6. Critical region for autism, Angelman syndrome and Prader-Willi syndrome on 15q11-13.** Abbreviations are as follows: BP, break-point; IC, imprinting center; *MKRN3*, makorin ring finger protein 3; *MAGEL2*, MAGE-like 2; *NDN*, Necdin; *C15orf2*, chromosome 15 open reading frame 2; *SNRPN*, small nuclear ribonucleoprotein polypeptide N; *SNURF*, SNRPN upstream reading frame protein; *UBE3A*, ubiquitin protein ligase E3A isoform 3; *ATP10A*, ATPase Class V type 10A; *GABRB3*, γ-aminobutyric acid (GABA) A receptor beta; *GABRA5*, GABA A receptor alpha; *GABRG3*, GABA A receptor gamma; *OCA2*, oculocutaneous albinism II; *HERC2*, hect domain and RLD 2. See text for further details. The figure is modified from Veenstra-VanderWeele and Cook (2004) and Nicholls and Knepper (2001) with additional data from the UCSC Human Genome Browser (May 2004 assembly).

*Turner syndrome.* Another chromosome of frequent aberrations in autism is chromosome X, where a wide range of different abnormalities has been observed (Gillberg 1998). An interesting observation was made in girls with Turner syndrome, which affects females who possess only one copy of the X-chromosome. Skuse (2000) has reported that 6.4% of the girls with Turner syndrome who had inherited complete maternal X-chromosome developed autism. By contrast, none of the girls with Turner syndrome and paternally inherited X-chromosome had autism in a total sample of 221 Turner girls. The girls with Turner syndrome who had inherited paternal X-

chromosomes also did significantly better in a test of social-communicative skills than Turner girls with maternally inherited X-chromosomes. Similarly, in the general population girls have superior social-communicative skills compared to boys. Based on these observations, an imprinted-X liability threshold model for autism was proposed. According to this model, genetic vulnerability to autism is due primarily to the effects of the autosomal loci. Females with a normal karyotype have an imprinted protective locus that is expressed from a paternally inherited X-chromosome, and therefore females have a higher threshold for the expression of autistic features. The X-chromosome in males is maternally inherited and, consequently, males lack the protective effect. According to this model, the lack of a paternal X-chromosome would also predispose to poor social-communicative - skills in general (Skuse et al. 1997; Skuse 2000).

*Balanced translocations.* The chromosomal breakpoints caused by balanced translocations may provide a short cut to susceptibility gene identification, as exemplified by the cases of *DISC1/DISC2* (disrupted in schizophrenia 1 and 2) in schizophrenia, *FOXP2* (forkhead box P2) in developmental verbal dyspraxia, and *DYX1C1* (dyslexia susceptibility 1 candidate 1; also known as *EKN1*) in dyslexia (Lai et al. 2000; Millar et al. 2000; Lai et al. 2001; Taipale et al. 2003). There have also been some attempts to identify candidate genes by such an approach in autism.

The long arm of chromosome 7 has been a subject of intense research in autism because of overlapping linkage findings from several genome-wide scans (see, 2.3.4). An autistic individual carrying a translocation between chromosomes 7 and 13, t(7;13)(q31.3;q21), was found in a study by Vincent and colleagues (2000). A novel gene denoted as *RAY1* (later known as suppression of tumorigenicity 7, *ST7*) was found to be interrupted by the translocational breakpoint, but the analysis of 27 unrelated autistic individuals failed to show phenotype-specific variants of this gene. Sultana and colleagues (2002) characterized a novel gene (*AUTS2*) located at 7q11.2, which was disrupted by a translocation in MZ twins with autism. No evidence for autism specific mutations or association was observed in a subsequent analysis of 65 families with autism (Sultana et

al. 2002). Also several other translocations for the 7q region have been reported in the autistic patients (Ashley-Koch et al. 1999; Warburton et al. 2000; Tentler et al. 2001).

The Neurobeachin (*NBEA*) gene was found to be disrupted in an autistic patient with a *de novo* translocation t(5;13)(q12.1;q13.2) (Castermans et al. 2003). There are also three additional autism cases reported to carry a chromosomal abnormality involving the *NBEA* locus on 13q13.3 (Ritvo et al. 1988; Steele et al. 2001; Smith et al. 2002). One of the autism genome-wide scans has identified the *NBEA* locus as a potential susceptibility locus for autism (Barrett et al. 1999).

A balanced t(X;8)(p22.13;q22.1) translocation has been reported to disrupt the gastrin-releasing peptide receptor (*GRPR*) located at Xp22.13 in a female patient with autism (Ishikawa-Brush et al. 1997). Subsequently, Thomas and colleagues (1999) reported three females with Xp deletions involving the GRPR locus. It is of interest that the *NLGN4* gene is located at Xp22.13 in the vicinity of the breakpoint reported by Ishikawa-Brush et al. (1997) and within the deleted region reported by Thomas et al. (1999).

Finally, an apparently balanced t(2;8)(q35;q21.2) translocation, which involved partial deletion of the paired box 3 (*PAX3*) gene at 2q25 and a breakpoint at the matrix metalloproteinase (*MMP16*) gene at 8q21.2, has been reported in a boy with developmental delay and autism (Borg et al. 2002).

### *2.3.4 Genome-wide scans and other linkage studies*

So far, a total of 10 genome-wide scans have been published in ASDs (IMGSAC 1998; Barrett et al. 1999; Philippe et al. 1999; Risch et al. 1999; Buxbaum et al. 2001; IMGSAC 2001b; Liu et al. 2001; Shao et al. 2002b; Yonan et al. 2003; Cantor et al. 2005). In addition, two genome-wide scans have been performed in the Finnish population, as described in the experimental part of this thesis (studies I and II). Most of

the studies have included not only the strict phenotypic criteria for autism, but also broader autism phenotype involving individuals with AS and other PDDs. Also, several analyses focusing on specific chromosomal regions have been published (Ashley-Koch et al. 1999; Bass et al. 2000). Furthermore, there have been efforts to dissect the autism samples to dichotomous sub-groups, such as phrase speech delay (PSD) and obsessive-compulsive behaviours, or to analyse quantitative traits associated with autism (Bradford et al. 2001; Buxbaum et al. 2001; Alarcon et al. 2002; Shao et al. 2002a; Buxbaum et al. 2004; Alarcon et al. 2005). It is also of interest that overlapping sample is used in (i) the three studies by the International Molecular Genetic Study of Autism Consortium (IMGSAC) (IMGSAC 1998; IMGSAC 2001a; IMGSAC 2001b), (ii) the studies by Barrett and colleagues (1999) and Bradford and colleagues (2001), (iii) the three studies by Shao and colleagues (2002a; 2002b; 2003), and (iv) the studies, which have used the sample provided by the Autism Genetic Resource Exchange (AGRE) (Buxbaum et al. 2001; Liu et al. 2001; Alarcon et al. 2002; Yonan et al. 2003; Buxbaum et al. 2004; Alarcon et al. 2005; Cantor et al. 2005). Thus, linkage evidence from these studies should not be considered as independent. The sample sizes have ranged from 51 families and 100 affected individuals in the study by the Paris Autism Research International Sibpair Study (Philippe et al. 1999) to 345 families with 732 affected individuals in a study by Yonan and colleagues (2003). Study samples have consisted mostly of Caucasian families, but commonly a small proportion of families with different ethnic origin has also been included. Generally, non-parametric linkage methods, especially sib-pair methods, have been favoured. The published linkage studies in autism have been listed in Table 4. The most promising findings have been summarized in brief below.

**Table 4. Full-genome scans in autism.** The three best loci from each scan are presented (given that MLS > 1.5). See text for further details.

| Study | Sample size | Ethnicity | Phenotypes | Best loci | Maximum LOD score | Statistics |
|---|---|---|---|---|---|---|
| IMGSAC 1998 | 87 ASP | 100% Caucasian | autism, AS, PDD | 7q32-34 | 2.53 | ASPEX |
| | | | | 4p16 | 1.55 | Genehunter |
| | | | | 16p13 | 1.51 | |
| Barrett et al. 1999 | 75 ASP | 95% Caucasian | autism | 13q22 | 3.0 | ASP/MLINK |
| | | | | 13q12 | 2.3 | |
| | | | | 7q21 | 2.2 | |
| Philippe et al. 1999 | 51 ASP | 100% Caucasian | autism | 6q13 | 2.23 | MAPMAKER/SIBS |
| Risch et al. 1999 | 139 families | 88% Caucasian | autism | 1p13.2 | 2.15 | ASPEX |
| Buxbaum et al. 2001 | 95 families | not stated | autism, borderline | 2p32 | 2.39 | Genehunter NPL/HLOD |
| | | | autism, AS | 9q34 | 1.72 | |
| IMGSAC 2001a/ IMGSAC 2001b | 152 ASP (of which 83 analyzed in IMGSAC 1998) | 97% Caucasian | autism, AS, PDD-NOS atypical autism | 2q13 | 3.74 | ASPEX |
| | | | | 7q22 | 3.2 | |
| | | | | 16p13 | 2.93 | |
| | | | | 17q11 | 2.34 | |
| Liu et al. 2001 | 110 families | mixed USA | autism, AS; PDD | Xq25 | 2.67 | MAPMAKER/SIBS |
| | | | | 5p13 | 2.55 | |
| | | | | 19p13 | 2.53 | |
| Shao et al. 2002 | 99 families | 79% Caucasian | autism | Xq21 | 2.54 | ASPEX |
| | | | | 3p25 | 2.02 | Parametric MLOD |
| | | | | 7q34 | 1.66 | |

*Table 4 continued*

| Study | Sample size | Ethnicity | Phenotypes | Best loci | Maximum LOD score | Statistics |
|---|---|---|---|---|---|---|
| Yonan et al. 2003 | 345 families | mixed USA | autism, NQA, PDD | 17q11 | 2.83 | MAPMAKER/SIBS |
| | | | | 5p13 | 2.54 | |
| | | | | 11p11-13 | 2.24 | |
| Cantor et al. 2005 | 109 ASP | mixed USA | autism, NQA, PDD | 17q21 | 1.90 | MAPMAKER/SIBS |
| | | | | 3p13 | 1.80 | |

*Chromosome 2.* In the follow-up study of their original genome-wide scan, IMGSAC reported the highest non-parametric multipoint MLS of 4.80 at 2q31 with the strict diagnostic criteria for autism (IMGSAC 1998; IMGSAC 2001b). The best linkage evidence in the study by Buxbaum and colleagues (2001) was reported for 2q32, only some 5 cM distally to the finding by IMGSAC. The highest non-parametric MLS was 2.39 (Buxbaum et al. 2001). Also, Shao and colleagues (2002b) have reported modest linkage evidence for the long arm of chromosome 2.

*Chromosome 5.* Liu and colleagues (2001) reported their best locus at 5p13 with MLS of 2.55 in the initial analyses of the AGRE families. This region remained among the best loci in the follow-up study of the increased sample of 345 families with MLS 2.55 (Yonan et al. 2003). Also, Buxbaum and colleagues (2001) have reported modest linkage evidence in the vicinity (5p15) with MLS 1.65, but it should be borne in mind that AGRE families were also included in this study.

*Chromosome 7.* Linkage for the long arm of chromosome 7 was initially reported by IMGSAC (IMGSAC 1998) in 87 autism families with a multipoint MLS 2.53 at 7q32-7q34. Their follow-up analysis with a total of 170 multiplex families resulted in the highest multipoint MLS of 3.37 at 7q22, some 25 cM proximally to the original finding (IMGSAC 2001a). In addition, the studies by Barrett and colleagues (1999) (MLS = 2.2), Buxbaum and colleagues (2001) (MLS = 1.53), and Shao and colleagues (2002b) (MLS = 1.38) have reported modest evidence for linkage within the 7q21-34 region, which is currently denoted as autism susceptibility locus 1 (*AUTS1*).

*Chromosome 17.* The best locus in the analysis of the AGRE sample was observed at 17q11 with an MLS of 2.83 (Yonan et al. 2003). It was later suggested that this result was mainly due to increased sharing in male ASPs (Stone et al. 2004). This finding was later replicated in a subsequently collected sample of 91 AGRE families (Cantor et al. 2005). Interestingly, also the second largest family material by IMGSAC provided suggestive linkage evidence for this region with an MLS of 2.34 (IMGSAC 2001b). The

best marker in the IMGSAC study was intragenic for the serotonin transporter gene *SLC6A4*, which is considered as a strong functional candidate for autism (see, 2.3.5).

*Chromosome X.* Both family and cytogenetic studies indicate that the X-chromosome might be involved in the aetiology of autism. Surprisingly, there are only two studies, which have reported linkage for chromosome X. Liu and colleagues (2001) reported an MLS of 2.56 at Xq25 (Liu et al. 2001). However, decreased linkage evidence (MLS < 2) for this locus was reported in a follow-up study (Yonan et al. 2003). Shao and colleagues (2002b) reported an MLS of 2.52 at Xq21, which is located 20 cM proximally to the peak by Liu and colleagues.

*Other chromosomes.* Many of the loci that were among the best loci in independent studies have not been replicated by analyses of different samples. Risch and colleagues (1999) reported the highest LOD score in their sample for chromosome 1p13 with MLS of 2.15. This was the only locus that exceeded the MLS of 1.5 in their study (Risch et al. 1999). The best autosomal locus in the genome-wide scan by Shao and colleagues (2002b) was observed at 3p25 with MLS of 2.02. The best locus in the genome-wide scan by Philippe and colleagues (1999) was observed at 6q16 with MLS of 2.23. Barrett and colleagues (1999) found the highest MLS of 3.0 at 13q22. They also reported a second peak on chromosome 13 (13q12) with MLS 2.3 at 13q12 (Barrett et al. 1999). Finally, the third highest peak in the IMGSAC sample was observed at 16p13 with MLS of 2.92 (IMGSAC 2001b). Also the first analysis of AGRE sample supported this finding, but only marginal evidence for linkage was observed in the follow-up analysis of 345 families (Liu et al. 2001; Yonan et al. 2003).

*Subgroup analyses.* Re-analyses of the genome scan data have been performed by dividing the samples into component phenotypes in attempts to improve sample homogeneity. Although potentially helpful, the main disadvantage of this approach is that subgroup division necessarily compromises the sample size and thus the overall power to detect susceptibility loci. The primary sub-grouping considered to date has

been phrase speech delay (PSD) past 36 months, which occurs in about 50% of autistic disorder sibling pairs (Veenstra-VanderWeele and Cook 2004). Bradford and colleagues (2001) considered families with PSD ($n_{families}$=50) from the genome scan sample originally reported by Barrett and colleagues (1999). They showed that the linkage evidence obtained for the 7q21 and 13q22 were mainly attributable to the PSD group (Bradford et al. 2001). Buxbaum and colleagues (2001) incorporated analyses of PSD subgroup ($n_{families}$=49) in their genome-wide scan and observed increased linkage evidence at 2q32 with an NPL score of 3.32. Also Shao and colleagues (2002a) found increased evidence for linkage at 2q33 when the analyses were restricted only to families with PSD ($n_{families}$=45). The highest multipoint LOD score of 2.86 was observed some 20 cM distally to the best marker in a study by Buxbaum and colleagues. Buxbaum et al. (2004) also analyzed a subset of autism families having obsessive-compulsive behaviours ($n_{families}$=62). They reported the best linkage evidence at 1q42 with an NPL score of 3.06, along with other putative loci at 6q14 and 19p13. Some of the regional subgroup analyses have provided increased linkage evidence for the 15q11-13 locus. Shao et al. (2003) showed that the LOD scores increased from 1.45 to 4.71 at the *GABRB3* locus, when analyses were restricted to a subset of 23 families with high scores in the insistence of sameness factor. Nurmi and colleagues (2003b), by contrast, showed increased LOD scores for this region when a subgroup of 21 families with savant skills was analyzed.

*QTL analyses.* Alarcon and colleagues (2002) used 152 families from the AGRE sample (Geschwind et al. 2001) and considered three quantitative measures from the Autism Diagnostic Interview – Revised (ADI-R) screening questionnaire (Lord et al. 1994): age at first word, age at first phrase, and repetitive and stereotyped behaviour. The best linkage evidence was observed at 7q34-36 when age at first word was used as the quantitative trait (p=0.002). This locus is located somewhat (3-38 cM) distally to the best markers of the *AUTS1* locus reported in the IMGSAC sample (IMGSAC 1998; IMGSAC 2001a). Some evidence for repetitive and stereotyped behaviour was also observed at the same region (p=0.007) (Alarcon et al. 2002). In an expanded sample of a

total of 291 families, Alarcon and colleagues (2005) reported the highest peaks for age at first word for 3q ($p<0.001$) and 17q ($p=0.002$).

*Summary of linkage analyses.* It is clear from the review above that there is no consistent linkage evidence across the studies for any chromosomal region. Another striking observation is that only one study reported LOD scores exceeding the proposed threshold for significant linkage regardless of the fact that the aetiology of autism is thought to have a remarkably strong genetic influence (IMGSAC 2001b). Such observations are thought to reflect truly extensive genetic heterogeneity of the disorder, but also methodological differences between the studies. It is encouraging, however, that the two largest data sets provided evidence for linkage at 17q11-12, which might indicate that the reliability of the analyses increases when larger samples are employed (IMGSAC 2001b; Yonan et al. 2003; Cantor et al. 2005). Putative overlapping linkage findings have been reported also for 7q and 2q. However, even if some of the findings most probably reflect true genetic susceptibility factors for autism, it will be difficult to restrict the critical susceptibility locus, especially when obvious functional candidate genes for autism do not exist.

### *2.3.5 Candidate genes*

The list of positional or hypothesis-driven candidates that have been analyzed in autism is exhausting. Commonly, both positive and negative associations have been reported, which is characteristic for all candidate gene studies in complex neuropsychiatric disorders. Some of the most commonly analyzed candidates or genes with promising findings have been reviewed as examples below.

