

2000

A linkage study of autism using multipoint sib-pair analysis

Tamara Rogers
Edith Cowan University

Follow this and additional works at: <https://ro.ecu.edu.au/theses>



Part of the [Genetic Phenomena Commons](#)

Recommended Citation

Rogers, T. (2000). *A linkage study of autism using multipoint sib-pair analysis*. Edith Cowan University.
Retrieved from <https://ro.ecu.edu.au/theses/1536>

This Thesis is posted at Research Online.
<https://ro.ecu.edu.au/theses/1536>

Edith Cowan University

Copyright Warning

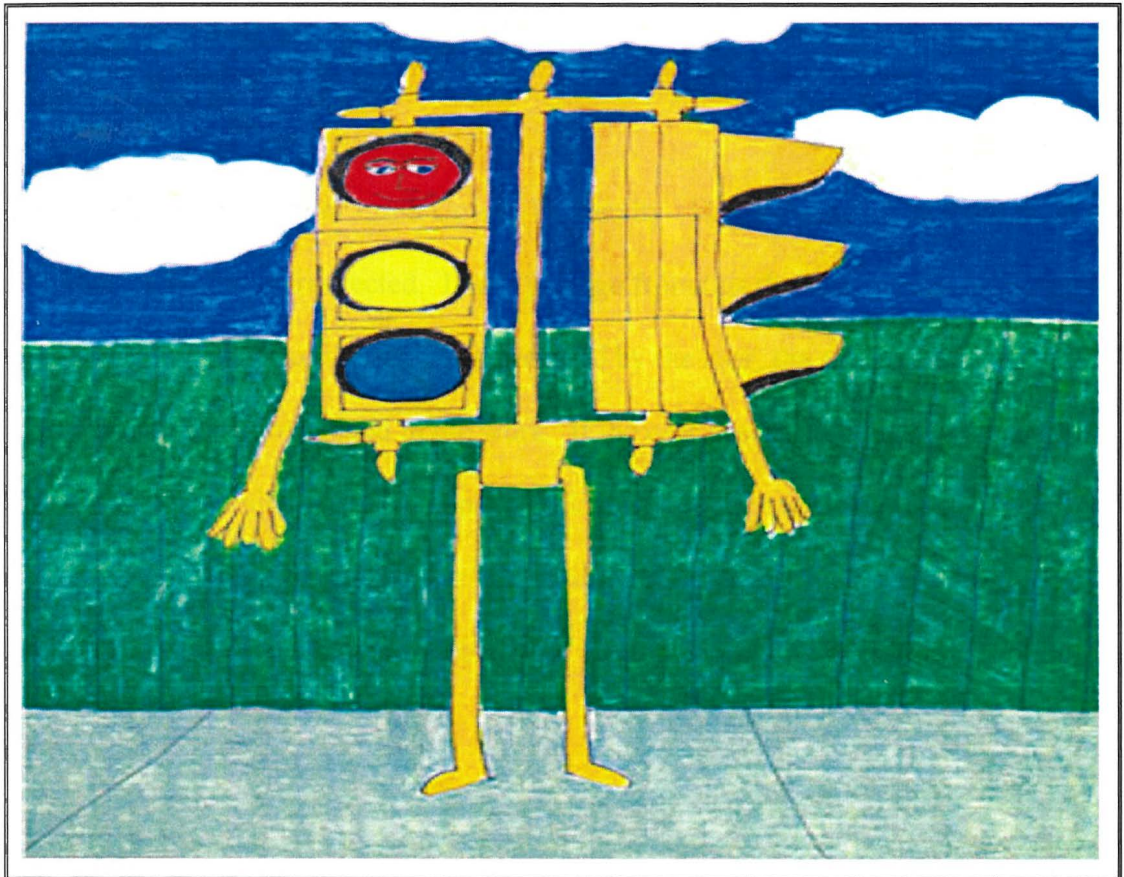
You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

A LINKAGE STUDY OF AUTISM USING MULTIPOINT SIB-PAIR ANALYSIS



("I am a traffic light" by B Moran)

Tamara Rogers

**EDITH COWAN UNIVERSITY
LIBRARY**

(Bachelor of Science in Biotechnology with Honours in Veterinary Biology)

This thesis is presented for the Doctorate in Philosophy in Human Biology at Edith Cowan University

Supervisors: Associate Professor L. Kalaydjieva (Human Biology, Edith Cowan University, Joondalup)

Associate Professor J. Hallmayer (Graylands Hospital / University of Western Australia)

Edith Cowan University, Western Australia, October 2000.

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

ABSTRACT

Autism is a severe developmental disorder that was first described by Kanner in 1943. It is characterised by four major criteria: marked social deficits, delay in language development, a restricted range of stereotyped repetitive behaviours and onset of the disease within the first three years of life.

The last decade of research has provided support for a strong genetic basis in the aetiology of autism. Firstly, a number of genetic conditions, such as fragile X syndrome, chromosome 15 anomalies and tuberous sclerosis, have been associated with autism. Secondly, family studies have demonstrated that the recurrence risk for autism among siblings of autistic children is 50-200 times higher compared to the risk of autism in the general population (4-5 in 10 000 births). Thirdly, twin studies show a significantly higher concordance rate among monozygotic twins compared to dizygotic twins.

To explore the theory of a biological basis of autism, molecular genetic strategies were employed in this study. Two collaborative groups (Stanford University in the U.S.A. and Edith Cowan University/Graylands Hospital in Perth, Australia) scanned the entire human genome for autism susceptibility genes in 90 American multiplex families, making this the largest genome screen in autism to date. Candidate regions were also run in an additional group of 41 Australian multiplex families. One hundred of the total of 519 markers were analysed as part of this PhD project. Since autism is most likely a complex disorder involving interacting genes with unknown modes of inheritance, non-parametric linkage analysis utilising the multipoint sib-pair approach is the best method presently available for identifying these susceptibility genes.

Using this method and given the power of the study to reliably detect genes of moderate effect, findings from this PhD project and the overall genome scan provide substantial evidence for the absence of such genes in autism. Regions that may, however, contain these disease-predisposing loci were identified. These include regions on chromosomes 1p and 22q. Results from other smaller genome scanning studies also indicate these two candidate regions as possibly containing loci involved in the aetiology of autism. Our results could also not exclude chromosomes 7q, 9q, 10q, 11p and 17, regions that have previously been suggested by other studies to contain susceptibility loci. Overall, our data were most consistent with a model of ≥ 15 susceptibility loci, each with a small effect on autism.

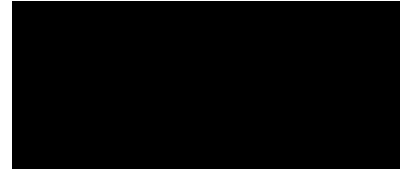
These results are important for future studies. In particular, they have shown the requirement for a large sample size to detect genes of small effect contributing to a disorder. Whilst it may be difficult for individual research groups to recruit an adequate sample size, collaboration with other groups may provide sufficient power to detect these susceptibility loci. By combining our results with those from other linkage studies on autism, we may also be able to eliminate some genetic models for this disease and provide an insight into the boundaries of the autistic phenotype.

With advances in the Human Genome Project and the development of rigorous statistical approaches to the analysis of multifactorial diseases, the molecular characterisation of genes that have a small effect on the overall aetiology of complex polygenic disorders, such as autism, could soon be feasible.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

- (i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;
- (ii) contain any material previously published or written by another person except where due reference is made in the text; or
- (iii) contain any defamatory material.