*SLC6A4/5-HTT.* 5-hydroxytryptamine transporter gene (*5-HTT*; solute carrier family 6 (neurotransmitter transporter, serotonin) member 4, *SLC6A4*) is probably one of the most intensively studies candidates in neuropsychiatric diseases. The 5-HTT is a target of selective serotonin reuptake-inhibitor (SSRI) drugs that are effective in treating anxiety,

depression, obsessive-compulsive behaviours, and also autistic behaviour (Glatt and Freimer 2002; Hollander et al. 2003b). Increased platelet and whole-blood serotonin levels have been reported approximately in one third of autistic patients and it has been suggested that this increase might be genetically determined (e.g. Abramson et al. 1989; Leventhal et al. 1990; Piven et al. 1991; Leboyer et al. 1999). It has therefore been hypothesized that the aetiology of autism could also be connected to the serotonergic system.

The *SLC6A4* gene has two functional variants, insertion/deletion polymorphism in the promoter region (5-HTT gene-liked promoter region, *5-HTTLPR*) and VNTR in the second intron, which have been primarily considered in the candidate gene analyses of neuropsychiatric diseases. Evidence for association between *SLC6A4* and autism was initially proposed by Cook and colleagues (1997), who reported preferential transmission of a short variant of the *5-HTTLPR* in a sample of 86 trios of mostly Caucasian origin. Two subsequent studies have reported preferential transmission of the short allele (Conroy et al. 2004; McCauley et al. 2004b), whereas two other studies have reported preferential transmission of the long allele instead of the short variant (Klauck et al. 1997; Yirmiya et al. 2001). Most of the studies, however, have failed to show an association between 5-HTT variants and autism (Maestrini et al. 1999; Zhong et al. 1999; Persico et al. 2000; Tordjman et al. 2001; Betancur et al. 2002; Kim et al. 2002a; Persico et al. 2002; Coutinho et al. 2004; Mulder et al. 2005). Also, haplotype analyses have revealed contradictory results (Cook et al. 1997; Betancur et al. 2002; Kim et al. 2002a; Conroy et al. 2004; McCauley et al. 2004b).

Recently, several studies have tested if the *SLC6A4* promoter polymorphisms are associated with the blood or platelet 5-HT levels in families with autism. Like in the other analyses concerning *SLC6A4* variants and autism, the conclusions have ranged from modest or a lack of association to significant contribution of the *SLC6A4* to the 5-HT levels (Anderson et al. 2002; Betancur et al. 2002; Persico et al. 2002; Coutinho et al. 2004). Furthermore, some authors have tested whether *5-HTTLPR* polymorphisms

could be associated with rigid-compulsive behaviours in individuals with ASDs but, again, with conflicting results (McCauley et al. 2004b; Mulder et al. 2005).

The analyses of *SLC6A4* variants in autism provide a typical example of candidate gene analyses in complex neuropsychiatric diseases, in which it is often difficult to interpret the results across the studies. The reported associations have generally been modest and, noteworthy, multiple-testing corrected p-values were reported in none of the studies. Furthermore, only few studies have considered more than two polymorphism or LD across the *SLC6A4* gene (Kim et al. 2002a; Conroy et al. 2004). Finally, the sample sizes have generally been modest ranging from 35 (Yirmiya et al. 2001) to 134 (Persico et al. 2002) families of mixed origin, indicating that most of the studies have been underpowered. It is therefore evident that there is a high probability for both type I and II errors (Ott 2004). Taken together, there is currently no convincing evidence that *SLC6A4* variants could confer susceptibility to autism or have an effect on hyperserotonemia in families with autism.

*Candidate genes on 7q*. Numerous candidate genes have been studied at 7q based on the recurrent linkage findings for this chromosome in autism. *RELN* gene is located at 7q22.1 and it encodes a protein called reelin. Hypothesis about the involvement of reelin in the pathogenesis of autism have been based on the relevant cellular function, deficits of reelin protein in post-mortem brains of autistic subjects, and *reeler* mouse with deleted *RELN* gene showing similar structural brain abnormalities that has been observed in autistic subjects. In addition, RELN gene variants have shown putative association in several other neuropsychiatric diseases such as schizophrenia and bipolar disorder (Fatemi 2002; Fatemi 2004). So far, seven association studies have been performed to analyze the role of the *RELN* gene in autism. Most of the studies have focused on the polymorphic CGG repeat located at the 5'UTR region of *RELN* gene. Two of these studies have shown evidence for association (Persico et al. 2001; Skaar et al. 2004) at the CGG repeat, whereas the others (Krebs et al. 2002; Zhang et al. 2002; Bonora et al. 2003; Devlin et al. 2004; Li et al. 2004) have yielded negative results.

The *FOXP2* gene located at 7q31 is mutated in a monogenic severe speech and language disorder (MIM 602081) (Fisher et al. 1998; Lai et al. 2000; Lai et al. 2001). The rationale to analyze *FOXP2* in autism comes from its physical location at autism susceptibility locus and from the phenotypic overlap of developmental language disorder and autism (Folstein and Mankoski 2000). Although one report of nominally positive association exists (Gong et al. 2004), the sequence and association analyses have generally failed to support the involvement of this gene in the etiology of autism (Newbury et al. 2002; Wassink et al. 2002; Gauthier et al. 2003).

A homeobox transcription factor Engrailed2 (*En2*) is located close to the telomeric end of chromosome 7 at 7q36.3. The mouse mutants of *En2* have been reported to have similar cerebellar morphological abnormalities as autistic patients, and thus *En2* has been analyzed as a candidate for autism. Recently, Gharani and colleagues (2004) reported a relatively strong association between the intronic SNPs of En2 and autism. An earlier study by Petit and colleagues (1995) reported the association of one *En2* variant in a case-control study. On the other hand, Zhong and colleagues (2003) failed to demonstrate an association in their family-based association test.

*Candidate genes on 15q11-13.* The chromosome 15q11-13 has been a focus of several candidate gene studies based on the frequent chromosomal aberrations in this region in the autistic patients (see, 2.3.3) and also due to existence of relevant functional candidates in this chromosomal region. There are also some studies, which have reported modest evidence for linkage on this region (Bass et al. 2000; Nurmi et al. 2003b; Shao et al. 2003), but generally 15q11-13 has not been among the best loci in the genome-wide scans.

The most widely studied candidates have been members of the γ-aminobutyric acid ($GABA_A$) receptor gene cluster, which contains genes coding for the α5, β3 and γ3 receptor units. GABA is the main inhibitory neurotransmitter in the human brain,

binding and acting through a complex series of $GABA_A$ receptors. It is also of interest that the GABAergic system is involved in the development of epilepsy (Baulac et al. 2001; Wallace et al. 2001), which is observed in 30% of autistic patients (Gillberg and Billstedt 2000). As noted earlier, the Angelman syndrome gene, *UBE3A*, is also located within this locus. Yet another maternally imprinted gene located at 15q11-13, *ATP10C* (ATPase, Class V, type 10A), has been studied in autism.

Most of the studies at 15q11-13 have included the *GABRB3* gene, for which both positive (Cook et al. 1998; Buxbaum et al. 2002; McCauley et al. 2004a) and negative (Maestrini et al. 1999; Salmon et al. 1999; Martin et al. 2000; Menold et al. 2001) association findings have been reported. Some studies have reported nominal association evidence at *GABRG3* (Menold et al. 2001) or *GABRA5* (McCauley et al. 2004a). However, all the reported associations for the components of the $GABA_A$ gene cluster have been modest. The association studies of the *UBE3A* and *ATP10C* loci have provided modest evidence for association in some studies (Nurmi et al. 2001; Nurmi et al. 2003a) but not in others (Kim et al. 2002b). Mutation analyses have failed to detect any pathogenic mutations in these genes (Veenstra-VanderWeele et al. 1999; Kim et al. 2002b).

*Other relevant candidates.* The best linkage evidence obtained in the French genome-wide scan for autism was observed at marker D6S283, which is intragenic for the glutamate receptor 6 gene (*GluR6/GRIK2*)(Philippe et al. 1999). Glutamate is the principal excitatory neurotransmitter in the brain and it is directly involved in cognitive functions. Therefore, glutamate receptors are candidates for several neuropsychatric diseases. A follow-up study of the genome-scan revealed increased IBD sharing of maternal alleles at the *GRIK2* locus. Furthermore, a TDT analysis resulted in promising association at two *GRIK2* variants. These associations were more pronounced when maternal transmission to male autistic subjects was tested suggesting an imprinting effect. Also, haplotype analysis with the two SNPs revealed a relatively strong association (Jamain et al. 2002). Modest association for this gene has been observed also

in the Chinese autistic trios (Shuang et al. 2004). More recently, evidence for association at *GRIN2A* gene (Glutamate receptor, ionotrophic, NMDA 2A) located on 16p13 was reported in the IMGSAC sample of 239 families and a replication set of 91 trios (Barnby et al. 2005).

Finally, Ramoz and colleagues (2004) performed mutation screening in the candidates at the 2q24-33 locus and identified two SNPs that showed relatively strong association with autism in a sample of 411 families. The identified SNPs are located within the *SLC25A12* gene, which encodes the mitochondrial aspartate/glutamate carrier (AGC1)(Ramoz et al. 2004). However, there are currently no published replication studies attempting to estimate the significance of this finding in independent samples.

# 3. AIMS OF THE PRESENT STUDY

This study aimed to dissect the genetic complexity of ASDs by following aims:

1) To use genome-wide approach to localize genetic loci, which harbour susceptibility genes for autism spectrum disorders in a well-characterized set of Finnish families.
2) To perform a genome-wide scan in the Finnish families with clinically well defined Asperger syndrome.
3) To perform combined data analysis of the genome-wide scans performed in the US and Finnish samples of autism families.
4) To analyse positional and hypothesis-driven candidate genes in Finnish families to assess their role in the genetic background of autism.

# 4. SUBJECTS AND METHODS

## 4.1 Laboratory procedures and analysis methods

Methods used in this study have been described in detail in the original articles (I-V) and listed in Table 5.

**Table 5. Methods used in the present study.**

| Method | Original publication |
|---|---|
| **Laboratory procedures** | |
| DNA extraction | I, II, III, IV, V |
| Polymerase Chain Reaction (PCR) | I, II, III, IV, V |
| Agarose gel electrophoresis | III, IV |
| Sequencing | III, IV |
| Electrophoresis, ABI377/ABI3730 | I, II, III, IV, V |
| Allele-specific primer extension – based microarray | IV |
| Sequenom MALDI-TOF mass spectrometry | V |
| **Analysis programs** | |
| Genescan 3.1 | I, II |
| Genotyper 2.0 | I, II |
| Genemapper 3.0 | III, IV |
| SNPSnapper | IV |
| Pedcheck 1.1 | I, II, III, IV, V |
| Sequencher 4.0.5 | III, IV |
| **Statistical methods** | |
| Downfreq 2.1 | I, II, III, V |
| MLINK/LINKAGE | I, II, IV |
| Homog 3.35 | I, II, IV |
| Analyze | I, II, IV |
| Genehunter v.2.1_r3beta;v.1.3 | I, II, III |
| Pseudomarker | III, IV |

*Table 5 continued.*

| Method | Original publication |
|---|---|
| Simwalk 2.81 | I |
| HRR | I |
| TDT-LRT | I |
| Gamete-competition | I |
| FBAT | IV, V |
| Dislamb | I |
| Haploview | IV |
| Genepop3.3; 3.4 | I, V |
| Mendel 4.0 | I |

## 4.2 Study subjects

### *4.2.1 Finnish families with autism, AS and dysphasia (I, III, IV,V)*

The family material used in the studies I, III, IV, and V was collected mainly via Helsinki University Hospital, Jyväskylä Central Hospital, and Kuopio University Hospital. These families included individuals diagnosed as having autism, AS, or developmental dysphasia. The diagnostic procedure used in the recruiting centres follows a highly uniform scheme. Initially, patients have been referred from outpatient clinic to the child neurological department for a two-week observation period involving a multidisciplinary battery of structured observations and tests. This period involves extensive diagnostic examinations, including neurological examinations, assessment of developmental history as well as psychological and neuropsychological evaluations using e.g. the Wechsler Intelligence Scale for Children – Revised (WISC-R) and Developmental Neuropsychological Assessment (NEPSY). Evaluations by speech pathologists, occupational therapists, and physiotherapists were performed using structured methods appropriate for the age and the developmental level of the child. Other specialists were consulted whenever it was considered necessary. During the

observation period, Childhood Autism Rating Scale (CARS) and the Psychoeducational Profile Revised (PEP-R) screening instruments were used to monitor autistic behavior and developmental profile. Both parents and a team of two nurses performed the CARS evaluation independently. Magnetic resonance imaging, electroencephalogram, cytogenetic screening, and a laboratory test package for detecting several metabolic and other disorders were routinely performed for all the patients. Families with associative medical conditions, such as fragile X syndrome, chromosomal aberrations, neurocutaneous syndromes, and profound mental retardation, were excluded from the sample. Particular attention was paid to the classification of the patients with developmental dysphasia. Reassessment was performed by a specialist for all the individuals diagnosed as having dysphasia. A total of 11/15 of patients with dysphasia had receptive language disorder (F80.2), which is similar to the DSM-IV diagnosis for mixed receptive-expressive language disorder. Three patients had the expressive type of dysphasia (F80.1) similar to the DSM-IV diagnosis for expressive language disorder. In one patient with dysphasia diagnosis, the specific diagnosis remained unspecified in the reassessment.

A stable team consisting of professionals with long experience carried out the diagnostic procedure and a consensus opinion of this team was required for the autism diagnosis. Finally, the diagnosis was set by a child neurologist in charge according to ICD-10 (World Health Organization 1993) and DSM-IV (American Psychiatric Association 1994) so that the criteria for both systems were fulfilled.

### *4.2.2 Finnish AS families (II)*

Diagnostic examinations in the AS sample used in study II consisted of detailed, structured interview, which was based on the ICD-10 (World Health Organization 1993) and DSM-IV (American Psychiatric Association 1994) criteria for Asperger syndrome. In some cases, the Autism Spectrum Screening Questionnaire (ASSQ) (Gillberg and Gillberg 1989; Ehlers and Gillberg 1993; Ehlers et al. 1999) and the Asperger Syndrome

Diagnostic Interview (ASDI) (Gillberg et al. 2001) was used to collect additional information. Neuropediatricians or research nurses with long experience of ASDs conducted all the interviews. All the interviews performed by a research nurse were further evaluated by a neuropediatrician. Structured interviews with one or several family members were used to collect reliable anamnestic information of adults included in the sample. The diagnostic interview also included questions on hypersensitivity to external stimuli, face recognition difficulties (prosopagnosia), motor clumsiness and sleeping and eating disorders, which are not diagnostic criteria but have been reported to be frequently present in individuals with AS (Nieminen-von Wendt 2004; Nieminen-von Wendt et al. 2005).

### *4.2.3 Autism Genetic Resource Exchange (AGRE) families (IV)*

The Autism Genetic Resource Exchange (AGRE) is a large central repository of DNA samples for genetic studies of autism. Families are recruited through a variety of methods, including physician referral, Web site contact, and family meetings and seminars (Geschwind et al. 2001; Liu et al. 2001; Yonan et al. 2003). The diagnostic assessment in the AGRE sample is based on the Autism Diagnostic Interview-Revised (ADI-R) (Lord et al. 1994), which is based on the classifications of the ICD-10 (World Health Organization 1993) and DSM-IV (American Psychiatric Association 1994). To be scored as affected, individuals must meet criteria in all three content areas of the ADI-R, including (i) quality of social interaction, (ii) repetitive, restricted and stereotyped interests and behaviour, and (iii) age at onset <3 years (Lord et al. 1994). AGRE family collection included three diagnostic classes based on the ADI-R questionnaire: Individuals that met the criteria for the ADI-R were considered to be in the "narrow" autism category, the "not quite autism" (NQA) category includes individuals who are no more than one point away from meeting criteria in all domains but do not meet age-at-onset criteria, and, lastly, individuals who might be considered as having AS or pervasive developmental disorder-not otherwise specified (PDD-NOS) were included in the "spectrum" category.

### *4.2.4 Diagnostic classes and sample sizes*

Three liability classes were generated for the sample of ASDs. The first class (LC1) included individuals with strictly defined autism, the second class (LC2) had also individuals with AS, and the third class (LC3) included family members with dysphasia as affected (studies I, IV, V). In studies III, IV and V, only families with at least one individual fulfilling the criteria for LC1 were selected for the analyses, whereas in study I some families with only broader classification (LC2 and LC3) were also included. The association analyses performed in studies IV and V were performed in a sample of 100 families with 122 affected individuals in LC1.

In the AS sample (study II), the first liability class (LC1) included individuals fulfilling the strict criteria for AS according to ICD-10. Family members who had AS–like features but who failed to meet one of the required criteria for diagnosis were included as affected individuals in a broad phenotypic category (LC2).

For study III, we generated three new liability classes. The first liability class (LC1) was restricted only to individuals with a diagnosis of narrow autism in both samples. In the second LC (LC2), also individuals with NQA from the AGRE sample were assigned as affected, and the third liability class (LC3) included a broad spectrum of autistic disorders from both samples. The individuals diagnosed as having AS (LC2 in study I) in the Finnish sample were included in the LC3. Description of sample sizes in different diagnostic classes is shown in Table 6.

**Table 6. Description of families used in studies I-V.** The numbers refer to the total sample. The number of families and affected individuals in the primary scans are shown in parentheses.

| Study | Liability Class | Families | Affected Individuals |
|---|---|---|---|
| I | 1 | 18 (12) | 39 (27) |
| | 2 | 28 (18) | 72 (41) |
| | 3 | 38 (19) | 87 (47) |
| II | 1 | 17 (13) | 72 (62) |
| | 2 | 17 (13) | 82 (66) |
| III[a] | 1 (AGRE) | 188 | 402 |
| | 2 (AGRE) | 218 | 468 |
| | 3 (AGRE) | 288 | 621 |
| | 1 (Finnish) | 18 | 38 |
| | 2 (Finnish) | 18 | 38 |
| | 3 (Finnish) | 26 | 55 |
| | 1 (Combined) | 206 | 438 |
| | 2 (Combined) | 236 | 504 |
| | 3 (Combined) | 314 | 676 |
| IV[b] | 1 | 19 | 41 |
| | 2 | 26 | 56 |
| | 3 | 33 | 74 |
| IV, V (association analyses)[c] | 1 | 100 | 122 |

a) For a detailed sample description of the first stage, see the study III. Note that one individual included in LC1 in study I was changed as unknown.

b) Only families with at least one individual fulfilling the criteria for LC1 in the study I were included. One novel family with three individuals in LC1 was included. Furthermore, one individual included in LC1 and one in LC2 in the study I were changed as unknown.

c) Association analyses in studies IV and V included the 33 families used in linkage analysis of the study IV and 67 additional families with single affected (LC1). Only individuals fulfilling criteria for LC1 were assigned as affected in the analyses resulting in a total number of 122 affected individuals in LC1.