Tamara Rogers

(BScHons)

ACKNOWLEDGEMENTS

I wish to make the following acknowledgements for assistance given during this research:

To the National Institute of Mental Health, National Institutes of Health of the U.S.A., the National Health and Medical Research Council of Australia and the APEX Foundation for supplying the funding for this project.

To my supervisors, Joachim Hallmayer and Luba Kalaydjieva, for their advice, encouragement and guidance.

To Wayne Hill, Amelia Nikolova and the U.S.A. collaborators, for providing the clinical data on the autistic individuals.

To Boyd Salmon, for his assistance in establishing the cell lines and extracting the DNA from the subjects used in this study.

To David Chandler, for his technical assistance and advice, as well as for his ongoing encouragement.

A special thank you to the families who participated in this study.

Finally, to my family and friends for their support and continual help throughout the research.

TABLE OF ABBREVIATIONS

Adenosine-tri-radioactive-phosphate	P ³² - ATP
American Psychiatric Association	APA
Applied Biosystems Inc.	ABI
Autism Diagnostic Interview	ADI
Autism Diagnostic Observation Schedule	ADOS
Base pair	bp
Centimorgan	cM
The Centre d'Étude du Polymorphisme Humain	CEPH
Deoxyribonucleic acid	DNA
Diagnostic and Statistical Manual of Mental Disorders	DSM
Distilled water	dH ₂ O
Ethylenediaminetetraacetic acid	EDTA
Expected Maximum Lod Score	EMLS
Greater than	>
Identity-by-descent	IBD
Intelligence Test Scores	IQ
International Classification of Diseases	ICD
Less than	<
Logarithm of odds	Lod
Magnesium chloride	MgCl
Maximum Lod Score	MLS
Millicurie	mCi

Most-likely Lod Score	mLOD
Microlitre (10^{-6})	μ l
Millilitre (10^{-3})	ml
Millimolar (10^{-3})	mM
Nanometre (10^{-9})	nm
National Institutes of Health	NIH
Neurofibromatosis	NF
Optical Density	OD
Photomultiplier Tube	PMT
Polymerase Chain Reaction	PCR
Polymorphic Information Content	PIC
Restriction Fragment Length Polymorphism	RFLP
Revolutions per minute	rpm
Ribonucleic acid	RNA
Room temperature	RT
Solution	Soln.
Standard Deviation	SD
Tris-acetate Electrophoresis Buffer	TAE
Tris-borate Electrophoresis Buffer	TBE
Tris-EDTA	TE
N,N,N',N' - tetramethylethylenediamine	TEMED
Unit	U
World Health Organisation	WHO

LIST OF TABLES

TABLE	TITLE	PAGE
1	Diagnostic criteria for autistic disorder.....	7.
2	Prevalence studies of autism by population.....	11.
3	Systematic twin studies on autism.....	30.
4	Predicted values of λ_M	45.
5	EMLS for multipoint analysis, 10cM map.....	58.
6	ADI area scores for children in autism multiplex families.....	89.
7	Expected proportion of genome excluded	117.
8	Most-likely lod scores for chromosome 19 (U.S.A. families).....	127.
9	Most-likely lod scores for chromosome 20 (U.S.A. families).....	130.
10	Most-likely lod scores for chromosome 21 (U.S.A. families).....	134.
11	Most-likely lod scores for chromosome 22 (U.S.A. families).....	137.
12	Two-point lod scores for chromosome 22 (U.S.A. families).....	138.
13	Most-likely lod scores for chromosome 6 (U.S.A. families).....	140.
14	Most-likely lod scores for chromosome 15 (U.S.A. families).....	143.
15	Most-likely lod scores for chromosome 1p (U.S.A. families).....	145.
16	Two-point lod Scores for chromosome 1p (U.S.A. families).....	145.
17	Most-likely lod scores for chromosome 1q (U.S.A. families).....	147.
18	Most-likely lod scores for chromosome 10q (U.S.A. families).....	149.
19	Most-likely lod scores for chromosome 17p (U.S.A. families).....	152.
20	Most-likely lod scores for chromosome 18q (U.S.A. families).....	154.
21	Most-likely lod scores for chromosome 1p (All families).....	157.

22 Two-point lod scores for chromosome 1p (All families).....157.

23 Most-likely lod scores for chromosome 22 (All families)..... 163.

24 Two-point lod scores for chromosome 22 (All families).....166.

25 Most-likely lod scores for chromosome 22 (Combined data)..... 167.

26 Most-likely lod scores for chromosome 7 (Australian families)..... 169.

27 TDT tests on chromosome 1p using all families..... 171.

28 TDT tests on chromosome 22 using all families..... 172.

29 Chromosomal regions showing positive linkages and associations to
autism..... 174.

30 Maximum lod scores at regions previously suggestive of containing
susceptibility loci for autism..... 176.

31 Susceptibility loci not replicated in this genome scan..... 189.

32 Comparison of number of sib-pairs required for linkage vs. association
studies.....212.

33 Chromosomal positions of some serotonin receptors..... 215.

LIST OF FIGURES

FIGURE	TITLE	PAGE
1	The co-morbidity of mental retardation, epilepsy and autistic disorder..	14.
2	An NF1 pedigree in which each member has been typed for the IF10 polymorphism.....	49.
3	Lod score plotted against recombination frequency to determine the most-likely recombination frequency for a pair of loci.....	50.
4	The number of known polymorphic markers from 1981 till 1995.....	56.
5	Power to detect linkage at differing λ_s	66.
6	Agarose gel of genomic DNA.....	98.
7	Outline of the genotyping method.....	103.
8	Loading samples on to the Perkin Elmer 310 DNA Analyzer.....	106.
9	ABI Prism™ 310 DNA Analyzer.....	109.
10	ABI Prism™ 373 DNA Sequencer.....	111.
11	Tamra GeneScan-500 size standard.....	115.
12	Flow chart of the linkage study on autism.....	123.
13	Chromosome 19 multipoint results using U.S.A. families.....	128.
14	Chromosome 20 multipoint results using U.S.A. families.....	131.
15	Chromosome 21 multipoint results using U.S.A. families.....	133.
16	Chromosome 22 multipoint results using U.S.A. families.....	136.
17	Chromosome 6 multipoint results using U.S.A. families.....	141.
18	Chromosome 15 multipoint results using U.S.A. families.....	142.
19	Chromosome 1p multipoint results using U.S.A. families.....	146.

20 Chromosome 1q multipoint results using U.S.A. families.....148.

21 Chromosome 10q multipoint results using U.S.A. families.....150.

22 Chromosome 17p multipoint results using U.S.A. families.....153.

23 Chromosome 18q multipoint results using U.S.A. families.....155.

24 Chromosome 1p multipoint results using all families..... 158.

25 Chromosome 22 multipoint results using all families..... 162.

26 Chromosome 22 multipoint results (paternal inheritance) using all
families.....164.

27 Chromosome 22 multipoint results (maternal inheritance) using all
families.....165.

28 Chromosome 7q multipoint results using Australian families.....170.

29 Cytogenetic map of chromosome 22 showing positive linkage results.178.

30 Cytogenetic map of chromosome 1 showing positive linkage results
from this study..... 181.

31 Cytogenetic maps of chromosomal regions showing $MLS \geq 1$ 183.

32 Cytogenetic maps of chromosomal regions most likely to contain
susceptibility loci for autism..... 193.

33 Power to detect linkage using different statistical programmes.....204.

34 False-positive rates for different statistical programmes..... 205.

34 Autistic family with multiple affected individuals 211.

TABLE OF CONTENTS

	PAGE
ABSTRACT.....	ii.
DECLARATION.....	iv.
ACKNOWLEDGEMENTS.....	v.
TABLE OF ABBREVIATIONS.....	vi.
LIST OF TABLES.....	viii.
LIST OF FIGURES.....	x.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1	Introduction.....	1.
1.2	Historical Overview of Autism.....	2.
1.3	Clinical Diagnosis of Autism.....	3.
1.4	Epidemiology.....	8.
1.4.1	Prevalence.....	8.
1.4.2	Incidence.....	12.
1.4.3	Sex Ratios.....	12.
1.4.4	Mental Retardation and Epilepsy in Autism.....	13.
1.5	Aetiological Models of Autism.....	15.
1.5.1	Neurological Conditions Associated with Autism.....	16.
1.5.1.1	Prenatal, Perinatal and Neonatal Factors.....	16.
1.5.1.2	Associated Disorders.....	17.
1.5.2	Autism as a Genetic Disorder.....	27.
1.5.2.1	Family Studies.....	28.