# 5. RESULTS AND DISCUSSION

## 5.1 Genome-wide scan of ASDs in the Finnish population (I)

### *5.1.1 Primary scan*

We first performed a genome-wide scan in 38 Finnish families to identify genetic susceptibility loci for ASDs. The primary scan involved a genome-wide scan of 369 microsatellites in a sample of 19 families with a total of 47 affected individuals. In the linkage analyses, nine regions on chromosomes 1, 3p, 3q, 9, 12, 14, 17, 18, and 21 yielded in $Z_{max\ rec} > 1.0$ with the strict diagnostic criteria (Table 7). These loci were selected for follow-up analyses in the second stage. In addition to loci identified using strict diagnostic classification (LC1), some evidence for chromosome Xq13 was observed by using the broadest diagnostic classification (LC3). The best evidence was observed at DXS7132 in both two-point ($Z_{max\ dom}$=1.89, LC3) and multipoint ($NPL_{all}$=2.75, p=0.0027, LC3) analyses.

### *5.1.2 Fine-mapping*

In the second stage, the nine selected chromosomal regions were analysed further by an additional set of 60 markers. A total of 19 additional families with autism, AS, or dysphasia were included in this stage. The best linkage evidence was observed for 3q25-27. The highest two-point LOD score was observed at D3S3037 ($Z_{max\ dom}$=4.31, LC2) and the highest parametric multipoint LOD score of 4.81 (LC2, dominant model) was observed at the same marker. Some supporting evidence for this locus was obtained from haplotype and association analyses (see, study I). Several markers genotyped at 1p13-1q23 yielded $Z_{max} > 1.0$. The best two-point LOD score was observed at D1S1675 ($Z_{max\ dom}$=2.63, LC2), whereas in the multipoint analysis the highest LOD score of 2.63 (LC1, dominant model) was observed near D1S1653 located some 13 cM distally to D1S1675. Analyses at other fine mapping loci resulted in highly suggestive LOD scores. In addition to findings on 1p13-q23 and 3q25-27, LOD scores > 1.5 were observed only at 3p24, 14q12, and 17q25.

**Table 7. The highest two-point LOD scores LC1 in the primary scan.** The LOD scores are maximized over recombination fractions (θ) ranging from 0 to 0.5. The α-value ranging from 0 to 1.0 indicates the proportion of linked families.

| Marker | Position (cM) | Model | $Z_{max}$ | θ-value | α-value |
|---|---|---|---|---|---|
| D1S1675 | 149 | Dom | 1.11 | 0.00 | 1.00 |
| D3S3038 | 45 | Dom | 1.67 | 0.00 | 0.77 |
| D3S4009 | 137 | Rec | 1.60 | 0.00 | 0.59 |
| D3S3554 | 153 | Dom | 1.90 | 0.00 | 1.00 |
| D3S3053 | 182 | Rec | 1.10 | 0.04 | 0.65 |
| D3S2427 | 188 | Dom | 1.77 | 0.00 | 1.00 |
| D3S2418 | 216 | Rec | 1.40 | 0.12 | 1.00 |
| D9S158 | 162 | Rec | 1.62 | 0.00 | 0.64 |
| D12S2078 | 150 | Dom | 1.30 | 0.00 | 1.00 |
| D14S297 | 32 | Dom | 1.55 | 0.00 | 1.00 |
| D17S784 | 117 | Rec | 1.43 | 0.04 | 0.51 |
| D18S59 | 0 | Rec | 1.22 | 0.00 | 0.40 |
| D21S1440 | 36 | Rec | 1.29 | 0.00 | 0.44 |

### *5.1.3 Discussion of autism genome-wide scan*

In study I, we identified a major susceptibility locus for ASDs on 3q25-27 with convincing linkage evidence. This finding was supported by putative haplotype sharing of two families known to be related and by modest association findings using family-based tests (study I). Generally, the 3q25-27 region has not been among the most promising regions in the autism scans (see, 2.3.4), which suggests that different autism predisposing alleles may have enriched in the Finnish population due to founder effect, genetic drift and isolation (Peltonen et al. 1999). The second best locus was identified on

1p13-q23. In contrast to the 3q locus, some evidence for the nearby region has earlier been reported by Risch and colleagues (1999), and more recently by Bartlett and colleagues (2005). The main shortcomings of this study are the small sample size and large number of statistical tests. Although no permutation analyses were performed, it is clear that with the current sample size there is a high risk for type I and II errors. Despite these limitations, the solid linkage findings for 3q25-27 and 1p13-q23 provide excellent starting point for the follow-up studies.

### 5.2 Genome-wide scan of AS (II)

Most of the genome-wide scans in ASDs have focused on families, which are ascertained through a proband with childhood autism and other ASDs exist only in the probands' siblings. In the sample of Finnish families with ASDs we identified a group of families with AS, which were distinguishable from the families used in study I. First, AS was present without other ASDs. Second, the risk in the proband's relatives seemed to be much higher than the risk in the families ascertained through a proband with childhood autism. Finally, a seemingly dominant mode of inheritance of AS was detected. Thus, these 17 families were selected as a separate group for the second genome-wide scan in study II. The pedigree structures are shown in Figure 7.

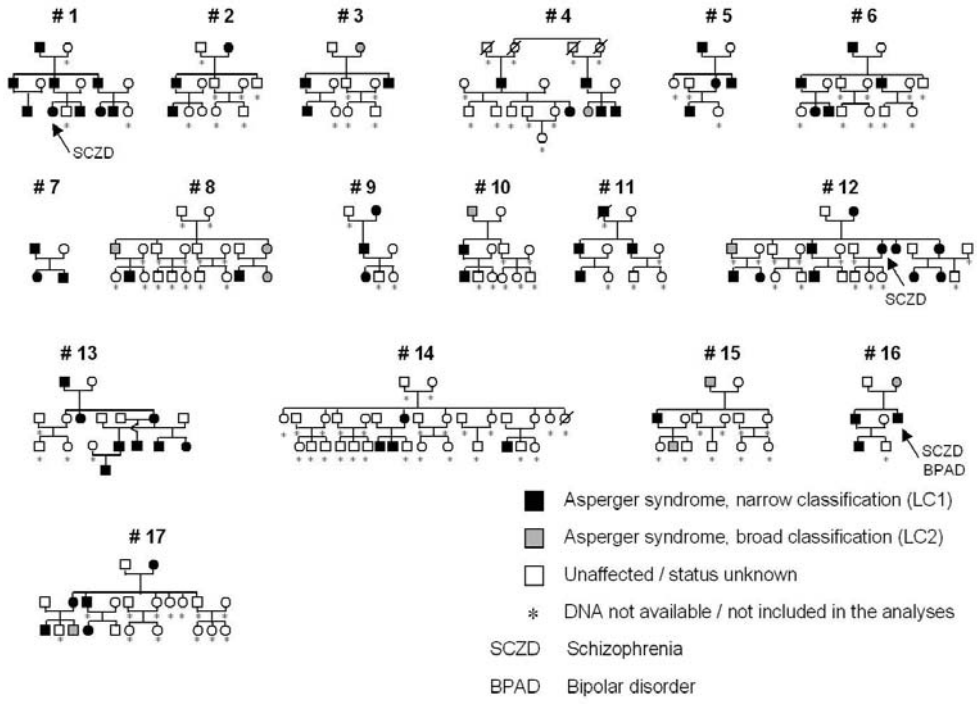


**Figure 7. Pedigree structures for the 17 families in the AS genome-wide scan (II).** Families 3, 7, 8, and 15 were included only in the second stage of the study.

### *5.2.1 Primary scan*

A genome-wide scan using a total of 415 microsatellites was performed in 13 multigenerational AS families. In the primary scan, we covered the loci at 1q21-23, 2q31-32, 3q25-27, 6q16, 7q, 13q12-22, and 16p13 with increased marker density based on the linkage findings from the earlier autism genome-wide scans. The two-point LOD scores were calculated by using the two liability classes and both dominant and recessive models. A total of 13 markers exceeded two-point LOD score of 1.5 in the primary scan. The best two-point LOD score was observed at 1q23.3 with D1S484 ($Z_{max\ dom}$=3.17).

Also, three other markers, D1S1653, D1S2705, and D1S1677, spanning 12 cM exceeded the two-point LOD score of 1.5 in this region. Other loci with LOD scores > 1.5 were identified at 3p24.1, 3q28, 4p15.1, 4q32.3, 6p25.1, 6q16.1-q16.3, 13q32.1, and 18p11.32, as shown in Table 8. These loci were selected for additional analyses in the fine mapping stage.

### *5.2.2 Fine mapping*

In the second stage, the best loci observed in the first stage were analysed by using the complete sample of 17 AS families. We added 22 new markers at these loci, the total number of markers being 54 with an average intermarker distance of 3.2 cM. Altogether, 8 out of 20 microsatellites spanning 19.7 cM at 1q21-23 resulted in the two-point LOD score > 1.0, the best marker being D1S484 ($Z_{max\ dom}$=3.58, LC1). The highest multipoint $NPL_{all}$ score at this locus was 1.33 (p=0.044, LC1). Eight markers covering 31 cM were analysed at 3p14-24. The best marker was D3S2432 ($Z_{max\ dom}$=2.50, LC1), but three other markers at this region also exceeded the LOD score of 1.0. The results from the multipoint analyses were consistent with the two-point results, the highest $NPL_{all}$ score being 3.32 (p=0.0029, LC1) at D3S2432. Two out of six markers genotyped at 13q31-33 exceeded the two-point LOD score of 1.0 in the second stage (D13S793, $Z_{max\ dom}$=1.59, LC2; D13S1271, $Z_{max\ dom}$=1.16, LC1). Increased linkage evidence was observed in the multipoint analyses with the $NPL_{all}$ of 2.86 (p=0.0025, LC1). Generally decreased LOD scores were observed for all the other loci included in the second stage analyses. However, markers with LOD scores > 1.0 were identified at all the fine mapped loci excluding 6p25.1 and 18p11.32, as shown in Table 8.

**Table 8. The highest two-point LOD scores from stages I and II of study II.** The markers with $Z_{max} > 1.5$ are presented.

| Marker | Position | Stage I | | | Stage II | | |
|---|---|---|---|---|---|---|---|
| | (cM) | LC | Model | Zmax | LC | Model | Zmax |
| D1S2715 | 159 | - | - | - | 1 | Dom | 1.75 |
| D1S2721 | 161 | - | - | - | 1 | Dom | 2.88 |
| D1S1653 | 164 | 1 | Dom | 2.47 | 1 | Dom | 1.53 |
| D1S484 | 170 | 1 | Dom | 3.17 | 1 | Dom | 3.58 |
| D1S2705 | 171 | 1 | Dom | 1.85 | 2 | Dom | 1.97 |
| APOA2 | 171 | - | - | - | 1 | Dom | 2.75 |
| D1S1677 | 176 | 2 | Dom | 1.78 | 2 | Dom | 1.21 |
| D3S2432 | 58 | 1 | Dom | 1.84 | 1 | Dom | 2.50 |
| D3S3685 | 68 | - | - | - | 1 | Dom | 1.54 |
| D3S2398 | 209 | 1 | Rec | 1.72 | 1 | Rec | 1.42 |
| D4S2408 | 46 | 2 | Rec | 1.99 | 2 | Rec | 1.04 |
| D4S2368 | 168 | 1 | Dom | 2.23 | 1 | Dom | 2.82 |
| D6S1668 | 9 | 1 | Rec | 1.66 | 1 | Rec | 0.85 |
| D6S1056 | 103 | 2 | Dom | 1.66 | 2 | Dom | 1.19 |
| D6S1671 | 108 | 2 | Dom | 1.73 | 2 | Dom | 1.07 |
| D13S793 | 76 | 1 | Dom | 1.74 | 2 | Dom | 1.59 |
| D18S59 | 0 | 1 | Rec | 1.71 | 1 | Rec | 0.98 |

### *5.2.3 Discussion of the AS genome-wide scan*

The most solid linkage evidence in the AS sample was observed at 3p14-24. Some earlier evidence for the nearby region was also observed in study I, in which two-point LOD score of 1.72 was observed with broad diagnostic classification (LC3) at D3S3659. This marker is located some 10 cM distally to the best marker D3S2432 in the AS sample. In their genome-wide scan of autism, Shao and colleagues (2002b) observed two-point LOD score 2.02 at D3S3680, which is located 22 cM distally to D3S2432.

The results at 1q21-23 are much more difficult to interpret. Some of the markers yielded promising results in the two-point analyses, but the multipoint evidence was not compelling. Multipoint analyses are sensitive to incorrect marker ordering or small number of genotyping errors. On the other hand, two-point analyses may fail to recognize all the recombinations due to missing genotype data, which may lead to false positive peaks at some individual markers. It is interesting, however, that this locus was among the best loci also in study I. This locus has also been the subject of intensive study in schizophrenia based on promising linkage findings (Brzustowicz et al. 2000; Gurling et al. 2001). In schizophrenia samples, some association evidence has been reported for *RGS4* and *CAPON* genes, which are both located at 1q23.3 (Chowdari et al. 2002; Brzustowicz et al. 2004).

Yet another putative susceptibility locus for AS was observed at 13q31-33. This locus has been among the best loci in schizophrenia studies, as it has also been in studies of bipolar disorder (Blouin et al. 1998; Brzustowicz et al. 1999; Detera-Wadleigh et al. 1999; Kelsoe et al. 2001; Badner and Gershon 2002). In the regional candidate gene analyses, the locus for the *G72/G30* gene at 13q33.2 has been reported to be associated with both schizophrenia and bipolar disorder (Chumakov et al. 2002; Hattori et al. 2003). Some shared features for some schizophrenia spectrum disorders and AS exist and some association between bipolar disorder and AS have also been suggested (DeLong and Dwyer 1988; Gillberg and Billstedt 2000; Wolff 2000). In our AS sample, three

individuals had an overlapping diagnosis of schizophrenia and one individual had bipolar disorder. Many family studies have found increased risk of depression, bipolar disorder and schizophrenia spectrum disorders in the relatives of patients with AS (DeLong and Dwyer 1988; Ghaziuddin 2005). Furthermore, one large-scale epidemiological study reported that risk for ASDs was markedly increased in children with a parental history schizophrenia-like psychosis or affective disorder (Larsson et al. 2005). It is therefore possible that some trait components with common genetic origin might be shared across different diagnostic groups. However, it is equally true that many studies have found only little overlap of ASDs and schizophrenia (Volkmar and Cohen 1991; Gillberg & Billstedt 2000; Cederlund and Gillberg 2004).

Taken together, the two genome-wide scans performed in studies I and II revealed a putative shared susceptibility locus at 1q21-23. Furthermore, in study I some linkage evidence emerged for the best locus in AS sample at 3p14-24. Similarly, some linkage evidence for the best locus at 3q25-27 in study I was observed also in the AS sample (Table 8). The shared loci with the linkage studies of other neuropsychiatric disorders at 1q21-23 and 13q31-33 might suggest the possibility of shared aetiological factors, but extensive studies are needed either to prove or disapprove these hypotheses.

### 5.3 Combined data analysis for autism loci (III)

The credibility of a linkage signal must be confirmed by an independent replication. In study III, we aimed (i) to pool raw genotype data from the Autism Genetic Resource Exchange (AGRE) and Finnish autism samples (study I) to reveal potential susceptibility loci common for both study samples, (ii) to analyze the best loci obtained in the AGRE sample in the study sample ascertained from the Finnish population, and (iii) to analyze the best loci that emerged from the Finnish sample in the AGRE families. This study design has a potential to confirm evidence for earlier identified susceptibility loci, but it could also identify novel loci for which the genetic effect is too small to be detected in a

single study, as exemplified elsewhere (Cavanaugh 2001; Demenais et al. 2003; Pajukanta et al. 2003).

### *5.3.1. Initial combined data analysis*

As the first stage, we combined raw genotype data from study I and from the study by Yonan and colleagues (Yonan et al. 2003). A total of 572 markers were included in the combined data set, of which 265 (46%) were common to both genome scans. Markers on nine loci (1p12-q25, 3p24-26, 4q21-31, 5p15-q12, 6q14-21, 7q33-36, 8q22-24, 17p12-q21 and 19p13-q13) exceeded the $NPL_{all}$ score of 1.64 ($p<0.05$) in the initial combined data analysis as shown in Table 9. In three of these, 1p12-q25, 3p24-26, 6q14-21, both datasets contributed to the linkage, whereas at the other loci linkage evidence emerged mainly from one data set only. On chromosome 1q, we observed the highest $NPL_{all}$ score of 2.25 ($p=0.012$) close to D1S1677 (LC3). On chromosome 3p, the best linkage evidence in the combined sample was detected at D3S2403 with $NPL_{all}$ of 2.10 ($p=0.014$, LC2). Finally, on 6q the overlapping linkage signals in the AGRE and Finnish samples produced increased linkage evidence at D6S1021 ($NPL_{all}$ = 2.47, $p=0.0046$, LC1).

**Table 9. Results from combined data analysis.** The multipoint $NPL_{all}$ scores > 1.64 ($p<0.05$) in the initial combined data analysis and the corresponding $NPL_{all}$ scores from separate data sets are presented. Results from the second stage are shown where available. Marker positions from pter (in cM) are from the Marshfield genetic map.

| | | Combined | | | | AGRE | | | | Finnish | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus (stage) | LC | Location (cM) | Marker | $NPL_{all}$ | p-value | Location (cM) | Marker | $NPL_{all}$ | p-value | Location (cM) | Marker | $NPL_{all}$ | p-value |
| 1p12-q25(I) | 3 | 176 | D1S1677 | 2.25 | 0.012 | 176 | D1S1677 | 1.66 | 0.048 | 169 | D1S2771 | 2.27 | 0.013 |
| 1p12-q25(II) | 3 | 176 | D1S1677 | 1.82 | 0.033 | 176 | D1S1677 | 1.33 | 0.088 | 169 | D1S2771 | 2.27 | 0.013 |
| 3p24-26 (I) | 2 | 29 | D3S3691 | 2.10 | 0.014 | 29 | D3S3691 | 1.79 | 0.032 | 37 | D3S2403 | 1.72 | 0.018 |
| 3p24-26 (II) | 2 | 29 | D3S3691 | 2.20 | 0.011 | 29 | D3S3691 | 1.79 | 0.032 | 37 | D3S2403 | 1.62 | 0.028 |
| 4q21-31 (I) | 3 | 107 | D4S1591 | 2.53 | 0.005 | 107 | D4S1591 | 2.67 | 0.003 | | | NS | |
| 5p15-q12 (I) | 3 | 57 | D5S1490 | 1.82 | 0.032 | 57 | D5S1490 | 2.04 | 0.019 | 19 | D5S807 | 2.09 | 0.02 |
| 6q14-21 (I) | 1 | 112 | D6S1021 | 2.47 | 0.0046 | 101 | D6S1056 | 2.04 | 0.017 | 112 | D6S1021 | 1.51 | 0.038 |
| 6q14-21 (II) | 1 | 112 | D6S1021 | 2.05 | 0.014 | 101 | D6S1056 | 2.04 | 0.017 | 112 | D6S1021 | 0.57 | 0.222 |
| 7q33-36 (I) | 2 | 165 | D7S483 | 2.31 | 0.008 | 165 | D7S483 | 2.33 | 0.008 | | | NS | |
| 8q22-24 (I) | 3 | 132 | D8S1832 | 2.07 | 0.018 | 132 | D8S1832 | 2.2o | 0.012 | | | NS | |
| 17p12-q21 (I) | 3 | 50 | D17S1294 | 2.39 | 0.007 | 50 | D17S1294 | 2.71 | 0.0028 | 62 | D17S1299 | 0.32 | 0.369 |
| 17p12-q21 (II) | 3 | 50 | D17S1294 | 2.38 | 0.0076 | 50 | D17S1294 | 2.71 | 0.0028 | 62 | D17S1299 | 0.54 | 0.287 |
| 19p13-q13 (I) | 1 | 42 | D19S714 | 1.77 | 0.032 | 42 | D19S714 | 1.78 | 0.033 | | | NS | |

### *5.3.2. Follow-up analyses on chromosomes 1, 3, 6 and 17*

In the follow-up stage, we analyzed additional markers on 1p12-q25, 3p24-26 and 6q14-21, where overlapping linkage peaks were observed in the first stage. Complete samples from both data sets were included at these loci. After follow-up analyses, the most promising overlapping linkage signal was observed at 3p24-26 with $NPL_{all}$ of 2.20 (p=0.011, LC2) at D3S3691 in the combined sample. The best $NPL_{all}$ scores for this locus in the separate analyses of Finnish and AGRE sample were 1.62 (p=0.028) at D3S2403 and 1.79 (p=0.032) at D3S3691, respectively. For the 1p12-q25 locus, slightly decreased linkage evidence was obtained in the second stage. In the combined sample, the best linkage evidence was observed at D1S1677 ($NPL_{all}$ = 1.82, p=0.033, LC3). The best findings in the individual samples were observed at D1S1677 ($NPL_{all}$ = 1.33, p=0.088) and D1S2771 ($NPL_{all}$ = 2.27, p=0.013) in the AGRE and Finnish samples, respectively. Analysing of the complete Finnish sample at 6q14-21 resulted in decreased linkage evidence compared to the initial stage (D6S1056, $NPL_{all}$=0.57, p=0.222, LC1). Therefore, the highest $NPL_{all}$ score of 2.05 (p=0.014) in the combined sample is mainly contributed by AGRE families.

The best loci observed in the separate analyses of Finnish and AGRE samples were 3q26-28 and 17p12-21, respectively. Therefore, we analyzed the 3q26-28 locus in the complete sample of AGRE families and the 17p12-q21 locus in the complete sample of Finnish families. At 3q25-27, the AGRE data set produced the highest linkage evidence at D3S1763 ($NPL_{all}$ = 1.30, p=0.089, LC1). With this same diagnostic classification, the Finnish sample produced the highest $NPL_{all}$ of 2.53 (p=0.0014) close to D3S3037, which is located around 14 cM distally to D3S1763. This suggests the potential overlap of the peaks identified in two samples. Using the narrow disease classification, the best linkage evidence in the combined sample was $NPL_{all}$ of 1.88 (p=0.024) at D3S1763. In the current analysis, the strongest linkage evidence in the Finnish sample was observed at D3S3037 ($NPL_{all}$ = 3.37, p=0.00059) using a broad classification.

Consistently with the first stage analysis, no support for linkage at 17p12-q21 locus emerged from the Finnish data set after analyzing the complete sample at this locus in the second stage (D17S1299, $NPL_{all}$ = 0.54, p=0.287, LC3). The best linkage evidence in both AGRE and combined samples were observed at marker D17S1294 with $NPL_{all}$ of 2.71 (p=0.0028, LC3) and $NPL_{all}$ of 2.38 (p=0.0076, LC3), respectively.

### *5.3.3. Sequence analyses of the OXTR gene*

The most promising shared locus for the AGRE and Finnish families was observed on 3p24-26, where both AGRE and Finnish samples contributed to the linkage (Figure 8). The 1-NPL-drop support interval on 3p24-26 extends around 14 cM (22-36 cM from ptel) corresponding approximately to 8.2 Mb of genomic sequence. A functional candidate for autism, oxytocin receptor gene (*OXTR*), is located only 29 kb distally to the best marker D3S3691. We analyzed the entire protein coding sequence and flanking splice sites of this gene in 22 probands with autism (10 from the Finnish sample, 12 from the AGRE sample) by direct sequencing. The probands were selected from the families that produced the highest LOD scores for D3S3691. A total of four sequence variants were identified, but none of these are obvious functional variants (Table 10). All variants excluding 712G>A are present also in healthy controls and are listed in public SNP databases (dbSNP). The 712G>A variant changes a hydrophobic alanine to hydrophilic threonine in the cytosolic side of the protein and it was inherited from the father to both affected siblings in the family. This amino acid is not located within known functional domains and it is not conserved between species suggesting that amino acid substitution at this position may not influence the functional properties of the receptor (Gimpl and Fahrenholz 2001). We also tested association with D3S3691 and the flanking markers D3S4545 and D3S2403 by using Pseudomarker, but found no evidence for association with either of the data sets (data not shown).

**Table 10.** The *OXTR* variants detected in 22 autistic probands.

| Variant number | Exon | Nucleotide change | Amino acid change | Frequency in 44 chromosomes |
|---|---|---|---|---|
| 1 | exon 3 | 171C>T | N57N | 34 |
| 2 | exon 3 | 652G>A | A218T | 6 |
| 3 | exon 3 | 690C>T | N230N | 11 |
| 4 | exon 3 | 712G>A | A238T | 1 |

### *5.3.4. Discussion of combined data analysis*

Study III was initiated to test, whether we could see shared linkage region(s) among Finnish and AGRE family collections. One of the main motivations was also the more detailed analyzes of earlier identified susceptibility loci in the independent study samples. We observed the most promising shared locus for the two samples on 3p24-26 (Figure 8). Also, some other studies with ASDs have provided linkage evidence for the nearby region. Most notably, Shao and colleagues (2002b) reported the second best locus in their genome-wide scan at D3S3680, which is located 7 cM proximal to the best marker in the combined data analysis, and only 1 cM away from the best marker in the Finnish data set. The best locus in study II was observed at D3S2432, which is located some 29 cM proximal to the best marker in the combined sample and 21 cM proximal to the best marker in the Finnish sample.

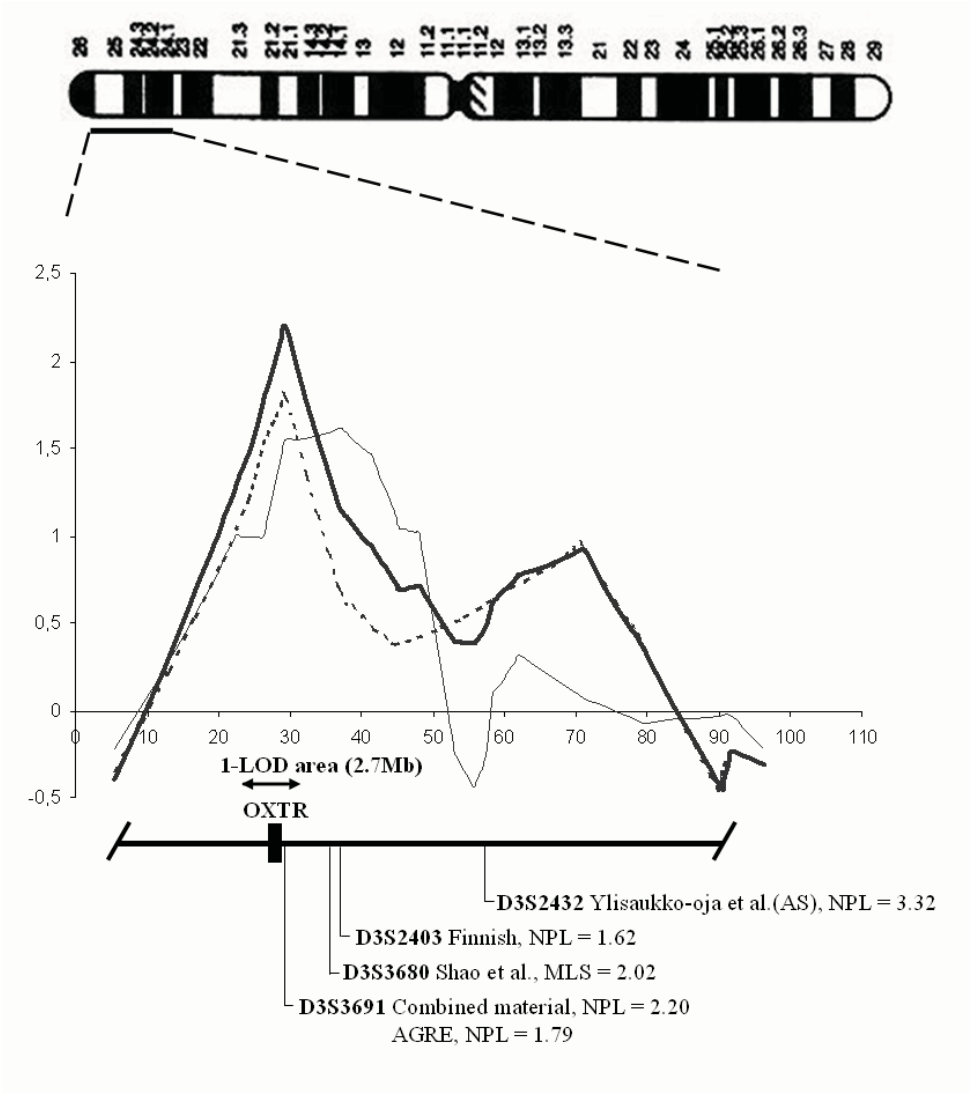


**Figure 8. Linkage findings for 3p in study III and in earlier studies.** The linkage curves have been indicated as follows: thick solid line = combined sample, thin solid line = Finnish sample, and thin dashed line = AGRE sample.

The 3p24-26 region contains the oxytocin receptor gene (*OXTR*), which is thought to be an excellent functional candidate for autism. In addition to its well-recognized role in a

variety of reproductive-related functions, oxytocin appears to have a role as a neurotransmitter or neuromodulator in the brain. Animal models have provided convincing evidence for its involvement in the basis of social attachment (Ferguson et al. 2000; Insel and Young 2001; Winslow and Insel 2002). In autism, reduced oxytocin plasma levels as well as deficits in oxytocin peptide processing have been proposed (Modahl et al. 1998; Green et al. 2001). There is also some evidence that infusion of synthetic oxytocin would ameliorate repetitive behaviors in adults with autistic symptoms (Hollander et al. 2003a).

As a preliminary analysis of *OXTR* gene, we sequenced the entire protein coding sequence in a sample of 22 autistic probands, but no obvious functional variants were detected. However, a more detailed study involving a dense set of markers is clearly needed in future to define the complete haplotype structure of this locus, and to test the association of *OXTR* variants with autism.

Only modest support for the earlier identified susceptibility loci on 1q21-23, 3q25-27 (study I) or 17q12 (Yonan et al. 2003) was observed in the independent study samples. These results support further the hypothesis of the enrichment of different autism predisposing alleles in the Finnish population due to the founder effect, genetic drift and isolation. Similarly, identification of only suggestive linkage signals in the large sample of 314 families proves once again that autism is genetically a highly heterogeneous disorder. However, this study indicates that the 3p24-26 locus deserves more attention and highlights the *OXTR* as a target for comprehensive molecular genetic studies in future.

### 5.4 Analysis of neuroligins as candidates for autism (IV)

Neuroligin genes have an essential role in the synaptogenesis and mutations in some of these genes have been shown to be causative for ASDs or MR, as reviewed in 2.3.2. Two out of the five members of the neuroligin gene family, namely *NLGN1* at 3q26 and *NLGN3* at Xq13, are located within best linkage peaks observed in study I, and were

therefore considered as primary candidates for autism in our sample. We also wanted to include *NLGN4* (located at Xp22) and its Y-chromosomal homologue *NLGN4Y* (Yq11) to the analyses based on the reported truncating mutations.

A total of 30 probands ($n_{male} = 26$; $n_{female} = 4$) were selected for the mutation analyses of the coding regions and the splice sites of *NLGN1*, *NLGN3*, *NLGN4*, and *NLGN4Y* genes. One proband from all the families, where affected individuals shared at least one haplotype IBD on Xq13 and/or 3q25-27, were selected for the analyses. In addition, two probands from the families producing positive LOD scores for 3q25-27 in the AS sample (study II) were included in the mutation analyses. A total of 27 individuals had a diagnosis of autism (LC1) and 3 individuals were diagnosed as having AS (LC2). We identified a total of six variants in the *NLGN1*: two silent coding sequence variants (K494K [1482G>A] and P818P [2454C>T]), three intronic variants, and one 5' UTR variant. None of these seemed to affect protein structure or were predicted to affect splicing (GeneSplicer Web Interface, the Berkeley Drosophila Genome Project, and the NetGene2 Server). We could not establish any sequence variants in *NLGN4* or *NLGN4Y* and only one rare variant was present in *NLGN3*. This was a silent Y74Y (222C>T) mutation in the first protein coding exon that was present in one affected male. Two of the variants identified in *NLGN1* were common (493-45A>G corresponds to rs3853390 and 1482G>A corresponds to rs7646919 in the dbSNP) and these were included in the SNP genotyping stage. Taken together, no obvious functional mutations were present in any of the genes analyzed. The identified variants have been listed in Table 11.

**Table 11. Sequence variants identified in the mutation analysis of the *NLGN1*, *NLGN3*, *NLGN4*, and *NLGN4Y* genes.** A total of 30 individuals ($n_{male}$ = 26; $n_{female}$ = 4) with ASDs were included in the analyses.

| Variant | Gene | Mutation type | Nucleotide change | Amino acid change | Exon/ Intron | Frequency in the autism sample |
|---|---|---|---|---|---|---|
| 1 | *NLGN1* | 5'UTR | -24C>T | - | exon 3 | 1 (1.7 %) |
| 2 | *NLGN1* | intronic | 492+67A>G | - | intron 3 | 1 (1.7 %) |
| 3 | *NLGN1* | intronic | 493-45A>G | - | intron 3 | 11 (18.3 %) |
| 4 | *NLGN1* | silent | 1482G>A | K494K | exon 8 | 7 (11.7 %) |
| 5 | *NLGN1* | intronic | 1771-45A>G | - | intron 8 | 1 (1.7 %) |
| 6 | *NLGN1* | silent | 2454C>T | P818P | exon 9 | 1 (1.7 %) |
| 7 | *NLGN3* | silent | 222C>T | Y74Y | exon 2 | 1 (2.9 %) |

Excluding the possibility of coding sequence mutations is not sufficient when candidate genes are evaluated in the complex diseases. Therefore, we employed a dense set of microsatellite and SNP markers to further dissect the role of *NLGN1*, *NLGN3*, and *NLGN4* genes. We genotyped a total of 16 microsatellites and 20 SNPs as follows: 6 microsatellites and 18 SNPs in *NLGN1*, 6 microsatellites and 2 SNPs in *NLGN3*, and 4 microsatellites in *NLGN4* (Figure 9). Intragenic and flanking microsatellites were selected from the Marshfield Medical Research Foundation map and UCSC Human Genome Browser (July 2003 assembly). Since no intragenic markers were available for the *NLGN3* gene, we generated two novel markers entitled ms.NLGN3-3 and ms.NLGN3-4, which were identified by using the Baylor College of Medicine Sequence Launcher's repeat masker algorithm (Smith et al. 1996). The UCSC Human Genome Browser, Celera Discovery System and dbSNP were used for selecting the SNPs for the analyses.

We confirmed the presence of linkage at the *NLGN1* and *NLGN3* loci, the best markers being D3S2421 ($Z_{max}$=2.58, LC1, dominant model) and DXS7117 ($Z_{max}$=2.39, LC2, dominant model), respectively. Also several other marker at these loci resulted in LOD score > 1.5, and thus the linked chromosomal region extends to overlap with *NLGN1* and *NLGN3*. By contrast, none of the four microsatellites analyzed at the *NLGN4* locus resulted in a LOD score > 1.5 (Table 12). Multipoint analyses were not performed, because the presence of LD between markers may lead to inflated LOD scores in analyses, which assume linkage equilibrium (Schaid et al. 2004). In association analyses, four out of 36 markers tested showed a trend towards association in a sample of 100 families including a total of 122 patients diagnosed as having autism (LC1). In *NLGN1*, the best association was observed at rs1488545 (FBAT, p=0.002; PSEUDOMARKER, p=0.041, recessive model). Also a nearby marker rs1352416, which was in LD with rs1488545 (D' = 1.0), showed some trend towards association (FBAT, p=0.004; PSEUDOMARKER, p=0.07). However, the microsatellite D3S1565 located only ~500 bp from rs1488545 did not show any evidence for positive association. Similarly, the analysis of two-marker haplotype constructed from rs1488545 and rs1352416 revealed less evidence for association compared with the evidence obtained for rs1488545 alone (rs1488545-rs1352416, p=0.01). Two of the X-chromosomal markers resulted in a nominal association in the dominant PSEUDOMARKER analysis. These were DXS7132 (p = 0.014) located at the 5' side of *NLGN3* and DXS996 (p=0.031) located within *NLGN4* (Table 12).