1.5.2.2	Twin Studies.....	29.
1.5.2.3	Family Aggregation of Other Conditions.....	30.
1.5.2.4	Summary of Genetic Involvement in Autism.....	33.
1.6	Neurobiological Factors	34.
1.6.1	Neuropsychological	35.
1.6.2	Neuroanatomical.....	36.
1.6.2.1	Neuroimaging Studies.....	36.
1.6.2.2	Post-Mortem Studies.....	38.
1.6.3	Neurochemistry	38.
1.6.4	Summary of the Brain Mechanisms Involved in Autism....	40.
1.7	Mode of Inheritance.....	40.
1.8	Linkage Analysis.....	45.
1.8.1	Principles of Genetic Linkage Analysis in Mendelian Inheritance.....	46.
1.8.1.1	DNA Polymorphism.....	53.
1.8.1.2	Polymorphic DNA Markers.....	53.
1.8.1.3	Family Material.....	56.
1.8.2	Complex Disorders.....	58.
1.8.2.1	Defining the Disease Phenotype.....	61.
1.8.2.2	Linkage Analysis Methods Used for the Study of Complex Disorders.....	62.
1.8.3	Sib-Pair Methodology.....	63.
1.8.3.1	Power to Detect Linkage.....	64.
1.8.3.2	Multipoint Sib-Pair Analysis.....	67.

1.8.3.3	Genes Localised in Complex Disorders using Multipoint Sib-Pair Analysis.....	67.
1.8.4	Association Analysis.....	70.
1.8.5	Linkage and Association Studies in Autism.....	72.
1.8.5.1	HLA and 19 Autosomal Markers (Spence <i>et al.</i> , 1985).....	72.
1.8.5.2	HLA (Warren <i>et al.</i> , 1991; etc).....	73.
1.8.5.3	Chromosome 15 (Gillberg <i>et al.</i> , 1991b; etc)	73.
1.8.5.4	Fragile X (Hallmayer <i>et al.</i> , 1994).....	73.
1.8.5.5	X-Chromosome (Hallmayer <i>et al.</i> , 1996b).....	74.
1.8.5.6	X-Chromosome (Petit <i>et al.</i> , 1996).....	74.
1.8.5.7	Harvey-Ras Factors (Herault <i>et al.</i> , 1995).....	75.
1.8.5.8	Serotonin (Cook <i>et al.</i> , 1997b).....	75.
1.8.5.9	Whole Genome Scan (IMGSAC, 1998).....	76.
1.8.5.10	Whole Genome Scan (Bass <i>et al.</i> , 1998).....	77.
1.8.5.11	Whole Genome Scan (Philippe <i>et al.</i> , 1999).....	78.
1.9	Conclusion.....	78.
1.10	Rationale of the Study.....	80.
1.11	Research Aims and Significance.....	81.
1.11.1	Specific Aims.....	81.
1.11.2	Significance.....	84.

CHAPTER 2: SUBJECTS AND METHODS

2.1	Clinical Methods.....	85.
2.1.1	Ascertainment of Families.....	85.

2.1.2	Diagnosis.....	86.
2.1.3	Intelligence Test Scores.....	88.
2.1.4	Ethical Considerations.....	90.
2.2	Clinical Characteristics of Autism Multiplex Families.....	91.
2.2.1	U.S.A. Families.....	91.
2.2.2	Australian Families.....	92.
2.3	Laboratory Methods.....	93.
2.3.1	DNA Isolation.....	93.
2.3.1.1	Cell Lines.....	94.
2.3.1.2	Salting-Out Method of DNA Extraction.....	95.
2.3.1.3	Measurement of DNA Concentration.....	96.
2.3.1.4	Evaluation of DNA Samples.....	97.
2.3.2	Microsatellite Analysis.....	98.
2.3.2.1	Microsatellites.....	98.
2.3.2.2	Selection of Microsatellites and Synthesis of Primers.....	99.
2.3.2.3	PCR Amplification and Different Methods of Labelling the PCR Products for Subsequent Detection.....	101.
2.3.2.4	Electrophoretic Separation and Detection of the PCR Products of Different Length.....	106.
2.3.2.5	Preparation of Detection Systems for Analysis...	108.
2.3.2.6	Allele Assignment.....	113.
2.3.3	Verification and Processing of Genotyping Data.....	115.

2.4	Statistical Analysis.....	117.
2.4.1	Statistical Model.....	117.
2.4.2	Linkage Analysis.....	118.
2.4.3	Linkage Disequilibrium Analysis (LD).....	120.
2.4.4	Computer Programme.....	120.

CHAPTER 3: RESULTS

3.1	Linkage Results from U.S.A. Families.....	124.
3.1.1	Genotyping Error Rate.....	124.
3.1.2	Systematic Scan of Chromosomes 19-22.....	125.
3.1.2.1	Chromosome 19.....	126.
3.1.2.2	Chromosome 20.....	129.
3.1.2.3	Chromosome 21.....	132.
3.1.2.4	Chromosome 22.....	134.
3.1.3	Candidate Regions.....	139.
3.1.3.1	HLA Region on Chromosome 6.....	139.
3.1.3.2	Chromosome 15.....	139.
3.1.4	Duplication of Positive Regions Detected by the Stanford and Perth Research Groups.....	144.
3.1.4.1	Chromosome 1p.....	144.
3.1.4.2	Chromosome 1q.....	147.
3.1.4.3	Chromosome 10q.....	149.
3.1.4.4	Chromosome 15q.....	151.
3.1.4.5	Chromosome 17p.....	151.
3.1.4.6	Chromosome 18q.....	154.

3.2	Linkage Results from Australian Families.....	156.
3.2.1	Extension Study of Positive Regions Detected Using U.S.A. Families.....	156.
3.2.1.1	Chromosome 1p.....	156.
3.2.1.2	Chromosome 22.....	159.
3.2.2	Replication of Potential Susceptibility Regions.....	168.
3.3	Linkage Disequilibrium Results.....	171.

CHAPTER 4: DISCUSSION

4.1	Findings from the Study.....	173.
4.1.1	Genome Scan.....	173.
4.1.1.1	Susceptibility Loci.....	173.
4.1.1.2	Excluded Regions.....	182.
4.1.2	Candidate Regions.....	186.
4.2	Replication Studies.....	188.
4.3	The Genetics of Autism – An Overview of Findings to Date.....	191.
4.4	Reliability of Results.....	195.
4.4.1	Ascertainment.....	195.
4.4.2	Diagnostic Criteria.....	197.
4.4.3	Linkage Maps.....	199.
4.4.4	Reliability of Genotyping.....	202.
4.4.5	Aspex Programme.....	202.
4.5	Implications for Future Studies of the Genetics of Autism.....	206.
4.5.1	Refining the Phenotype Definition.....	207.
4.5.2	Alternative Family Structures.....	209.