**Table 12. The markers, which yielded best linkage or association evidence in *NLGN1*, *NLGN3* and *NLGN4*.** Markers with $Z_{max} > 1.5$ or $p<0.05$ are shown. NS refers to $Z_{max} < 1.0$ in linkage analysis and $p>0.05$ in association analysis. The abbreviations are as follows: Het = heterozygosity, LC = liability class, Dom = dominant model, Rec = recessive model.

| Marker | Het | Linkage Zmax | Model (LC) | FBAT p-value | Pseudomarker p-value |
|---|---|---|---|---|---|
| **NLGN1:** | | | | | |
| D3S1556 | 0.80 | 2.00 | Dom (2) | NS | NS |
| rs635255 | 0.34 | 1.95 | Rec (2) | NS | NS |
| rs3853390 | 0.26 | 1.4 | Dom (1) | NS | 0.025 (Dom) |
| rs1352416 | 0.22 | NS | | 0.004 | NS |
| rs1488545 | 0.23 | NS | | 0.002 | 0.041 (Rec) |
| hCV1196655 | 0.44 | 2.16 | Rec (2) | NS | NS |
| D3S2421 | 0.85 | 2.58 | Dom (1) | NS | NS |
| **NLGN3:** | | | | | |
| DXS7132 | 0.74 | 1.89 | Dom (2) | - | 0.014 (Dom) |
| ms.NLGN3-3 | 0.58 | 1.71 | Dom (2) | - | NS |
| DXS7117 | 0.37 | 2.39 | Dom (2) | - | NS |
| DXS6800 | 0.68 | 1.74 | Dom (2) | - | NS |
| **NLGN4:** | | | | | |
| DXS996 | 0.80 | NS | | - | 0.031 (Dom) |

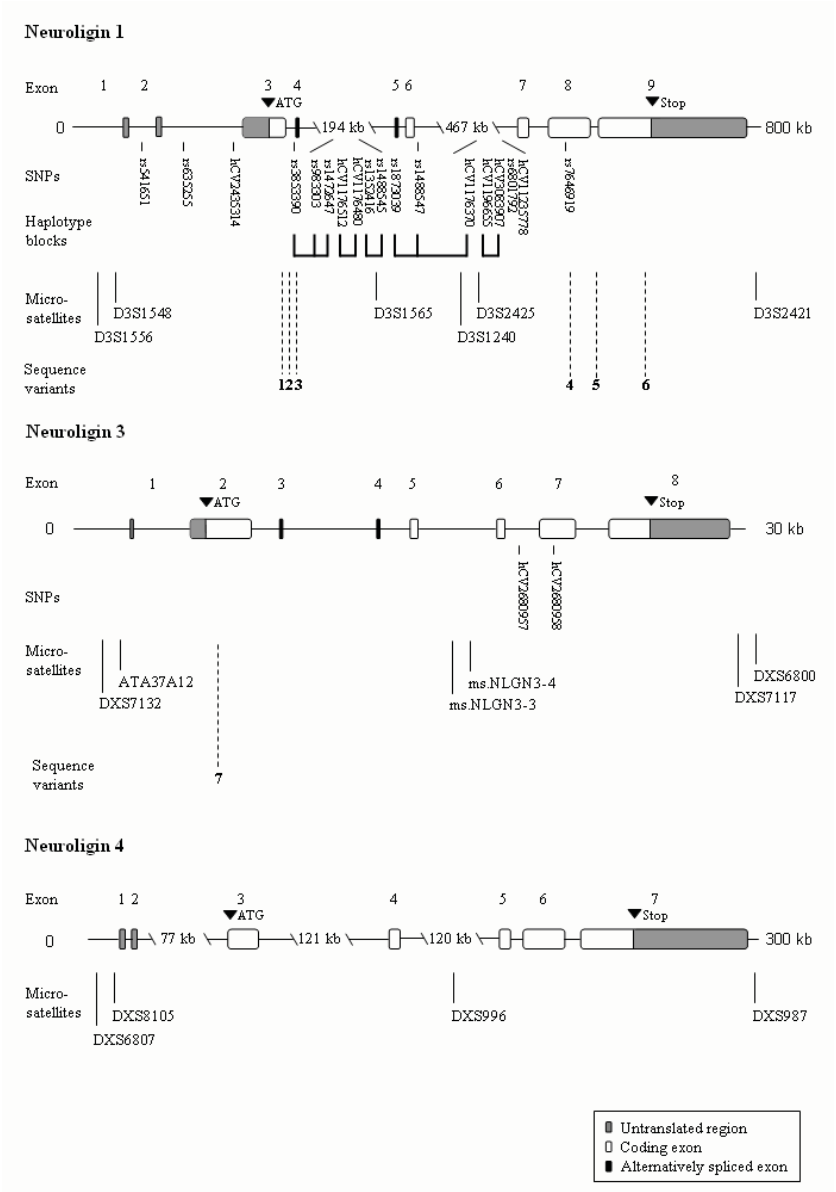


**Figure 9. Schematic overview of *NLGN1*, *NLGN3*, and *NLGN4* genes.** Genomic structures of the genes, analyzed SNPs and microsatellites as well as the identified sequence variants are presented. The sequence variant numbers correspond to the numbers in Table 11.

We identified a total of five haplotype blocks in *NLGN1* by using a solid spine of the LD method (D' > 0.8) of the HaploView - program (Figure 10)(Barrett et al. 2005). Block 1 involved four common haplotypes, Block 3 had two common haplotypes, and the remaining blocks three common haplotypes with a frequency > 5%. We tested haplotype association within these blocks but none of them yielded significant results (p>0.05). Data from 11 out of 18 SNPs analysed on *NLGN1* locus existed on the HapMap webpage. The LD pattern in the HapMap data (CEPH samples) was highly similar to the LD structure in the current Finnish sample (data not shown).

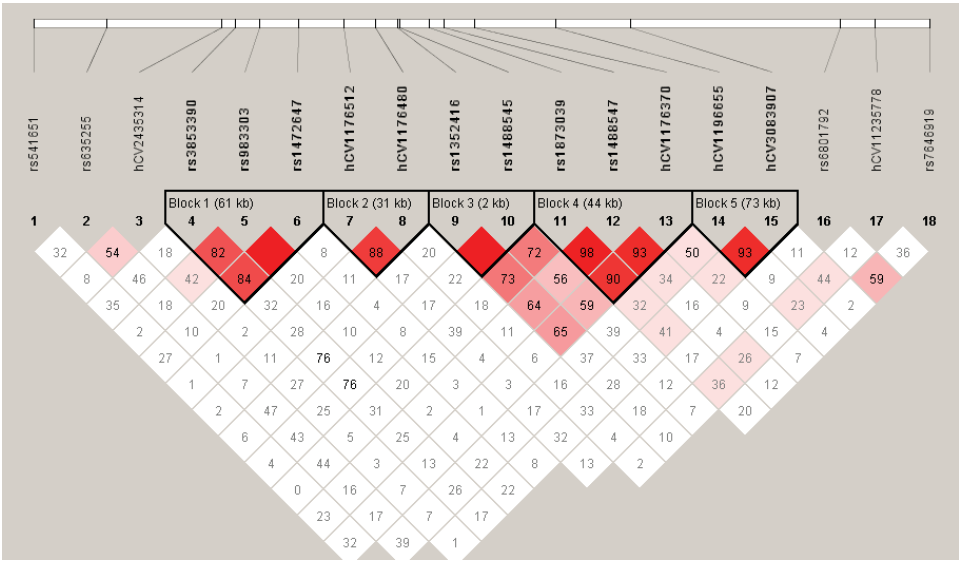


**Figure 10. Haplotype block structure of the *NLGN1* gene.** Haplotype blocks are defined by using the solid spine of LD method using D' of 0.8 as a threshold. D' values are shown within the boxes.

There are many reasons for the excitement aroused by the identification of neuroligin mutations by Jamain and colleagues (2003) and Laumonnier and colleagues (2004). First, these findings suggest that at least a small fraction of typical cases of autism might be caused by rare mutations with a major functional effect. Second, it is intriguing to notice that similar truncating *NLGN4* mutations seem to result in a wide variety of

phenotypes ranging from autism to the milder AS and to MR without autistic features. Third, neuroligin mutations link the aetiology of autism with mechanisms participating in the synaptogenesis and show that components involved in the synaptogenesis and synaptic structures are excellent functional candidates for future molecular genetic studies of autism and related disorders. Finally, functional analyses of the neuroligin pathway may eventually lead to better a understanding of the pathophysiology of ASDs.

We could not establish any functional mutations in any of the four neuroligin genes analyzed in 30 probands selected from families showing linkage for Xq13 and/or 3q26 loci. This implies that the coding sequence variants do not explain the linkage observed in our data set. Our findings are also consistent with the findings in other recent studies. Two other research groups have analyzed *NLGN3* and *NLGN4* in a total of 292 samples, but no causative mutations were identified (Gauthier et al. 2004; Vincent et al. 2004). In addition, one recent study reported rare missense variants in *NLGN4* but the atiological significance of these variants is not confirmed (Yan et al. 2005). Similarly, the two mutations described in the original report were identified among the sample of 158 autistic patients (Jamain et al. 2003). Thus, neuroligin mutations seem to be real but rare causes for autism. It is important to notice, however, that by the current study design we cannot exclude the existence of rare neuroligin mutations in some autism cases also in the Finnish population.

Although we identified a modest association at some of the markers analyzed, it is probable that they only account for random statistical fluctuation. The best association evidence was obtained for rs1488545 located in the fourth intron of *NLGN1* using the FBAT analysis program. However, none of the nearby markers or haplotype-based association analysis supported this finding. Also, the two X-chromosomal markers yielded only highly suggestive associations. It is not well established how p-values should be corrected for multiple testing when the tested markers are tightly correlated, but it is evident that none of the associations reported here would remain significant if such corrections were performed (Hennah et al. 2004). It is equally clear that more

stringent significance level criteria than the traditional $p<0.05$ are needed for convincing associations in the presence of multiple testing and low prior probability (Ott 2004). Taken together, our findings do not provide strong evidence for the involvement of any of these genes in the aetiology of autism in the Finnish sample.

### 5.5 Analysis of *DYX1C1* variants in the Finnish autism families (V)

Developmental dyslexia (MIM 127700) is the most common childhood learning disorder characterized by an unexpected difficulty in learning to read and write despite adequate intelligence, motivation, and education (Shaywitz 1998). Recently, the *DYX1C1* gene (dyslexia susceptibility 1 candidate 1) located at 15q21 has been identified as a putative candidate gene for dyslexia in a Finnish dyslexia sample (Taipale et al. 2003).

It has been suggested that *DYX1C1* variants might play a role in the broad spectrum of common childhood neuropsychiatric disorders involving learning and acquisition of specific competences (Grigorenko 2003). Autism is an extreme example of such a disorder, having e.g. severe impairment of speech and language development among its essential features. Therefore, we wanted to test whether the phenotype influenced by *DYX1* locus extends to autism by analysing the allelic distribution of the *DYX1C1* gene in a sample of 100 Finnish autism families with 122 affected family members. A total of four intragenic SNPs spanning ~68 kb of genomic sequence were analysed including –3G>A located on the 5'UTR (exon 2), rs3743204 in the second intron, 572G>A (G191E) in exon 5, and 1249G>T (E417X) in exon 10. We calculated family-based association for each individual marker as well as for two-, three-, and four-marker haplotypes constructed from the genotypes of the adjacent markers. None of these tests yielded positive associations as shown in Table 13. We also calculated the frequency of a two-marker haplotype of –3A/1249T shown to be positively associated with dyslexia. In the original report, 13% of cases with dyslexia were shown to carry this haplotype *versus* 5% of the controls, yielding an odds ratio of 2.8 ([CI95%]=1.2-6.5) (Taipale et al. 2003). The –3A/1249T haplotype frequency in the autism cases was 3%, which was equal to the

frequency in the unaffected founder individuals. Consistently with the association data, we did not observe linkage evidence with the *DYX1C1* intragenic SNPs or with the flanking microsatellites genotyped in the study I.

**Table 13. Results from the *DYX1C1* association analyses.** The analyses were performed by the FBAT program using the empirical variance option. Minor allele frequencies for each SNP are shown in parentheses.

| Variant | df | $\chi^2$ | p-value |
|---|---|---|---|
| **Single markers** | | | |
| –3G>A (0.04) | - | - | - |
| rs3743204 (0.09) | 1 | 0 | 1 |
| 572G>A (0.42) | 1 | 0.696 | 0.404 |
| 1249G>T (0.07) | 1 | 0.727 | 0.394 |
| **2 marker haplotypes** | | | |
| –3G>A + rs3743204 | 2 | 4.455 | 0.108 |
| rs3743204 + 572G>A | 2 | 1.167 | 0.558 |
| 572G>A + 1249G>T | 2 | 2.053 | 0.358 |
| **3 marker haplotypes** | | | |
| –3G>A + rs3743204 + 572G>A | 2 | 1.919 | 0.383 |
| rs3743204 + 572G>A + 1249G>T | 3 | 1.256 | 0.74 |
| **4 marker haplotype** | | | |
| –3G>A + rs3743204 + 572G>A + 1249G>T | 2 | 1.059 | 0.589 |

Despite the initial excitement of the results concerning *DYX1C1* as a dyslexia susceptibility gene, the replication attempts have generally been unsuccessful. After completing the present study, at least four replication studies have been published in dyslexia samples, which have failed to demonstrate significant association either with

dyslexia or with quantitative measures of the dyslexic phenotype (Scerri et al. 2004; Cope et al. 2005; Marino et al. 2005). Although some modest association has been reported, they have been for different SNPs or towards the opposite direction compared to original report (Scerri et al. 2004; Wigg et al. 2004). Population specific or diagnostic differences have been suggested to explain these differences (Marino et al. 2005), but it is also true that both the sample size and associations in the original study were quite modest (Taipale et al. 2003). Therefore, an association between *DYX1C1* variants and dyslexia cannot be considered as confirmed. Furthermore, we could not detect any evidence for association in the Finnish autism sample.

# 6. CONCLUDING REMARKS AND FUTURE PROSPECTS

To be successful, a genetic study needs a substantial contribution from experts in several fields. The foundation of this study has been collecting and characterising of more than 250 Finnish families with ASDs or related traits by a clinical team with a strong expertise of ASDs. Thus far, the clinical part of the study has involved the characterisation of the sample by using mainly conventional diagnostic nomenclatures of ICD-10 and DSM-IV. We are currently developing this area of research further towards more detailed phenotypic characterisation using novel screening instruments, which will enable more specific subgrouping and quantification of the phenotype. One obvious benefit of this study is that the sample is collected from the isolated population of Finland with the high quality national health care system and availability to monitor the health of children throughout the childhood years. This should provide the advantage of a relatively homogeneous genetic background and also ensure improved diagnostic homogeneity. A unique characteristic of the Finnish autism sample is surprisingly extensive relations among the families, which may prove an important feature in the future genetic mapping studies. Such founder effect has clearly provided an advantage of potentially increased IBD-sharing also in the studies presented here.

Sib-pair methodology has been the predominant mapping strategy in the autism studies. In the present study, we used our unique position in Finland to ascertain extensive families and extracted genetic information from entire pedigrees by using parametric methods. This approach should provide reasonably good power in linkage analysis even when a limited number of families are available. This was important because focusing on a small target population necessarily compromises the sample size available for mapping studies. The power of this study design is highlighted especially in study II, where up to 11 affected individuals per family were present.

Several putative susceptibility loci were identified in the present study. Studies I and II provide a solid starting point for the future fine-scale mapping studies, especially at the best loci 3q25-27 (study I) and 3p14-24 (study II). It is worth mentioning here, that some additional support for the best AS locus on 3p14-24 was recently observed in a subsequently collected sample of Finnish AS families (Rehnström et al. in press). Another promising finding was that studies I and II yielded some evidence for shared loci, especially on 1q21-23, but also on some other loci, such as on 3p. Support for the 3p region was observed also in study III, in which both Finnish and US families contributed to linkage for this locus. Although the studies of positional candidates (*NLGN1*, *NLGN3*, and *OXTR*) in studies III and IV did not yield promising results, they have been essential steps towards detailed characterisation of the identified linkage regions. Similarly, exclusions of other candidates, *NLGN4* and *DYX1C1* (studies IV and V), were based on the novel findings in the field of neuropsychiatric genetics and deserved detailed scrutiny in our sample. It is obvious, however, that with the sample size available it was possible to locate only variants with relatively strong effects.

Genetic research of autism has yielded only limited success in the past decade despite the continuing growth of interest and funding for autism research as well as technological and statistical advances in the field of human genetics. The lack of success reflects the enormous difficulty of the problem, which undoubtedly arises from truly heterogeneous background of the autism phenotype. The predominant hypothesis explaining the inheritance of autism is that numerous interacting genetic variants with subtle effects combined with contributions of some environmental risk factors result in the phenotype. This assumption is based on the evidence obtained from the twin and family studies, as reviewed in 2.3.1. Similarly, poor reproducibility in the numerous genome-wide scans performed so far is indicative of a genetically heterogeneous background of the disease (see, 2.3.4). It should be borne in mind, however, that some rare high-penetrance mutations or chromosomal aberrations are known to be causative for autism (reviewed in 2.3.2 and 2.3.3). Thus, the assumption of common predisposing genetic variants with subtle effect is not entirely correct.

Although technological advances and rapid accumulation of biological information provide excellent tools for disease gene mapping, all the efforts will presumably be wasted unless they are combined with intelligent research strategies. If a combination of common variants with minor effect truly predispose to autism, large-scale genome-wide association studies using large samples might prove useful. The mapping of rare mutations with a strong effect requires large pedigrees and extensive mutation screening of the positional candidates. Yet the alternative approach is to focus on population isolates or small sub-isolates, in which single causative variant is enriched in the affected individuals. Furthermore, characterisation and measurement of several trait components, clinical subtypes, or quantitative traits, will most likely result in biologically relevant phenotypes, which might be more readily mapped than the end-phenotype. It is also clear that the latest technological advantages, such as gene expression profiling, array-based copy number analysis, and high-throughput genotyping technologies, should be fully exploited. Therefore, the strategies to map autism genes should not be based only on one predominant hypothesis or approach, but rather, we should accept our ignorance of the true underlying genetic background and a variety of mapping strategies using different types of samples and methods should be employed.

Despite the difficulties experienced so far, it is certain that genetic studies will eventually be able to characterize factors important for developing autism. This information will pave the way towards identification of novel metabolic pathways underlying autism, and thus provide new insights into the aetiopathogenesis of this and related disorders. It is hoped that in more distant future, this type of research will eventually result in development of new diagnostic markers enabling earlier and more specific diagnosis, and provide new starting points for drug development.

# 7. ACKNOWLEDGEMENTS

This study was carried out in the Department of Molecular Medicine at the National Public Health Institute, Helsinki. I wish to thank Jussi Huttunen and Pekka Puska, the former and present heads of the institute, for providing excellent research facilities.

I have been privileged to work with two outstanding supervisors. The leading force of this study has been the scientific insight of Professor Leena Peltonen-Palotie. Her wide expertise and knowledge in the field of molecular genetics as well as the dedication and drive for scientific discovery can only be admired. I would like to thank Leena for creating a most inspiring working environment, and for all the support and responsibility she has given. It is amazing that she always has time for help and discussion with her students, no matter how busy she is. I want to thank Docent Irma Järvelä for giving me the opportunity to work in this project already when I was a second-year student. Irma is gratefully thanked for initiating the project, running it with enthusiasm, and for sharing many long, educative, and interesting discussions.

I wish to thank Professor Thomas Bourgeron for accepting the role as Opponent in my thesis defence.

Professor Peter Szatmari and Docent Tiinamaija Tuomi are thanked for reviewing this thesis and for their helpful comments. Donald J.M. Smart, B.Sc, MA, is thanked for language revision of this thesis.

This type of study would not have been possible without substantial contribution from people with clinical expertise. I have been privileged to work with several leading experts of autism spectrum disorders in Finland, including Professor Lennart von Wendt, Raija Vanhala, MD, PhD, Taina Nieminen-von Wendt, MD, PhD, and Reija Alen, MD. I would like to thank especially Lennu for being my unofficial third supervisor. The collaboration by numerous other clinicians, especially by Professor Raili Riikonen, has

been extremely valuable. The contribution by Susan Sarenius has been absolutely indispensable for this project. The genealogical work by Teppo Varilo, MD, PhD, has made our autism sample unique. Teppo is also thanked for being excellent company in the lab.

This study has involved several collaborative projects with other research groups. Professor Daniel Geschwind, Rita Cantor, PhD, and especially Maricela Alarcón, PhD, from UCLA are thanked for the collaboration in the study III. Professor Juha Kere, Cecilia Lindgren, PhD, and Myriam Peyrard-Janvid, PhD, from Karolinska Institutet are thanked for the collaboration in the study V. Professor Risto Näätänen, Docent Teija Kujala and Tuulia Lepistö, MA, are thanked for the collaboration in the LIFE2000 consortium.

Senior researchers in our laboratory, Drs. Anu Jalanko, Ismo Ulmanen, Marjo Kestilä, Janna Saarela, Iiris Hovatta, Anu Loukola, Tiina Paunio, Markus Perola, Juha Saharinen and many others, have helped me in many ways during the past years. Their interest towards my work is gratefully appreciated.

Elli Kempas is thanked for producing vast amount of data as well as for all the help and support during my thesis work. This work would have taken years longer without Elli's effort. I have been extremely lucky to have Karola Rehnström as a fellow worker. Karola is thanked especially for taking care of numerous responsibilities, running several important branches of the study, making the sequences I should have made, and colouring everything in pink. Mari Auranen is thanked for introducing me to the world of autism genetics and building the foundations for the project. It has been a pleasure to follow the enthusiasm and drive of the most recent members of the autism/MR team, Helena Kilpinen and Anne Koivisto.

I wish to thank Mikko Kuokkanen, soon PhD, for his support and friendship. It has been great to have someone to discuss about the meaning of doing science. Mikko is also

thanked for frequently beating me up in badminton and tennis. Nabil is thanked for long discussions concerning everything from gene mapping to life in Libya. Heli is thanked for bringing up the female point of view in our room. Susanna, Hanna, Ilona and other members of the Irma's group are thanked for the cheerful company.

I would like to thank Joni for serious (?) discussions about science as well as for sharing many memorable congress trips. Pekka is acknowledged for sharing his superior technical expertise and for his attempts to make me to do sports. I wish to thank Jesper for teaching me the basics of linkage analysis and Tero H for helping me with them. William is thanked for proving that you can do good science even in the field of psychiatric genetics. The funny people at the Department of Molecular Medicine make it a special place to work. I'd like to thank Mira, Iita, Niklas, Henna, Tintti, Annina, Juho, Teemu, Jari, Juha, Markus, Emma, Heidi, Jussi, Minna, Sisko, Ritva, Arja, Marika, and Elina, just to mention few (and I know that I missed many important persons). Heli and Susanna are thanked for excellent sequencing service. Sari K, Sari M, Sami, Mika and Tuija are thanked for helping in numerous practical matters.

I wish to thank Apeli, Jussi, Oltsu, Teemu, Perttu & Emma, Miikka, Outi, & Emilia, Inga, Rami & Veera, and Heli & Pete for showing that there is life outside the lab. Sinikka, Markku and the others are thanked for always being so helpful. Many thanks go to my grandparents.

I would like to express my deepest gratitude to my parents for their endless support. Especially, I wish to thank my mother whose work is all too rarely acknowledged. My younger brother, Petri, is thanked for being the best friend one could hope for. He is also thanked for keeping me updated in everything sports-related during the years I've spent in the lab. I want to thank also my older brother Arto and his spouse Tiina. Arto has always been my idol and the clever one in our family.

Our lovely dog Jaffa is thanked for taking me out even when I did not know that I needed it.

The past over five years of thesis work have involved a lot of fun and excitement, but also days filled with frustration and depression. It has been a tremendous source of strength to have someone to share all the ups and downs. Thus, my deepest gratitude goes to my wife Sanna. This work would have never been completed without her love, support and care. The moments with Eemil have brought a tremendous amount of joy, laughter, and wonders to our life.

I wish to thank all the families who participated in this study.

This study has been financially supported by the Academy of Finland, the Pediatric Research Foundation (Ulla Hjelt Fund), the Päivikki and Sakari Sohlberg Foundation, Sigrid Juselius Foundation, Biomedicum Helsinki Foundation and Helsinki University Foundation. The support from the Cure Autism Now Foundation for the future work is gratefully acknowledged.

Kirkkonummi, October 20th, 2005

Tero Ylisaukko-oja

# 8. ELECTRONIC DATABASE INFORMATION

Autism Genetic Resource Exchange, http://www.agre.org/

Baylor College of Medicine (BCM) Sequence Launcher,
http://searchlauncher.bcm.tmc.edu/

Berkeley Drosophila Genome Project, http://www.fruitfly.org/seq_tools/splice.html

Celera Discovery System, http://www.celeradiscoverysystem.com/

dbSNP, www.ncbi.nlm.nih.gov/SNP

Ensembl Genome Browser, http://www.ensembl.org/

GeneSplicer Web Interface, http://www.tigr.org/tdb/GeneSplicer/gene_spl.html

Genetic Power Calculator, http://statgen.iop.kcl.ac.uk/gpc/

Genome Database, http://www.gdb.org

International HapMap Project, http://www.hapmap.org

Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics

Multiplex PCR primer designer, http://apps.bioinfo.helsinki.fi/mpd

NetGene2 Server, http://www.cbs.dtu.dk/services/NetGene2/

Online Mendelian Inheritance in Man, www.ncbi.nlm.nih.gov/Omim

Primer 3, http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

SNPSnapper, http://www.bioinfo.helsinki.fi/SNPSnapper/

UCSC Human Genome Browser, http://genome.ucsc.edu/

# 9. REFERENCES


Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279-92

Abramson RK, Wright HH, Carpenter R, Brennan W, Lumpuy O, Cole E, Young SR (1989) Elevated blood serotonin in autistic probands and their first-degree relatives. J Autism Dev Disord 19:397-407

Abreu PC, Greenberg DA, Hodge SE (1999) Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. Am J Hum Genet 65:847-57

Alarcon M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH (2005) Quantitative genome scan and ordered-subsets analysis of autism endophenotypes support language QTLs. Mol Psychiatry Epub Apr 12.

Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH (2002) Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. Am J Hum Genet 70:60-71

Allen G, Harvald B, Shields J (1967) Measures of twin concordance. Acta Genet Stat Med 17:475-481

Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet 69:936-50

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Washington, DC

Anderson GM, Gutknecht L, Cohen DJ, Brailly-Tabard S, Cohen JH, Ferrari P, Roubertoux PL, Tordjman S (2002) Serotonin transporter promoter variants in autism: functional effects and relationship to platelet hyperserotonemia. Mol Psychiatry 7:831-6

Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, Winchester E, Lander ES, Kruglyak L (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. Am J Hum Genet 69:582-9

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

Arnheim N, Calabrese P, Nordborg M (2003) Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. Am J Hum Genet 73:5-16

Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh MB, Aylsworth AS, Vance JM, Gilbert JR, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (1999) Genetic studies of autistic disorder and chromosome 7. Genomics 61:227-36.

Asperger H (1944) Die autistischen Psychopathen im Kindersalter. Archiv für Psychiatrie und Nervenkrankheiten 1:76-136

Asperger H (1991) Autistic psychopathy in childhood (trans.). In: Frith U (ed) Autism and Asperger syndrome. Cambridge University Press, Cambridge, UK, pp 37-92

Badner JA, Gershon ES (2002) Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. Mol Psychiatry 7:405-11

Baieli S, Pavone L, Meli C, Fiumara A, Coleman M (2003) Autism and phenylketonuria. J Autism Dev Disord 33:201-4

Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M (1995) Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 25:63-77.

Bailey A, Phillips W, Rutter M (1996) Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. J Child Psychol Psychiatry 37:89-126

Barnby G, Abbott A, Sykes N, Morris A, Weeks DE, Mott R, Lamb J, Bailey AJ, Monaco AP (2005) Candidate-Gene Screening and Association Analysis at the Autism-Susceptibility Locus on Chromosome 16p: Evidence of Association at GRIN2A and ABAT. Am J Hum Genet 76

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-5

Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL, Childress D, et al. (1999) An autosomal genomic screen for autism. Collaborative linkage study of autism. Am J Med Genet 88:609-15.

Bartlett CW, Goedken R, Vieland VJ (2005) Effects of updating linkage evidence across subsets of data: reanalysis of the autism genetic resource exchange data set. Am J Hum Genet 76:688-95

Bass MP, Menold MM, Wolpert CM, Donnelly SL, Ravan SA, Hauser ER, Maddox LO, Vance JM, Abramson RK, Wright HH, Gilbert JR, Cuccaro ML, DeLong GR, Pericak-Vance MA (2000) Genetic studies in autistic disorder and chromosome 15. Neurogenetics 2:219-26

Baulac S, Huberfeld G, Gourfinkel-An I, Mitropoulou G, Beranger A, Prud'homme JF, Baulac M, Brice A, Bruzzone R, LeGuern E (2001) First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. Nat Genet 28:46-8

Berney TP, Ireland M, Burn J (1999) Behavioural phenotype of Cornelia de Lange syndrome. Arch Dis Child 81:333-6

Betancur C, Corbex M, Spielewoy C, Philippe A, Laplanche JL, Launay JM, Gillberg C, Mouren-Simeoni MC, Hamon M, Giros B, Nosten-Bertrand M, Leboyer M (2002) Serotonin transporter gene polymorphisms and hyperserotonemia in autistic disorder. Mol Psychiatry 7:67-71

Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, et al. (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet 20:70-3.

Bolliger MF, Frei K, Winterhalter KH, Gloor SM (2001) Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression. Biochem J 356:581-8

Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, Rutter M (1994) A case-control family history study of autism. J Child Psychol Psychiatry 35:877-900.

Bonora E, Beyer KS, Lamb JA, Parr JR, Klauck SM, Benner A, Paolucci M, Abbott A, Ragoussis I, Poustka A, Bailey AJ, Monaco AP (2003) Analysis of reelin as a candidate gene for autism. Mol Psychiatry 8:885-92

Boomsma D, Busjahn A, Peltonen L (2002) Classical twin studies and beyond. Nat Rev Genet 3:872-82

Borg I, Squire M, Menzel C, Stout K, Morgan D, Willatt L, O'Brien PC, Ferguson-Smith MA, Ropers HH, Tommerup N, Kalscheuer VM, Sargan DR (2002) A cryptic deletion of 2q35 including part of the PAX3 gene detected by breakpoint mapping in a child with autism and a de novo 2;8 translocation. J Med Genet 39:391-9

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

Bowman EP (1988) Asperger's syndrome and autism: the case for a connection. Br J Psychiatry 152:377-82.

Bradford Y, Haines J, Hutcheson H, Gardiner M, Braun T, Sheffield V, Cassavant T, Huang W, Wang K, Vieland V, Folstein S, Santangelo S, Piven J (2001) Incorporating language phenotypes strengthens evidence of linkage to autism. Am J Med Genet 105:539-47

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861-9

Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS (2000) Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. Science 288:678-82.

Brzustowicz LM, Honer WG, Chow EW, Little D, Hogan J, Hodgkinson K, Bassett AS (1999) Linkage of familial schizophrenia to chromosome 13q32. Am J Hum Genet 65:1096-103.

Brzustowicz LM, Simone J, Mohseni P, Hayter JE, Hodgkinson KA, Chow EW, Bassett AS (2004) Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. Am J Hum Genet 74:1057-63

Burgoine E, Wing L (1983) Identical triplets with Asperger's syndrome. Br J Psychiatry 143:261-5.

Buxbaum JD, Silverman J, Keddache M, Smith CJ, Hollander E, Ramoz N, Reichert JG (2004) Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19. Mol Psychiatry 9:144-50

Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, Reichert J, Cook EH, Jr., Fang Y, Song CY, Vitale R (2002) Association between a GABRB3 polymorphism and autism. Mol Psychiatry 7:311-6

Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, Hollander E, Lawlor BA, Fitzgerald M, Greenberg DA, Davis KL (2001) Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. Am J Hum Genet 68:1514-20.

Cantor RM, Kono N, Duvall JA, Alvarez-Retuerto A, Stone JL, Alarcon M, Nelson SF, Geschwind DH (2005) Replication of autism linkage: fine-mapping peak at 17q21. Am J Hum Genet 761050-1056.

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

Carney RM, Wolpert CM, Ravan SA, Shahbazian M, Ashley-Koch A, Cuccaro ML, Vance JM, Pericak-Vance MA (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. Pediatr Neurol 28:205-11

Castermans D, Wilquet V, Parthoens E, Huysmans C, Steyaert J, Swinnen L, Fryns JP, Van de Ven W, Devriendt K (2003) The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. J Med Genet 40:352-6

Cavanaugh J (2001) International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 68:1165-71

Cawthon RM, Weiss R, Xu GF, Viskochil D, Culver M, Stevens J, Robertson M, Dunn D, Gesteland R, O'Connell P, et al. (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. Cell 62:193-201

Cederlund M, Gillberg C (2004) One hundred males with Asperger syndrome: a clinical study of background and associated factors. Dev Med Child Neurol 46: 652-660.

Chakrabarti S, Fombonne E (2001) Pervasive developmental disorders in preschool children. Jama 285:3093-9.

Chakravarti A (1999) Population genetics--making sense out of sequence. Nat Genet 21:56-60

Charman T (2002) The prevalence of autism spectrum disorders. Recent evidence and future challenges. Eur Child Adolesc Psychiatry 11:249-56

Chih B, Afridi SK, Clark L, Scheiffele P (2004) Disorder-associated mutations lead to functional inactivation of neuroligins. Hum Mol Genet 13:1471-7

Chih B, Engelman H, Scheiffele P (2005) Control of excitatory and inhibitory synapse formation by neuroligins. Science 307:1324-8

Chowdari KV, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T, Deshpande SN, B KT, Ferrell RE, Middleton FA, Devlin B, Levitt P, Lewis DA, Nimgaonkar VL (2002) Association and linkage analyses of RGS4 polymorphisms in schizophrenia. Hum Mol Genet 11:1373-80

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

Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 65:1170-7

Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580-1

Comoletti D, De Jaco A, Jennings LL, Flynn RE, Gaietta G, Tsigelny I, Ellisman MH, Taylor P (2004) The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. J Neurosci 24:4889-93

Comoletti D, Flynn R, Jennings LL, Chubykin A, Matsumura T, Hasegawa H, Sudhof TC, Taylor P (2003) Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta. J Biol Chem 278:50497-505

Conroy J, Meally E, Kearney G, Fitzgerald M, Gill M, Gallagher L (2004) Serotonin transporter gene and autism: a haplotype analysis in an Irish autistic population. Mol Psychiatry 9:587-93

Cook EH, Jr., Courchesne R, Lord C, Cox NJ, Yan S, Lincoln A, Haas R, Courchesne E, Leventhal BL (1997) Evidence of linkage between the serotonin transporter and autistic disorder. Mol Psychiatry 2:247-50

Cook EH, Jr., Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ, Lincoln A, Nix K, Haas R, Leventhal BL, Courchesne E (1998) Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. Am J Hum Genet 62:1077-83

Cope NA, Hill G, van den Bree M, Harold D, Moskvina V, Green EK, Owen MJ, Williams J, O'Donovan MC (2005) No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. Mol Psychiatry 10:237-8

Coutinho AM, Oliveira G, Morgadinho T, Fesel C, Macedo TR, Bento C, Marques C, Ataide A, Miguel T, Borges L, Vicente AM (2004) Variants of the serotonin transporter gene (SLC6A4) significantly contribute to hyperserotonemia in autism. Mol Psychiatry 9:264-71

Curatolo P, Porfirio MC, Manzi B, Seri S (2004) Autism in tuberous sclerosis. Eur J Paediatr Neurol 8:327-32

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

Dasgupta B, Gutmann DH (2003) Neurofibromatosis 1: closing the GAP between mice and men. Curr Opin Genet Dev 13:20-7

Davis S, Weeks DE (1997) Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. Am J Hum Genet 61:1431-44

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, et al. (2002) A first-generation linkage disequilibrium map of human chromosome 22. Nature 418:544-8

Dean C, Scholl FG, Choih J, DeMaria S, Berger J, Isacoff E, Scheiffele P (2003) Neurexin mediates the assembly of presynaptic terminals. Nat Neurosci 6:708-16

DeLong GR, Dwyer JT (1988) Correlation of family history with specific autistic subgroups: Asperger's syndrome and bipolar affective disease. J Autism Dev Disord 18:593-600.