4.5.2.1 Cousin Pairs..... 209.

4.5.2.2 Pairs Discordant for Disease..... 209.

4.5.2.3 Extended Families..... 211.

4.5.3 Association Studies..... 212.

4.5.3.1 Candidate Gene Approach..... 213.

4.5.3.2 Single Nucleotide Polymorphisms..... 215.

4.6 Conclusions and Outlook..... 216.

CHAPTER 5: BIBLIOGRAPHY..... 218.

CHAPTER 6: APPENDICES

6.1 Microsatellite Information and Analysis Conditions..... 246.

6.2 Solution and Buffer Protocols..... 250.

6.3 Published Papers..... 255.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Autism is a severe, lifelong neuropsychiatric disorder with a profound impact on the affected individuals, their families and society at large. The causes of autism remain unknown, however a large number of observations suggest that genetic factors may play a role in its aetiology. This PhD project explores this hypothesis via the search for susceptibility genes for autism throughout the human genome.

The introduction and literature review begin by giving a historical perspective on the study of autism and the changes in diagnostic criteria throughout the last 40 years, resulting in what is thought to be a more valid method of diagnosis (ICD-10 & DSM-IV). The findings from epidemiological research in autism are then described, followed by an account of the aetiological models of autism. Specific reference is made to the genetic model and diseases thought to co-occur with autism that could point to candidate regions in the genome. A brief review on the brain mechanisms that are thought to be involved in autism follows. Although there is extensive literature on this subject, only areas relevant to the genetics of autism are reviewed. The introduction concludes by suggesting the modes of inheritance which are most pertinent to this disorder. Using these models, the methods of linkage analysis that could be applied to locate susceptibility loci are described.

1.2 Historical Overview of Autism

Children exhibiting symptoms similar to autism have been recognised for a long time, with the earliest reports made at the beginning of the 19th century (Vaillant, 1962). Clear descriptions have been given by Haslam (1809), Itard (translation into English by Lane 1976), and Critchley and Earl (1932). However, it was Leo Kanner (1943) and Hans Asperger (1944), independent of each other, who first described the syndrome and assigned to it a particular name including reference to autism. Kanner was the first to outline the common characteristics and special qualities of the behaviour of children with autism. His original paper outlined 11 children carefully observed over 5 years. These individuals all shared “fascinating peculiarities”; including delayed echolalia, pronoun reversal, failure to use speech to communicate, an anxious desire to preserve sameness, repetitive behaviours, failure to anticipate being lifted, a general lack of awareness of other people’s existence or feelings and a lack of ability to play imaginatively with other children. He also described these children as having a normal intellectual potential (based mainly on their “intelligent physiognomies”) and good physical health. Although his assumptions about normal intelligence were refuted in future studies due to the occurrence of mentally retarded individuals with autism, Kanner’s description of autism continues to be one of the most vivid representations of the behaviour of autistic children.

Asperger’s (1944) definition of the syndrome (autistic psychopathy) was very similar to the disease Kanner described at the symptom level, however, there were some notable differences. Asperger’s definition did not specify an age of onset, whereas to be diagnosed with Kanner’s definition of autism the onset of the disease had to be within the first 3 years of life. Also, the children described by Asperger all

developed speech before school age, they typically had large vocabularies and reasonable grammar, and, although socially isolated, they were not unaware of the existence of others. Asperger described the children he studied as odd in appearance in contrast to the alert, attractive look often mentioned as typical of the children Kanner studied. According to the International Classification of Diseases (ICD-10) the syndromes described by Kanner and Asperger are two separate disorders, namely autism and Asperger syndrome respectively. The main criteria that distinguish these two disorders are the lack of mental retardation as well as the absence of a language delay in individuals with Asperger syndrome. It has also been noted that Asperger syndrome is at least five times more common than the autistic disorder (Gillberg, 1989; Ehlers and Gillberg, 1993). Whether Asperger disorder and autism represent aetiologically separate entities, rather than different degrees of severity of the same condition, is not yet known (Volkmar *et al.*, 1994; Frith, 1991). With the identity of the genes involved in the “core” syndrome of autism, it is hoped that the boundary of the autistic phenotype can be determined.

1.3 Clinical Diagnosis of Autism

The classification and diagnosis of autism have changed over time as the key features and associated characteristics of the disorder have been revised. Until recently, there were two predominant models for the diagnosis of psychiatric disorders, namely “the great professor” principle and “the consensus of experts”. In the beginning of this century, many great men in psychiatry, including Kraepelin and Bleuler, developed and published their own diagnostic systems of the major psychoses

(Kraepelin and Diefendorf, 1923; Bleuler and Brill, 1924). Whether these diagnostic criteria were accepted or rejected relied on many factors (e.g. clarity of thought and clinical applicability). Ultimately, however, the validity of these systems rested on the personal authority of their originators which was generally only feasible within the region of the country in which they lived (Kendler, 1990). This led to the use of different diagnostic systems in different countries or even different regions within the same country. With the development of national and international psychiatric organisations, committees were formed to resolve specific diagnostic issues. This was generally achieved by the “consensus of experts” approach, although sometimes the issue would be put to a vote. Examples of diagnostic criteria formulated using this approach are the Diagnostic and Statistical Manual of Mental Disorders I and II (DSM-I and DSM-II; American Psychiatric Association (APA), 1952 and 1968)

Kanner and Eisenberg (1956) were the first to define a set of diagnostic criteria for autism. These became known as “Kanner’s criteria” and included the following:

- (I) profound lack of affective contact with other people,
- (II) an anxiously obsessive desire for the preservation of sameness in the child’s routines and environment,
- (III) a fascination for objects, which are handled with skill in fine motor movements,
- (IV) mutism or a kind of language that does not seem intended for inter-personal communication,
- (V) good cognitive potential shown in feats of memory or skills on performance tests.

They also emphasised that onset of autism should be from birth or before 30 months of age, and that the first two criteria were essential for the diagnosis. They stated in this paper that, if a child was diagnosed with the first two features, the rest of the typical clinical picture would also be found.

Rutter (1978) developed the second set of diagnostic criteria. He suggested four criteria for defining childhood autism. These were:

- (I) impaired social development which has a number of special characteristics out of keeping with the child's intellectual level,
- (II) delayed and deviant language development that also has certain defined features and is out of keeping with the child's intellectual level,
- (III) "insistence on sameness" as shown by stereotyped play patterns, abnormal preoccupations or resistance to change,
- (IV) Onset before 30 months

The American Psychiatric Association and the World Health Organisation (WHO) have also published their definitions of autism in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD), respectively. The development of DSM-III (American Psychiatric Association, 1980) saw a change in the way autism was defined, mainly due to the increased knowledge obtained through research. These organisations used available scientific information in the development and evaluation of proposed changes and used diagnostic criteria that would greatly facilitate the reliability and validity of future studies (Kendler, 1990).

DSM-III and DSM-III-R (American Psychiatric Association, 1987) contained criteria for autism which were quite different from each other and from those of the International Classification of Diseases (ICD). The DSM-III-R diagnostic criteria of

autism tended to overdiagnose this disorder compared with the ICD-10 (World Health Organisation, 1990) and the DSM-III. DSM-III-R also elaborated the symptoms of autistic disorder, discarded the age criterion and did not contain the childhood-onset pervasive developmental disorder subgroup. The criteria used in ICD-10 and, recently, DSM-IV (American Psychiatric Association, 1994) are different again from their previous versions, however, they are essentially similar to each other and will allow future studies conducted in different countries to be more comparable.