Demenais F, Kanninen T, Lindgren CM, Wiltshire S, Gaget S, Dandrieux C, Almgren P, Sjogren M, Hattersley A, Dina C, Tuomi T, McCarthy MI, Froguel P, Groop LC (2003) A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes. Hum Mol Genet 12:1865-73

Dempfle A, Loesgen S (2004) Meta-analysis of linkage studies for complex diseases: an overview of methods and a simulation study. Ann Hum Genet 68:69-83

Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, Rollins DY, Moses T, Sanders AR, Karkera JD, Esterling LE, Zeng J, Ferraro TN, Guroff JJ, Kazuba D, Maxwell ME, Nurnberger JI, Jr., Gershon ES (1999) A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. Proc Natl Acad Sci U S A 96:5604-9.

Devlin B, Bennett P, Dawson G, Figlewicz DA, Grigorenko EL, McMahon W, Minshew N, Pauls D, Smith M, Spence MA, Rodier PM, Stodgell C, Schellenberg GD (2004) Alleles of a reelin CGG repeat do not convey liability to autism in a sample from the CPEA network. Am J Med Genet B Neuropsychiatr Genet 126:46-50

Dresbach T, Neeb A, Meyer G, Gundelfinger ED, Brose N (2004) Synaptic targeting of neuroligin is independent of neurexin and SAP90/PSD95 binding. Mol Cell Neurosci 27:227-35

Durner M, Vieland VJ, Greenberg DA (1999) Further evidence for the increased power of LOD scores compared with nonparametric methods. Am J Hum Genet 64:281-9

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 25:320-3

Ehlers S, Gillberg C (1993) The epidemiology of Asperger syndrome. A total population study. J Child Psychol Psychiatry 34:1327-50.

Ehlers S, Gillberg C, Wing L (1999) A screening questionnaire for Asperger syndrome and other high- functioning autism spectrum disorders in school age children. J Autism Dev Disord 29:129-41.

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51 ( Pt 3):227-33

Farrer LA, Cupples LA (1998) Determining the genetic component of a disease. In: Haines J, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley-Liss, Inc., New York, pp 93-130

Fatemi SH (2002) The role of Reelin in pathology of autism. Mol Psychiatry 7:919-20

Fatemi SH (2004) Reelin glycoprotein: structure, biology and roles in health and disease. Mol Psychiatry

Feinstein C, Reiss AL (1998) Autism: the point of view from fragile X studies. J Autism Dev Disord 28:393-405

Ferguson JN, Young LJ, Hearn EF, Matzuk MM, Insel TR, Winslow JT (2000) Social amnesia in mice lacking the oxytocin gene. Nat Genet 25:284-8

Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168-70

Folstein S, Rutter M (1977a) Genetic influences and infantile autism. Nature 265:726-8

Folstein S, Rutter M (1977b) Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry 18:297-321

Folstein S, Santangelo S (2000) Does Asperger syndrome aggregate in families? In: Klin A, Volkmar F, Sparrow S (eds) Asperger Syndrome. The Guilford Press, New York, US, pp 159-171

Folstein SE, Mankoski RE (2000) Chromosome 7q: where autism meets language disorder? Am J Hum Genet 67:278-81

Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet 2:943-55

Fombonne E, Tidmarsh L (2003) Epidemiologic data on Asperger disorder. Child Adolesc Psychiatr Clin N Am 12:15-21, v-vi

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

Gauthier J, Bonnel A, St-Onge J, Karemera L, Laurent S, Mottron L, Fombonne E, Joober R, Rouleau GA (2004) NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. Am J Med Genet 132B:74-75.

Gauthier J, Joober R, Mottron L, Laurent S, Fuchs M, De Kimpe V, Rouleau GA (2003) Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. Am J Med Genet 118A:172-5

Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, Ducat L, Spence SJ (2001) The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. Am J Hum Genet 69:463-6

Gharani N, Benayed R, Mancuso V, Brzustowicz LM, Millonig JH (2004) Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. Mol Psychiatry 9:474-84

Ghaziuddin M (2005) A family history of Asperger syndrome. J Autism Dev Disord 35:177-182.

Ghaziuddin M, Weidmer-Mikhail E, Ghaziuddin N (1998) Comorbidity of Asperger syndrome: a preliminary report. J Intellect Disabil Res 42 ( Pt 4):279-83

Gillberg C (1989) Asperger syndrome in 23 Swedish children. Dev Med Child Neurol 31:520-31.

Gillberg C (1998) Chromosomal disorders and autism. J Autism Dev Disord 28:415-25

Gillberg C, Billstedt E (2000) Autism and Asperger syndrome: coexistence with other clinical disorders. Acta Psychiatr Scand 102:321-30.

Gillberg C, Rastam M, Wentz E (2001) The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new structured clinical interview. Autism 5:57-66

Gillberg IC, Gillberg C (1989) Asperger syndrome--some epidemiological considerations: a research note. J Child Psychol Psychiatry 30:631-8.

Gillberg IC, Gillberg C, Rastam M, Johansson M (1996) The cognitive profile of anorexia nervosa: a comparative study including a community-based sample. Compr Psychiatry 37:23-30

Gimpl G, Fahrenholz F (2001) The oxytocin receptor system: structure, function, and regulation. Physiol Rev 81:629-83

Glatt CE, Freimer NB (2002) Association analysis of candidate genes for neuropsychiatric disease: the perpetual campaign. Trends Genet 18:307-12

Gong X, Jia M, Ruan Y, Shuang M, Liu J, Wu S, Guo Y, Yang J, Ling Y, Yang X, Zhang D (2004) Association between the FOXP2 gene and autistic disorder in Chinese population. Am J Med Genet B Neuropsychiatr Genet 127:113-6

Göring HH, Terwilliger JD (2000) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet 66:1310-27

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

Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM (2004) Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. Cell 119:1013-26

Green L, Fein D, Modahl C, Feinstein C, Waterhouse L, Morris M (2001) Oxytocin and autistic disorder: alterations in peptide forms. Biol Psychiatry 50:609-13

Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet 63:870-9

Grigorenko EL (2003) The first candidate gene for dyslexia: Turning the page of a new chapter of research. Proc Natl Acad Sci U S A 100:11190-2

Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. Am J Hum Genet 68:661-73.

Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204-11

Hata Y, Davletov B, Petrenko AG, Jahn R, Sudhof TC (1993) Interaction of synaptotagmin with the cytoplasmic domains of neurexins. Neuron 10:307-15

Hattori E, Liu C, Badner JA, Bonner TI, Christian SL, Maheshwari M, Detera-Wadleigh SD, Gibbs RA, Gershon ES (2003) Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. Am J Hum Genet 72:1131-40

Hennah W, Varilo T, Paunio T, Peltonen L (2004) Haplotype analysis and identification of genes for a complex trait: examples from schizophrenia. Ann Med 36:322-31

Hirao K, Hata Y, Ide N, Takeuchi M, Irie M, Yao I, Deguchi M, Toyoda A, Sudhof TC, Takai Y (1998) A novel multiple PDZ domain-containing molecule interacting with N-methyl-D-aspartate receptors and neuronal cell adhesion proteins. J Biol Chem 273:21105-10

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95-108

Hollander E, Novotny S, Hanratty M, Yaffe R, DeCaria CM, Aronowitz BR, Mosovich S (2003a) Oxytocin infusion reduces repetitive behaviors in adults with autistic and Asperger's disorders. Neuropsychopharmacology 28:193-8

Hollander E, Phillips AT, Yeh CC (2003b) Targeted treatments for symptom domains in child and adolescent autism. Lancet 362:732-4

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9:301-6

Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajarvi R, Kokko-Sahin ML, Ekelund J, Lonnqvist J, Peltonen L (1998) Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. Mol Psychiatry 3:452-7

Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, Sudhof TC (1995) Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell 81:435-43

Ichtchenko K, Nguyen T, Sudhof TC (1996) Structures, alternative splicing, and neurexin binding of multiple neuroligins. J Biol Chem 271:2676-82

IMGSAC (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. Hum Mol Genet 7:571-8.

IMGSAC (2001a) Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q. Hum Mol Genet 10:973-82.

IMGSAC (2001b) A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. Am J Hum Genet 69:570-81.

Insel TR, Young LJ (2001) The neurobiology of attachment. Nat Rev Neurosci 2:129-36

Irie M, Hata Y, Takeuchi M, Ichtchenko K, Toyoda A, Hirao K, Takai Y, Rosahl TW, Sudhof TC (1997) Binding of neuroligins to PSD-95. Science 277:1511-5

Ishikawa-Brush Y, Powell JF, Bolton P, Miller AP, Francis F, Willard HF, Lehrach H, Monaco AP (1997) Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3' to the SDC2 gene. Hum Mol Genet 6:1241-50

Jamain S, Betancur C, Quach H, Philippe A, Fellous M, Giros B, Gillberg C, Leboyer M, Bourgeron T (2002) Linkage and association of the glutamate receptor 6 gene with autism. Mol Psychiatry 7:302-10

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34:27-9

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29:217-22

Jiang Y, Lev-Lehman E, Bressler J, Tsai TF, Beaudet AL (1999) Genetics of Angelman syndrome. Am J Hum Genet 65:1-6

Jin P, Alisch RS, Warren ST (2004) RNA and microRNAs in fragile X mental retardation. Nat Cell Biol 6:1048-53

Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, McMahon WM, Petersen B, Jenson WR, Mo A (1991) Complex segregation analysis of autism. Am J Hum Genet 49:932-8

Jorde LB, Mason-Brothers A, Waldmann R, Ritvo ER, Freeman BJ, Pingree C, McMahon WM, Petersen B, Jenson WR, Mo A (1990) The UCLA-University of Utah epidemiologic survey of autism: genealogical analysis of familial aggregation. Am J Med Genet 36:85-8

Kaessmann H, Zollner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlen M, Paabo S (2002) Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 70:673-85

Kanner L (1943) Autistic disturbances of affective contact. Nervous Child 2:217-250

Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA, Flodman P, Khristich J, Mroczkowski-Parker Z, Brown JL, Masser D, Ungerleider S, Rapaport MH, Wishart WL, Luebbert H (2001) A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. Proc Natl Acad Sci U S A 98:585-90.

Kielinen M, Linna SL, Moilanen I (2000) Autism in Northern Finland. Eur Child Adolesc Psychiatry 9:162-7

Kim SJ, Cox N, Courchesne R, Lord C, Corsello C, Akshoomoff N, Guter S, Leventhal BL, Courchesne E, Cook EH, Jr. (2002a) Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. Mol Psychiatry 7:278-88

Kim SJ, Herzing LB, Veenstra-VanderWeele J, Lord C, Courchesne R, Leventhal BL, Ledbetter DH, Courchesne E, Cook EH, Jr. (2002b) Mutation screening and transmission disequilibrium study of ATP10C in autism. Am J Med Genet 114:137-43

Klauck SM, Poustka F, Benner A, Lesch KP, Poustka A (1997) Serotonin transporter (5-HTT) gene variants associated with autism? Hum Mol Genet 6:2233-8

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241-7

Krebs MO, Betancur C, Leroy S, Bourdel MC, Gillberg C, Leboyer M (2002) Absence of association between a polymorphic GGC repeat in the 5' untranslated region of the reelin gene and autism. Mol Psychiatry 7:801-4

Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17:21-4

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

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-63.

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

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435-8

Lai CS, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED, Vargha-Khadem F, Monaco AP (2000) The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am J Hum Genet 67:357-68

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-23

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241-7.

Lander ES (1996) The new genomics: global views of biology. Science 274:536-9

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

Larsson HJ, Eaton WW, Madsen KM, Vestergaard M, Olesen AV, Agerbo E, Schendel D, Thorsen P, Mortensen PB (2005) Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. Am J Epidemiology 161: 916-925.

Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lemonnier E, Calvas P, Laudier B, Chelly J, Fryns JP, Ropers HH, Hamel BC, Andres C, Barthelemy C, Moraine C, Briault S (2004) X-Linked Mental Retardation and Autism Are Associated with a Mutation in the NLGN4 Gene, a Member of the Neuroligin Family. Am J Hum Genet 74:552-7

Leboyer M, Philippe A, Bouvard M, Guilloud-Bataille M, Bondoux D, Tabuteau F, Feingold J, Mouren-Simeoni MC, Launay JM (1999) Whole blood serotonin and plasma beta-endorphin in autistic probands and their first-degree relatives. Biol Psychiatry 45:158-63

Leventhal BL, Cook EH, Jr., Morford M, Ravitz A, Freedman DX (1990) Relationships of whole blood serotonin and plasma norepinephrine within families. J Autism Dev Disord 20:499-511

Li J, Nguyen L, Gleason C, Lotspeich L, Spiker D, Risch N, Myers RM (2004) Lack of evidence for an association between WNT2 and RELN polymorphisms and autism. Am J Med Genet B Neuropsychiatr Genet 126:51-7

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397-401

Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D, Lord C, Iversen P, Hoh J, Ott J, Gilliam TC (2001) A genomewide screen for autism susceptibility loci. Am J Hum Genet 69:327-40.

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

Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659-85

Maestrini E, Lai C, Marlow A, Matthews N, Wallace S, Bailey A, Cook EH, Weeks DE, Monaco AP (1999) Serotonin transporter (5-HTT) and gamma-aminobutyric

acid receptor subunit beta3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. The International Molecular Genetic Study of Autism Consortium. Am J Med Genet 88:492-6

Marino C, Giorda R, Luisa Lorusso M, Vanzin L, Salandi N, Nobile M, Citterio A, Beri S, Crespi V, Battaglia M, Molteni M (2005) A family-based association study does not support DYX1C1 on 15q21.3 as a candidate gene in developmental dyslexia. Eur J Hum Genet 13:491-499

Martin ER, Menold MM, Wolpert CM, Bass MP, Donnelly SL, Ravan SA, Zimmerman A, Gilbert JR, Vance JM, Maddox LO, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (2000) Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. Am J Med Genet 96:43-8

McCauley JL, Olson LM, Delahanty R, Amin T, Nurmi EL, Organ EL, Jacobs MM, Folstein SE, Haines JL, Sutcliffe JS (2004a) A linkage disequilibrium map of the 1-Mb 15q12 GABA(A) receptor subunit cluster and association to autism. Am J Med Genet 131B:51-9

McCauley JL, Olson LM, Dowd M, Amin T, Steele A, Blakely RD, Folstein SE, Haines JL, Sutcliffe JS (2004b) Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. Am J Med Genet B Neuropsychiatr Genet 127:104-12

Menold MM, Shao Y, Wolpert CM, Donnelly SL, Raiford KL, Martin ER, Ravan SA, Abramson RK, Wright HH, Delong GR, Cuccaro ML, Pericak-Vance MA, Gilbert JR (2001) Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder. J Neurogenet 15:245-59

Meyer G, Varoqueaux F, Neeb A, Oschlies M, Brose N (2004) The complexity of PDZ domain-mediated interactions at glutamatergic synapses: a case study on neuroligin. Neuropharmacology 47:724-33

Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 9:1415-23

Miller W, Makova KD, Nekrutenko A, Hardison RC (2004) Comparative genomics. Annu Rev Genomics Hum Genet 5:15-56

Modahl C, Green L, Fein D, Morris M, Waterhouse L, Feinstein C, Levin H (1998) Plasma oxytocin levels in autistic children. Biol Psychiatry 43:270-7

Morton NE (1955) Sequental tests for the detection of linkage. Am J Hum Genet 7:277-318

Muhle R, Trentacoste SV, Rapin I (2004) The genetics of autism. Pediatrics 113:e472-86

Mulder EJ, Anderson GM, Kema IP, Brugman AM, Ketelaars CE, de Bildt A, van Lang ND, den Boer JA, Minderaa RB (2005) Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. Am J Med Genet B Neuropsychiatr Genet 133B:93-96

Munafo MR, Flint J (2004) Meta-analysis of genetic association studies. Trends Genet 20:439-44

Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CS, Baird G, Jannoun L, Slonims V, Stott CM, Merricks MJ, Bolton PF, Bailey AJ, Monaco AP (2002) FOXP2 is not a major susceptibility gene for autism or specific language impairment. Am J Hum Genet 70:1318-27

Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. Annu Rev Genomics Hum Genet 2:153-75

Nieminen-Von Wendt T (2004) On the origin and diagnosis of Asperger syndrome: A clinical, neuroimaging and genetic study. University of Helsinki, 94p.

Nieminen-von Wendt T, Paavonen JE, Ylisaukko-oja T, Sarenius S, Kallman T, Jarvela I, von Wendt L (2005) Subjective face recognition difficulties, aberrant sensibility, sleeping disturbances and aberrant eating habits in families with Asperger syndrome. BMC Psychiatry 5:20

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67-86

Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, Peltonen L (1995) Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. Am J Hum Genet 56:1088-95

Norio R (2003a) Finnish Disease Heritage I: characteristics, causes, background. Hum Genet 112:441-56

Norio R (2003b) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112:457-69

Nurmi EL, Amin T, Olson LM, Jacobs MM, McCauley JL, Lam AY, Organ EL, Folstein SE, Haines JL, Sutcliffe JS (2003a) Dense linkage disequilibrium mapping in the 15q11-q13 maternal expression domain yields evidence for association in autism. Mol Psychiatry 8:624-34, 570

Nurmi EL, Bradford Y, Chen Y, Hall J, Arnone B, Gardiner MB, Hutcheson HB, Gilbert JR, Pericak-Vance MA, Copeland-Yates SA, Michaelis RC, Wassink TH, Santangelo SL, Sheffield VC, Piven J, Folstein SE, Haines JL, Sutcliffe JS (2001) Linkage disequilibrium at the Angelman syndrome gene UBE3A in autism families. Genomics 77:105-13

Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, Sutcliffe JS (2003b) Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. J Am Acad Child Adolesc Psychiatry 42:856-63

Ott J (1999) Analysis of Human Genetic Linkage. The Johns Hopkins University Press, Baltimore

Ott J (2004) Association of genetic loci: Replication or not, that is the question. Neurology 63:955-8

Owen MJ, Holmans P, McGuffin P (1997) Association studies in psychiatric genetics. Mol Psychiatry 2:270-3

Pajukanta P, Allayee H, Krass KL, Kuraishy A, Soro A, Lilja HE, Mar R, Taskinen MR, Nuotio I, Laakso M, Rotter JI, de Bruin TW, Cantor RM, Lusis AJ, Peltonen L (2003) Combined analysis of genome scans of dutch and finnish families

reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. Am J Hum Genet 72:903-17

Pan D, Dong J, Zhang Y, Gao X (2004) Tuberous sclerosis complex: from Drosophila to human disease. Trends Cell Biol 14:78-85

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719-23

Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. Hum Mol Genet 8:1913-23

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. Nat Rev Genet 1:182-90

Persico AM, D'Agruma L, Maiorano N, Totaro A, Militerni R, Bravaccio C, Wassink TH, Schneider C, Melmed R, Trillo S, Montecchi F, Palermo M, Pascucci T, Puglisi-Allegra S, Reichelt KL, Conciatori M, Marino R, Quattrocchi CC, Baldi A, Zelante L, Gasparini P, Keller F (2001) Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. Mol Psychiatry 6:150-9

Persico AM, Militerni R, Bravaccio C, Schneider C, Melmed R, Conciatori M, Damiani V, Baldi A, Keller F (2000) Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples. Am J Med Genet 96:123-7

Persico AM, Pascucci T, Puglisi-Allegra S, Militerni R, Bravaccio C, Schneider C, Melmed R, Trillo S, Montecchi F, Palermo M, Rabinowitz D, Reichelt KL, Conciatori M, Marino R, Keller F (2002) Serotonin transporter gene promoter variants do not explain the hyperserotoninemia in autistic children. Mol Psychiatry 7:795-800

Petit E, Herault J, Martineau J, Perrot A, Barthelemy C, Hameury L, Sauvage D, Lelord G, Muh JP (1995) Association study with two markers of a human homeogene in infantile autism. J Med Genet 32:269-74

Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M, van Malldergerme L (1999) Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. Hum Mol Genet 8:805-12.

Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, et al. (2003) Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. Nat Genet 33:382-7

Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, Rutter M (1995) Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. Am J Hum Genet 57:717-26

Piven J, Tsai GC, Nehme E, Coyle JT, Chase GA, Folstein SE (1991) Platelet serotonin, a possible marker for familial autism. J Autism Dev Disord 21:51-9

Prange O, Wong TP, Gerrow K, Wang YT, El-Husseini A (2004) A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin. Proc Natl Acad Sci U S A 101:13915-20

Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124-37

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

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

Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19:149-50

Ramoz N, Reichert JG, Smith CJ, Silverman JM, Bespalova IN, Davis KL, Buxbaum JD (2004) Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. Am J Psychiatry 161:662-9

Rehnström K, Ylisaukko-oja T, Nieminen-von Wendt T, Sarenius S, Källman T, Kempas E, von Wendt L, Peltonen L, Järvelä I. Independent replication and initial fine mapping of 3q21-24 in Asperger syndrome. J Med Genet, in press.

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

Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17:502-10

Ringman JM, Jankovic J (2000) Occurrence of tics in Asperger's syndrome and autistic disorder. J Child Neurol 15:394-400

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222-8

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

Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, et al. (1999) A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65:493-507.

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

Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry 142:74-7

Ritvo ER, Mason-Brothers A, Menkes JH, Sparkes RS (1988) Association of autism, retinoblastoma, and reduced esterase D activity. Arch Gen Psychiatry 45:600

Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. Am J Hum Genet 65:876-84

Rogers SJ, Wehner DE, Hagerman R (2001) The behavioral phenotype in fragile X: symptoms of autism in very young children with fragile X syndrome, idiopathic autism, and other developmental disorders. J Dev Behav Pediatr 22:409-17

Sabaratnam M, Murthy NV, Wijeratne A, Buckingham A, Payne S (2003) Autistic-like behaviour profile and psychiatric morbidity in Fragile X Syndrome: a prospective ten-year follow-up study. Eur Child Adolesc Psychiatry 12:172-7

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-33

Salmon B, Hallmayer J, Rogers T, Kalaydjieva L, Petersen PB, Nicholas P, Pingree C, McMahon W, Spiker D, Lotspeich L, Kraemer H, McCague P, Dimiceli S, Nouri N, Pitts T, Yang J, Hinds D, Myers RM, Risch N (1999) Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. Am J Med Genet 88:551-6

Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D (1997) Empirical genomewide significance levels established by whole genome simulations. Genet Epidemiol 14:223-9

Scerri TS, Fisher SE, Francks C, MacPhie IL, Paracchini S, Richardson AJ, Stein JF, Monaco AP (2004) Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. J Med Genet 41:853-7

Schaid DJ, Guenther JC, Christensen GB, Hebbring S, Rosenow C, Hilker CA, McDonnell SK, Cunningham JM, Slager SL, Blute ML, Thibodeau SN (2004) Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility Loci. Am J Hum Genet 75:948-65

Scheiffele P, Fan J, Choih J, Fetter R, Serafini T (2000) Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. Cell 101:657-69

Shao Y, Cuccaro ML, Hauser ER, Raiford KL, Menold MM, Wolpert CM, Ravan SA, Elston L, Decena K, Donnelly SL, Abramson RK, Wright HH, DeLong GR, Gilbert JR, Pericak-Vance MA (2003) Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. Am J Hum Genet 72:539-48

Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, Ashley-Koch AA, Abramson RK, Wright HH, DeLong RG, Gilbert JR, Cuccaro ML, Pericak-Vance MA (2002a) Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. Am J Hum Genet 70:1058-61.

Shao Y, Wolpert CM, Raiford KL, Menold MM, Donnelly SL, Ravan SA, Bass MP, McClain C, von Wendt L, Vance JM, Abramson RH, Wright HH, Ashley-Koch A, Gilbert JR, DeLong RG, Cuccaro ML, Pericak-Vance MA (2002b) Genomic screen and follow-up analysis for autistic disorder. Am J Med Genet 114:99-105.

Shaywitz SE (1998) Dyslexia. N Engl J Med 338:307-12

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, Sunden SL, et al. (1995) A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837-44

Shemer R, Hershko AY, Perk J, Mostoslavsky R, Tsuberi B, Cedar H, Buiting K, Razin A (2000) The imprinting box of the Prader-Willi/Angelman syndrome domain. Nat Genet 26:440-3

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

Shuang M, Liu J, Jia MX, Yang JZ, Wu SP, Gong XH, Ling YS, Ruan Y, Yang XL, Zhang D (2004) Family-based association study between autism and glutamate receptor 6 gene in Chinese Han trios. Am J Med Genet B Neuropsychiatr Genet 131:48-50

Skaar DA, Shao Y, Haines JL, Stenger JE, Jaworski J, Martin ER, Delong GR, Moore JH, McCauley JL, Sutcliffe JS, Ashley-Koch AE, Cuccaro ML, Folstein SE, Gilbert JR, Pericak-Vance MA (2004) Analysis of the RELN gene as a genetic risk factor for autism. Mol Psychiatry

Skuse DH (2000) Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. Pediatr Res 47:9-16

Skuse DH, James RS, Bishop DV, Coppin B, Dalton P, Aamodt-Leeper G, Bacarese-Hamilton M, Creswell C, McGurk R, Jacobs PA (1997) Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. Nature 387:705-8

Smalley SL (1998) Autism and tuberous sclerosis. J Autism Dev Disord 28:407-14

Smith IM (2000) Motor functioning in Asperger syndrome. In: Klin A VF, Sparrow SS (ed) Asperger Syndrome. The Guilford Press, New York, pp 97-124

Smith M, Woodroffe A, Smith R, Holguin S, Martinez J, Filipek PA, Modahl C, Moore B, Bocian ME, Mays L, Laulhere T, Flodman P, Spence MA (2002) Molecular genetic delineation of a deletion of chromosome 13q12-->q13 in a patient with autism and auditory processing deficits. Cytogenet Genome Res 98:233-9

Smith RF, Wiese BA, Wojzynski MK, Davison DB, Worley KC (1996) BCM Search Launcher--an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. Genome Res 6:454-62

Song JY, Ichtchenko K, Sudhof TC, Brose N (1999) Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. Proc Natl Acad Sci U S A 96:1100-5

Speer MC (1998) Basic concepts in genetics. In: Haines J, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley-Liss, Inc., New York, pp 17-52

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983-9

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

Spielman RS, McGinnis RE, Ewens WJ (1994) The transmission/disequilibrium test detects cosegregation and linkage. Am J Hum Genet 54:559-60; author reply 560-3

Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, Napolitano C, Schwartz PJ, Joseph RM, Condouris K, Tager-Flusberg H, Priori SG, Sanguinetti MC, Keating MT (2004) Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell 119:19-31

Ssucharewa GE, Wolff S (1996) The first account of the syndrome Asperger described? Translation of a paper entitled "Die schizoiden Psychopathien im Kindesalter" by Dr. G.E. Ssucharewa; scientific assistant, which appeared in 1926 in the Monatsschrift fur Psychiatrie und Neurologie 60:235-261. Eur Child Adolesc Psychiatry 5:119-32

Steele MM, Al-Adeimi M, Siu VM, Fan YS (2001) Brief report: A case of autism with interstitial deletion of chromosome 13. J Autism Dev Disord 31:231-4

Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, Bohman M (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry 30:405-16.

Steffenburg S, Gillberg CL, Steffenburg U, Kyllerman M (1996) Autism in Angelman syndrome: a population-based study. Pediatr Neurol 14:131-6

Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH, Nelson SF (2004) Evidence for sex-specific risk alleles in autism spectrum disorder. Am J Hum Genet 75:1117-23

Strachan T, Read AP (2004) Human Molecular Genetics 3. Garland Publishing, New York

Sultana R, Yu CE, Yu J, Munson J, Chen D, Hua W, Estes A, Cortes F, de la Barra F, Yu D, Haider ST, Trask BJ, Green ED, Raskind WH, Disteche CM, Wijsman E, Dawson G, Storm DR, Schellenberg GD, Villacres EC (2002) Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics 80:129-34

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

Szatmari P, Bartolucci G, Bremner R (1989) Asperger's syndrome and autism: comparison of early history and outcome. Dev Med Child Neurol 31:709-20.

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324-8

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K, Kere J (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proc Natl Acad Sci U S A 100:11553-8

Tani P, Lindberg N, Nieminen-von Wendt T, von Wendt L, Alanko L, Appelberg B, Porkka-Heiskanen T (2003) Insomnia is a frequent finding in adults with Asperger syndrome. BMC Psychiatry 3:12

Tantam D (1988) Asperger's syndrome. J Child Psychol Psychiatry 29:245-55

Tentler D, Brandberg G, Betancur C, Gillberg C, Anneren G, Orsmark C, Green ED, Carlsson B, Dahl N (2001) A balanced reciprocal translocation t(5;7)(q14;q32) associated with autistic disorder: molecular analysis of the chromosome 7 breakpoint. Am J Med Genet 105:729-36

Terwilliger J, Ott J (1994) Handbook of human genetic linkage. The John Hopkins University Press, Baltimore

Terwilliger JD, Goring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 72:63-132.

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

The European Chromosome 16 Tuberous Sclerosis Consortium (1993) Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75:1305-15

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

Thomas NS, Sharp AJ, Browne CE, Skuse D, Hardie C, Dennis NR (1999) Xp deletions associated with autism in three females. Hum Genet 104:43-8

Tierney E, Nwokoro NA, Kelley RI (2000) Behavioral phenotype of RSH/Smith-Lemli-Opitz syndrome. Ment Retard Dev Disabil Res Rev 6:131-4

Tordjman S, Gutknecht L, Carlier M, Spitz E, Antoine C, Slama F, Carsalade V, Cohen DJ, Ferrari P, Roubertoux PL, Anderson GM (2001) Role of the serotonin transporter gene in the behavioral expression of autism. Mol Psychiatry 6:434-9

van den Oord EJ, Neale BM (2004) Will haplotype maps be useful for finding genes? Mol Psychiatry 9:227-36

van Slegtenhorst M, de Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, Lindhout D, et al. (1997) Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. Science 277:805-8

Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L (2000) Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. Eur J Hum Genet 8:604-12

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

Varoqueaux F, Jamain S, Brose N (2004) Neuroligin 2 is exclusively localized to inhibitory synapses. Eur J Cell Biol 83:449-56

Veenstra-VanderWeele J, Cook EH, Jr. (2004) Molecular genetics of autism spectrum disorder. Mol Psychiatry 9:819-32

Veenstra-VanderWeele J, Gonen D, Leventhal BL, Cook EH, Jr. (1999) Mutation screening of the UBE3A/E6-AP gene in autistic disorder. Mol Psychiatry 4:64-7

Veltman MW, Thompson RJ, Roberts SE, Thomas NS, Whittington J, Bolton PF (2004) Prader-Willi syndrome--a study comparing deletion and uniparental disomy cases with reference to autism spectrum disorders. Eur Child Adolesc Psychiatry 13:42-50

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. Science 291:1304-51

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905-14

Vincent JB, Herbrick JA, Gurling HM, Bolton PF, Roberts W, Scherer SW (2000) Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. Am J Hum Genet 67:510-4

Vincent JB, Kolozsvari D, Roberts WS, Bolton PF, Gurling HM, Scherer SW (2004) Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. Am J Med Genet 129B:82-4

Volkmar F, Klin A (2000) Diagnostic issues in Asperger syndrome. In: Klin A, Volkmar F, Sparrow S (eds) Asperger Syndrome. The Guilford Press, New York, US, pp 25-71

Volkmar FR, Klin A, Pauls D (1998) Nosological and genetic aspects of Asperger syndrome. J Autism Dev Disord 28:457-63.

Volkmar FR, Klin A, Schultz R, Bronen R, Marans WD, Sparrow S, Cohen DJ (1996) Asperger's syndrome. J Am Acad Child Adolesc Psychiatry 35:118-23.

Volkmar FR, Cohen DJ (1991) Comorbid association of autism and schizophrenia. Am J Psychiatry 148: 1705-1707.

Wallace RH, Marini C, Petrou S, Harkin LA, Bowser DN, Panchal RG, Williams DA, Sutherland GR, Mulley JC, Scheffer IE, Berkovic SF (2001) Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet 28:49-52

Wang N, Akey JM, Zhang K, Chakraborty R, Jin L (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet 71:1227-34

Warburton P, Baird G, Chen W, Morris K, Jacobs BW, Hodgson S, Docherty Z (2000) Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. Am J Med Genet 96:228-34

Wassink TH, Piven J, Patil SR (2001) Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. Psychiatr Genet 11:57-63

Wassink TH, Piven J, Vieland VJ, Pietila J, Goedken RJ, Folstein SE, Sheffield VC (2002) Evaluation of FOXP2 as an autism susceptibility gene. Am J Med Genet 114:566-9

Weaving LS, Christodoulou J, Williamson SL, Friend KL, McKenzie OL, Archer H, Evans J, Clarke A, Pelka GJ, Tam PP, Watson C, Lahooti H, Ellaway CJ, Bennetts B, Leonard H, Gecz J (2004) Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet 75:1079-93

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388-96

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

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

Wigg KG, Couto JM, Feng Y, Anderson B, Cate-Carter TD, Macciardi F, Tannock R, Lovett MW, Humphries TW, Barr CL (2004) Support for EKN1 as the susceptibility locus for dyslexia on 15q21. Mol Psychiatry 9:1111-21

Williams PG, Hersh JH (1998) Brief report: the association of neurofibromatosis type 1 and autism. J Autism Dev Disord 28:567-71

Wing L (1981) Asperger's syndrome: a clinical account. Psychol Med 11:115-29.

Winslow JT, Insel TR (2002) The social deficits of the oxytocin knockout mouse. Neuropeptides 36:221-9

Wise LH, Lanchbury JS, Lewis CM (1999) Meta-analysis of genome searches. Ann Hum Genet 63 ( Pt 3):263-72

Wiznitzer M (2004) Autism and tuberous sclerosis. J Child Neurol 19:675-9

Wolff S (2000) Schizoid personality in childhood and Asperger syndrome. In: Klin A, Volkmar F, Sparrow S (eds) Asperger Syndrome. The Guilford Press, New York, US, pp 278-305

World Health Organization (1993) International classification of diseases: Tenth revision. Chapter V. Mental and behavioural diseases. Diagnostic criteria for research. Author, Geneva

Xu J, Meyers DA, Pericak-Vance MA (1998) Lod score analysis. In: Haines J, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley Liss, Inc., New York, pp 253-272

Yan J, Oliveira G, Coutinho A, Yang C, Feng J, Katz C, Sram J, Bockholt A, Jones IR, Craddock N, Cook EH, Vicente A, Sommer SS (2005) Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychaitric patients. Mol Psychaitry 10:329-332.

Yeargin-Allsopp M, Rice C, Karapurkar T, Doernberg N, Boyle C, Murphy C (2003) Prevalence of autism in a US metropolitan area. Jama 289:49-55

Yirmiya N, Pilowsky T, Nemanov L, Arbelle S, Feinsilver T, Fried I, Ebstein RP (2001) Evidence for an association with the serotonin transporter promoter region polymorphism and autism. Am J Med Genet 105:381-6

Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA, Grunn A, Juo SH, Terwilliger JD, Liu J, Cantor RM, Geschwind DH, Gilliam TC (2003) A genomewide screen of 345 families for autism-susceptibility loci. Am J Hum Genet 73:886-97

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of human genetic and sequence-based physical maps. Nature 409:951-3

Zhang H, Liu X, Zhang C, Mundo E, Macciardi F, Grayson DR, Guidotti AR, Holden JJ (2002) Reelin gene alleles and susceptibility to autism spectrum disorders. Mol Psychiatry 7:1012-7

Zhang K, Akey JM, Wang N, Xiong M, Chakraborty R, Jin L (2003) Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. Hum Genet 113:51-9

Zhong H, Serajee FJ, Nabi R, Huq AH (2003) No association between the EN2 gene and autistic disorder. J Med Genet 40:e4

Zhong N, Ye L, Ju W, Brown WT, Tsiouris J, Cohen I (1999) 5-HTTLPR variants not associated with autistic spectrum disorders. Neurogenetics 2:129-31