Regardless of the diagnostic manual used, the diagnosis of autism rests on a set of behavioural criteria, invariably comprising communication, social interaction and behavioural deficits. Each manual emphasises slightly different sub-criteria for each class of these symptoms. However, they all agree that there must be severe problems: (1) in the ability to develop relationships with people; (2) in the development of speech, with abnormal language patterns and communicative use of speech after it develops; and (3) in coping with any forced change in routine. These three categories of problems are now often referred to as “the triad of social, language and behavioural impairment” (Wing, 1981). Some manuals (Kanner and Eisenberg, 1956; Rutter, 1978; APA, 1980; WHO, 1990; WHO, 1992; APA, 1994) require that the age of onset be before 30-36 months, whereas others specify only that signs of the triad be present from infancy or childhood (APA, 1987). The set of diagnostic criteria of the DSM-IV and the ICD-10 are the ones currently most often used. The criteria in the ICD-10 manual were used to diagnose the subjects for this study (Table 1).

The precision of these definitions will continue to evolve as new research clarifies the phenotype (visible characteristics of autism). At present, DSM-IV International Field Trials have indicated that the clinical diagnosis of autism remains

one of the most reliable diagnoses in psychiatric or developmental research (Volkmar *et al.*, 1994).

Table 1: Diagnostic Criteria for Autistic Disorder

- A. Qualitative impairments in reciprocal social interaction, as manifested by at least two of the following four:
 - (1) Failure adequately to use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction
 - (2) Failure to develop peer relationships that involve a mutual sharing of interests, activities, and emotions
 - (3) Lack of socio-emotional reciprocity
 - (4) Lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (eg. a lack of showing, bringing, or pointing out to other people objects of interest to the individual)

- B. Qualitative impairments in communication as manifested by at least one of the following four
 - (1) a delay in, or total lack of, development of spoken language that is not accompanied by an attempt to compensate through alternative modes of communication
 - (2) relative failure to initiate or sustain conversational interchange
 - (3) stereotyped and repetitive use of words or phrases
 - (4) lack of varied spontaneous make-believe or social imitative play

- C. Restricted, repetitive, and stereotyped patterns of behaviour, interests, and activities, as manifested by at least one of the following four:
 - (1) Encompassing preoccupation with stereotyped and restricted patterns of interest
 - (2) Apparently compulsive adherence to specific, non-functional, routines/rituals
 - (3) Stereotyped and repetitive motor mannerisms
 - (4) Preoccupations with part-objects or non-functional elements of play material

- D. At least six of the above twelve symptoms must be present.

- E. Developmental abnormalities (in at least one of language, social interaction, or play) must have been present in the first 3 years for the diagnosis to be made.

Table from ICD-10, WHO (World Health Organization) 1992.

1.4 Epidemiology

Epidemiology is the branch of medical science that deals with the incidence, distribution, and control of disease in a population (Webster's). *Prevalence* refers to the rate of the disorder present in the population at a given point of time and *Incidence* refers to the number of new cases occurring in the population during a given period (Bristol *et al.*, 1996). Epidemiological surveys of autism are important to identify the prevalence and/or incidence of this developmental disorder, and the requirements for specialised educational and medical services for this group of children. These surveys are also important to genetic studies, in order to show the increase in the prevalence rate within families affected by the disease compared to the general population. This allows the genetic component involved in aetiology of the disorder to be evaluated.

1.4.1 Prevalence

A number of population-based studies on the prevalence of autism have been completed, with the rates varying enormously across studies. Recent reviews of the epidemiological surveys of autism (Wing, 1993; Gillberg, 1995; Honda *et al.*, 1996; Fombonne *et al.*, 1997; Fombonne, 1999) show that rates vary from 0.7/10 000 (Treffert, 1970) to 21.1/10 000 (Honda *et al.*, 1996). Much of the variation across studies can be explained by the use of relatively small samples, and differences in ascertainment schemes and diagnostic criteria. A review by Honda *et al.* (1996) divided published prevalence studies of autism by population size (Table 2). Those

studies with populations of less than 50,000 showed an average prevalence of 13.4 per 10,000, whereas studies with populations of more than 50,000 had an average prevalence of approximately 5.8 per 10,000. Honda *et al.* (1996) suggested that the reason for the discrepancies between the two population sizes could be due to the higher number of overlooked cases in the larger populations. It is still not proven what population size is optimal for epidemiological studies of autism, and sizes less than 50,000 may be too small for this rare disorder.

Studies in the last few years have estimated the prevalence of autism to be typically around 0.1 in 10,000 children (Landgren *et al.*, 1996; Webb *et al.*, 1997; Arvidsson *et al.*, 1997). This is 2-5 times greater than the figure commonly reported in the past (2 to 5 in 10 000 children). The high rate reported in these more recent studies could be attributed to the researchers' increased awareness of the autism spectrum disorders and the inclusion of these disorders in their epidemiology studies. Autism spectrum disorders is a term used to describe all conditions involving severe autistic symptomatology, or those conditions with severe autistic symptomatology that do not meet operationalised criteria for autistic disorder/childhood autism (Gillberg, 1995).

In Gillberg's (1995) review of epidemiological studies of autism, he provided an estimate of the overall prevalence of autism in the population. In this report he reviewed all papers published in books or journals that included reference to the words "autism" and "prevalence", and were written in English or with comprehensive English summaries included. Only papers that met the following criteria were included in the review:

- The study used defined diagnostic criteria for autism;

- The population studied referred to a specific age-cohort from a geographically defined population;
- The initial screening for autistic individuals was made within a large population of children, including not only those who had already been clinically diagnosed as “autistic” or “psychotic”; and
- The final diagnosis had been made after clinical examination of possible cases by a research team or by primary care clinical staff who provided the information requested by the research team as part of the study.

Table 2 lists those 16 studies which met the criteria for inclusion and which had been published by 1st of June 1993. Publications in italics were those noted in the Honda *et al.* (1996) and the Fombonne (1999) reviews but were not included in Gillberg (1995). The mean prevalence across all 16 studies was 6.5 per 10 000 children. This mean was computed by adding all cases diagnosed with autism and dividing that figure by the total sum of all those individuals who participated in the original screening. If the prevalence figures were added and divided by 16 the figure would be 8.4 (± 4.4 (S.D.)). It is interesting to note that if the two U.S.A. studies were removed (the U.S.A.’s mobile population and the unevenly distributed health-care system around the country causes problems in population-based studies), the mean rate of reported autism in the studies published after 1985 would be 11.8 per 10 000 children. This figure is very close to the recent prevalence estimates of around 0.1 in 10,000 (Landgren *et al.*, 1996; Webb *et al.*, 1997; Arvidsson *et al.*, 1997). As previously mentioned, these studies also included individuals diagnosed with autism spectrum disorders. In this study, only children with the “core” syndrome of autism were included in the analysis and, in turn, a prevalence rate of 4-5 per 10 000 children born was used throughout.

Table 2: Prevalence Studies of Autism by Population (prevalence is given as n per 10 000 children born)

Study	Region	Population Size	Prevalence	Male:Female Ratio
Population size <50,000				
<i>Honda et al. (1996)</i>	<i>Japan</i>	8,537	21.1	2.6:1
Sugiyama and Abe (1989)	Japan	12,263	13.0	?
Bryson <i>et al.</i> (1988)	Canada	20,800	10.1	2.5:1
Matsuishi <i>et al.</i> (1987)	Japan	32,834	15.5	4.7:1
Ishii and Takahashi (1982)	Japan	34,987	16.0	6.0:1
Wing and Gould (1979)	England	35,000	4.9	16.0:1
Population size >50,000				
<i>Fombonne (1999)</i>	<i>Review</i>	>4,000,000	5.2	3.8:1
<i>Fombonne et al. (1997)</i>	<i>French</i>	325,347	5.4	1.8:1
<i>Fombonne & du Mazaubrun (1992)</i>	<i>French</i>	274,816	4.9	2.1:1
Gillberg <i>et al.</i> (1991a)	Sweden	78,106	11.5	2.8:1
Ritvo <i>et al.</i> (1989)	USA	184,822	4.0	3.7:1
Cialdella and Mamelie (1989)	France	135,180	10.8	2.0:1
Tanoue <i>et al.</i> (1988)	Japan	95,394	13.8	4.1:1
Burd <i>et al.</i> (1987)	USA	180,986	3.3	2.7:1
<i>Steinhausen et al. (1986)</i>	<i>Germany</i>	279,616	1.9	2.2:1
Steffenburg and Gillberg (1986)	Sweden	78,413	7.5	3.0:1
Gillberg (1984)	Sweden	125,584	4.0	1.8:1
<i>McCarthy et al. (1984)</i>	<i>Ireland</i>	65,000	4.3	1.3:1
Bohman <i>et al.</i> (1983)	Sweden	>50,000	6.1	1.6:1
Hoshino <i>et al.</i> (1982)	Japan	217,626	5.0	9.0:1
Brask (1972)	Denmark	>50,000	4.3	1.4:1
<i>Treffert (1970)</i>	<i>USA</i>	899,750	3.1	3.4:1
Lotter (1966)	England	78,000	4.5	2.6:1

(Gillberg, 1995; Honda *et al.*, 1996)

1.4.2 Incidence

Only one paper to date (Honda *et al.*, 1996) has reported the cumulative incidence of childhood autism. A measure of incidence can often be more useful than prevalence, but can be difficult to calculate in childhood disorders (Stein and Susser, 1981). In autism, problems generally arise due to the difficulty in identifying the exact age in which the first symptoms of the disease arise. Honda *et al.* (1996) overcame this problem by using cumulative incidence up to five years of age within a specified birth cohort. They reported a cumulative incidence of 16.2 per 10 000 in the birth cohort. This value was lower than the reported prevalence for this population, which according to the authors was a consequence of parents staying in areas where special facilities were available for their autistic children.

1.4.3 Sex Ratios

Autism is known to affect males more frequently than females. Papers on autism constantly report the male:female ratio to be 3-4:1. Population studies, however, show a variety of values ranging from 1.3:1 (McCarthy *et al.*, 1984) to 16:1 (Wing and Gould, 1979). In the reviews of sixteen population studies of autism, Wing (1993) and Gillberg (1995) found that the ratio was closer to 2-3:1 (Table 2). Wing (1993) pointed out, that “the exceptionally high ratio in the Wing and Gould (1979) study (16:1) was most probably a chance finding due to the small number of autistic children identified”. There have also been several studies suggesting that the ratio is associated with IQ levels, with an equal proportion of males to females seen

among autistic children with severe and profound mental retardation (Wing, 1981; Gillberg, 1984).

The reason for the higher vulnerability of males to autism is unknown. Many speculations have been put forth to explain the skewed sex ratios, and one rationale could be due to a genetic involvement. Skuse *et al.* (1997) recently presented evidence for one or more loci on the human X chromosome influencing social cognition. Males do not express these imprinted loci as they are only expressed from a paternally inherited X chromosome. Skuse *et al.* (1997) suggested that such a locus might offer some explanation of why boys are more vulnerable than girls to developmental disorders of speech and language. The X-chromosome was examined in detail within this study to determine whether any susceptibility loci were present on this chromosome. Alternatively, because girls have better developed superficial social and language skills early in life, the basic disabilities associated with autism could go unrecognised for longer periods of time. This would lead to an underestimate of the prevalence of autism in girls and elevate the difference in the sex ratio (Gillberg, 1995).

1.4.4 Mental Retardation and Epilepsy in Autism

Intellectual handicap (defined as IQ under 70) has been shown to be present in 75-90% of cases of autism (Gillberg, 1990) (Figure 1). It has also been shown to be the single most powerful predictor of outcome in autism (Venter *et al.*, 1992), since the severity of autistic symptoms is related to verbal IQ (Bolton *et al.*, 1994).

Although autism and mental retardation are strongly associated, the accompanying

risks and neurological findings are different. Mental retardation is associated with an increased rate of major pre- and perinatal abnormalities, while autism is associated with mainly minor abnormalities (Finnegan and Quarrington, 1979). In addition, although autism and mental retardation are associated with the same disorders (eg. infantile spasm and congenital rubella), other conditions that are usually associated with intellectual handicap rarely give rise to autism (eg. cerebral palsy). These different associations suggest that while there is some overlap in the specific disorders known to cause the two conditions, that overlap is small (Folstein and Rutter, 1988). Also, since it is difficult to distinguish between mental retardation and autism in children with low IQs (< 30), such individuals were removed from the study to prevent misdiagnosis (see section 2.1.3).

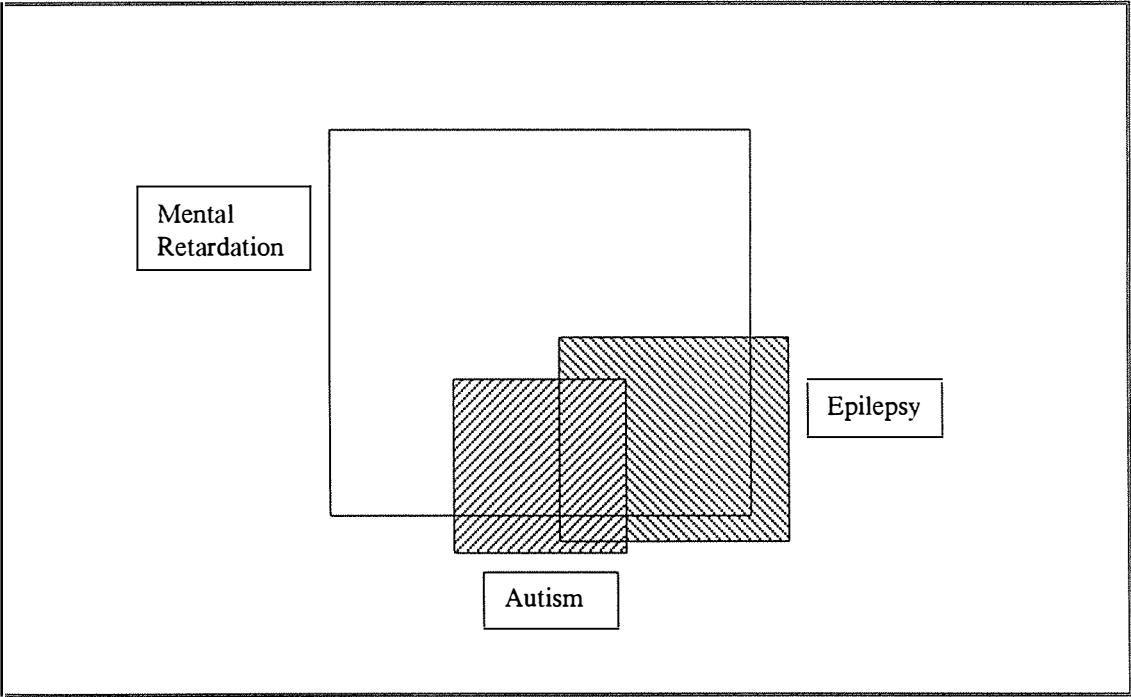


Figure 1: The co- morbidity of mental retardation, epilepsy and autistic disorder.

Epilepsy is also common in people with autism, particularly between early childhood and adolescence (Figure 1). About 20% of autistic individuals develop

non-febrile seizures before the age of about 3 years. Another 15-20% develop seizures around the time of puberty or later (Olsson *et al.*, 1988; Volkmar & Nelson, 1990; Gillberg and Coleman, 1992). The increased incidence of epilepsy and abnormal electroencephalograms (EEG) in autistic children, compared with other psychiatric patients, was the first piece of evidence that autism is a biological disorder of the brain rather than a psychogenic disorder (Ornitz, 1978). Although epilepsy is not strongly associated with IQ, it does seem to have a higher prevalence in autistic individuals who are mentally retarded (Bailey *et al.*, 1996). The prevalence of epilepsy among the individuals used in this study is not discussed as it is beyond the scope of this project.

1.5 Aetiological Models of Autism

The causes of autism still remain uncertain, and opinions about the aetiology have varied over the years. It was first put forth by Kanner (1943) that autism was due to an inborn defect, since the onset of the disorder occurs very early in life. Other views suggested that the condition is the outcome of parental neglect, rejection, or indifference (Bene, 1958; Bettelheim, 1967; Despert, 1956). Numerous clinical and epidemiological studies, however, have failed to show a positive association between autism and qualities of upbringing or frequency of stress experiences (Cantwell *et al.*, 1978; Bailey *et al.*, 1996). Accordingly, it seems unlikely that psychosocial environmental risk factors are involved in the aetiology of autism. Current research findings support Kanner's initial interpretations, and it is now generally accepted that autism has a biological cause.

Like other behavioural syndromes, autism is an aetiologically heterogeneous disorder. A number of neurological conditions, some traumatic, some infectious, and some genetic in aetiology, have been reported to be associated with autism. These conditions are presumably the cause of the syndrome in a small number of cases and so they will be reviewed below. Fragile X syndrome, chromosome 15 deletions and several other disorders are discussed in more detail; since these conditions are frequently reported to co-occur with autism, the chromosomal regions involved have been included in this study as candidate regions for susceptibility loci to autism. Although the exact figures are unknown, cases with a known or presumed aetiology constitute a small proportion (10-25%; see section 1.5.1.2) of all autistic children. The majority of cases are of unknown aetiology and several observations suggest that genetic factors may play a role. This theory, supported by family and twin studies as well as the presence of other related conditions within the family, is also reviewed in the following section.

1.5.1 Neurological Conditions Associated with Autism

1.5.1.1 Prenatal, Perinatal and Neonatal Factors

Pre-, peri- and neonatal complications have been reported to be more prevalent in individuals with autism than in control groups (Deykin and MacMahon, 1980; Finnegan and Quarrington, 1979; Torrey *et al.*, 1975). These factors include abnormal obstetrical presentation (Finnegan and Quarrington, 1979; Levy *et al.*, 1988), advanced maternal age (Gillberg and Gillberg, 1983; Tsai and Stewart, 1983),

low birth weight (Knobloch and Pasamanick, 1975), prematurity (Gilberg and Gilberg, 1983), use of medication during pregnancy (Deykin and MacMahon, 1980), viral illness (Deykin and MacMahon, 1979), low Apgar score (Finnegan and Quarrington, 1979), vaginal bleeding (Gilberg and Gillberg, 1983; Knobloch and Pasamanick, 1975; Torrey *et al.*, 1975) and hyperbilirubinemia (Finnegan and Quarrington, 1979). In studies controlling for birth order, no differences in the rates of pre-, peri- and neonatal problems between autistic subjects and control subjects were detected (Lord *et al.*, 1991; Piven *et al.*, 1993). The association of pre-, peri- and neonatal factors with autism has an uncertain significance (Rutter *et al.*, 1990), and it has been suggested that these factors could be consequences of underlying genetic problems in the fetus (Bailey *et al.*, 1995; Bolton *et al.*, 1994).

1.5.1.2 Associated Disorders

A number of pathological conditions have been reported to be associated with autism. These range from infectious diseases to chromosomal abnormalities, and neurocutaneous and metabolic disorders. Gillberg and Coleman (1996) recently reviewed all population studies on autism published in the English language with particular reference to co-morbidity. Seven studies met the criteria for inclusion in the survey. These criteria were similar to those used in Gillberg's (1995) paper for estimating the overall prevalence of autism in the population (see above). They found that approximately one in four people (24.4%) with autism have an associated disorder. There was also a trend for higher rates of associated disorders among individuals with severe mental retardation. It still remains unclear whether the associated disorders are aetiologically linked to the symptoms of autism or whether

the individuals have two separate and unrelated disorders (Gillberg and Coleman, 1996). Other studies regard additional disorders as rather rare (up to 10%) among autistic children (Ritvo *et al.*, 1990; Smalley, 1991; Rutter *et al.*, 1994). The reasons for the discrepancy could be due to differences in the stringency of diagnostic criteria for the associated conditions as well as ethnic differences in the aetiology of the disorders (Gillberg, 1992a).

Infectious Diseases

Pre- and postnatal infections associated with autism include encephalitis (Deykin and MacMahon, 1979; Markowitz, 1983), syphilis (Stubbs, 1978; Stubbs and Crawford, 1977), toxoplasmosis (Todd, 1986), CMV-infection (Stubbs *et al.*, 1984), herpes simplex (Ghaziuddin *et al.*, 1992a) and congenital rubella (Chess, 1971). The evidence for these associations is not convincing in the majority of cases.

The association between congenital rubella and autism was first shown by Chess (1971). However in a follow-up study, Chess (1977) noted that the autistic symptoms in these children with congenital rubella improved over time. A review of epidemiological studies in autism (Fombonne, 1998) reported a 0.9% median rate of congenital rubella, a value not significantly different from the rates in control groups. A recent Japanese survey of autism conducted in the aftermath of a rubella epidemic also failed to document any single case of autism associated with congenital rubella (Honda *et al.*, 1996).

Chromosomal Abnormalities

Autism has been linked to different chromosomal abnormalities. While most of these abnormalities involve sex chromosomes, structural autosomal defects have also been reported. Recent reviews of the literature (Gillberg, 1998; Lauritsen *et al.*, 1999) have reported abnormalities in each of the autosomes and of the gonosomes among individuals with autism / autism-like behaviour. These include a partial trisomy 8q (Ritvo *et al.*, 1990), a partial trisomy 6p21-6p25-ter (Burd *et al.*, 1988), a partial trisomy 15 (Gillberg *et al.*, 1991b; Schinzel, 1990), a 1q deletion (Murayama *et al.*, 1991) and several other defects such as a translocation t(3;12)(p26.3;q23.3) (Fahsold *et al.*, 1991) or a complex chromosome rearrangement involving chromosomes 1,7 and 21 (Lopreiato and Wulfsberg, 1992). Most of these abnormalities have only been reported in isolated cases of autism and it is not known whether they are associated with autistic behaviour by chance alone (Gillberg and Coleman, 1992). In a review of the literature on autosomal chromosome abnormalities from January 1987 to June 1998, Lauritsen *et al.* (1999) detected three interesting regions with multiple reports of cases with abnormalities in the same area. These include 15q11-13, 16q23 and 17p11.2, with chromosome 15 being the most frequently reported aberrant site. These regions are good candidates for containing susceptibility loci involved in the aetiology of autism, and were studied within the course of this study. Chromosome 15 abnormalities (Gillberg *et al.*, 1991b) together with Down syndrome (trisomy 21) (Wakabayashi, 1979), and fragile X (Hagerman, 1989), have also all been reported to co-occur with autism in a number of studies and are reviewed in detail below.

Chromosome 15 Aberrations

Aberrations, including duplications and to a lesser extent deletions, on the proximal long arm of chromosome 15, have been reported to be the single most common identifiable cause of autism (Schroer *et al.*, 1998). Kerbeshian *et al.* (1990) first reported an individual with autism who had a supernumerary inversion on chromosome 15. Since then, there have been a number of reports that associate deletions, uniparental disomy (UPDs), or supernumerary inverted duplications on chromosome 15 with autism. Autistic patients with these aberrations commonly have mental retardation and epilepsy as well (Baker *et al.*, 1994; Bunday *et al.*, 1994, Schroer *et al.*, 1998).

The most common abnormality involves duplications within the 15q11-q13 region (Gillberg *et al.*, 1991b, Martinsson *et al.*, 1994; Bunday *et al.*, 1994; Baker *et al.*, 1994; Flejter *et al.*, 1996; Cook *et al.*, 1997a). This is the critical region for Prader-Willi syndrome (PWS) and Angelman syndrome (AS). Both of these syndromes involve developmental abnormalities and mental retardation, and show abnormalities in the 4Mb-region spanning 15q11-q13. In the case of PWS, both alleles at this specific location on chromosome 15 have been inherited from the mother or the paternal allele has been deleted. AS, on the other hand, is the deficiency of normal maternal DNA at 15q11-q13 (Chan *et al.*, 1993). Although PWS and AS have not been shown to be directly associated with autism, duplications in this region have. All duplications of the PWS/AS critical region in autistic patients have been maternal in origin (Robinson *et al.*, 1993; Leana-Cox *et al.*, 1994; Schinzel *et al.*, 1994; Flejter *et al.*, 1996; Cook *et al.*, 1997a). These studies found that paternally inherited duplications of 15q11-q13 lead to a non-autistic phenotype,

whereas maternally inherited duplications usually lead to autism or atypical autism. This suggests that an imprinted gene at chromosome 15 may contribute to susceptibility to autism in some cases.

Down Syndrome

Wakabayashi (1979) first described the association of Down syndrome (DS) with autism. Since then, a number of studies have reported the occurrence of autism in individuals with Down syndrome (Wing and Gould, 1979; Gath and Gumley, 1986; Bregman and Volkmar, 1988; Myers and Pueschel, 1991; Gillberg *et al.*, 1986; Ghaziuddin *et al.*, 1992b; Howlin *et al.*, 1995; Ghaziuddin, 1997). The association reported in these studies could be explained by (1) a family history and genetic predisposition to autism (Ghaziuddin, 1997), (2) complicating physical conditions that occur in DS (such as brain damage) and may increase the risk of autism symptoms (Gillberg *et al.*, 1986), (3) the wide range of deficits caused by trisomy 21 that could contribute to the emergence of autistic features (Howlin *et al.*, 1995), and (4) neuropathological aberrations present in both disorders might reflect altered patterns of functional organisation that are common to both (Ghaziuddin *et al.*, 1992b). Overall the occurrence of Down syndrome in people with autism is relatively rare. Presently, it does not seem that Down syndrome, as such, predisposes to the development of autistic behaviour (Gillberg, 1993).

Fragile X

The association of fragile X syndrome and autism is well established, however the prevalence of this association differs significantly between studies. The fragile X syndrome is the most common cause of hereditary mental retardation. It is a recessive X-linked disease with incomplete penetrance and affects boys and girls at a ratio of 3:1 (Davies, 1991). Originally it was detected cytogenetically by the presence of the fragile site (FraX) Xq27.3. In 1991, Verkerk *et al.* identified the FMR-1 gene (fragile X mental retardation) that is responsible for this disorder, and the syndrome can be detected with molecular genetic methods.

Studies before the identification of the FMR-1 gene reported the rate of fragile X in individuals with autism to range from 0 to 25%, with an average of 7% (Bolton and Rutter, 1990). The wide variation between studies could be due to the ascertainment procedures utilised, to the use of small sample sizes, as well as different diagnostic methods (Folstein and Piven, 1991). In addition, the common fragile site at Xq27.2 could have produced frequent false-positive results in individuals with low levels of FraX (Xq27.3) expression (ie. 1% to 3%) (Ramos *et al.*, 1989; Sutherland and Baker, 1990).

More recent investigations, which use standard techniques and the same valid and reliable clinical diagnostic criteria (ADI, see section 2.1.2), have tended to report lower rates of co-occurrence between the two syndromes (Bailey *et al.*, 1993a; Piven *et al.*, 1991a). Bailey *et al.* (1993a) replicated the Piven *et al.* (1991a) study, and detected the fragile X anomaly in 1.6% of autistic individuals from a combined sample of autistic twins, hospital patients, and individuals from families with several members

affected by autism or related cognitive phenotypes. Bailey *et al.* (1993a) surmised that the similarity of the two findings supports the conclusion that the fragile X anomaly accounts for only a small proportion of the genetic influences in autism. It should also be noted that most medical centres currently test patients with autistic behaviour for fragile X. If positive, these individuals are usually diagnosed with fragile X or fragile X with autistic features instead of autism itself.

Neurocutaneous Disorders

Three neurocutaneous disorders have been described in association with autism or autistic symptomatology. These include tuberous sclerosis (mapped to chromosomes 9 and 16), neurofibromatosis (NF1 maps to chromosome 17 and NF2 to chromosome 22), and hypomelanosis of Ito (believed to reside on chromosome 15q) (Gillberg and Coleman, 1992). The association between hypomelanosis of Ito and autism has only been reported in case studies (Akefeldt and Gillberg, 1991; Zappella, 1993). Tuberous sclerosis and neurofibromatosis have been reported in epidemiological studies of autistic populations at rates greater than would be expected by chance.

Tuberous Sclerosis

Tuberous sclerosis (TSC) is an autosomal dominant disorder involving benign tumour growth in various organs, including skin, brain, eye, kidney, and heart (Gomez, 1988). It has a variable expression and exhibits genetic heterogeneity. This disorder, like fragile X, has been investigated in detail with respect to its association with autism. Smalley *et al.* (1992) reviewed the literature regarding the frequency of

autism in TSC and found rates ranging from 17-58% in studies using various diagnostic instruments. This related to a frequency of TSC among autistic individuals in the order of 0.4-3%. They also noted an increase in the estimated frequency (8-14%) in a subgroup of autistic subjects with seizures (Gillberg, 1991; Riikonen and Simell, 1990). Since then, three population-based studies of TSC have examined autism rates. Hunt and Shepherd (1993) found that 24% (5 of 21) of children with TSC in Western Scotland met the criteria for autism. These individuals were diagnosed using a 13-item scale, with a score of 7 or greater indicating autism. Gillberg *et al.* (1994), in their study in Western Sweden, found a rate of 61% (17 of 28) of individuals with TSC meeting the DSM-III-R (APA, 1987) criteria for autistic disorders. In a Californian study, Gutierrez *et al.* (1998) found the frequency of autism in their sample of 28 TSC probands to be 29%. These probands were all diagnosed using the revised Autism Diagnostic Interview, based on ICD-10 and DSM-IV criteria (see section 2.1.2; WHO, 1992; APA, 1994). In an overview of all of these epidemiological studies, Gutierrez *et al.* (1998) surmised that autism occurs in approximately 25% of individuals with TSC, with the majority of these people showing evidence of mental retardation as well. He also noted that infantile seizures and mental retardation are significant risk factors for the development of autism in TSC, for example in the Gillberg *et al.* (1994) sample, 13 of the 17 children with autism and TSC had a history of infantile spasms. Overall, if the population prevalence of TSC is taken to be 1/6,800 (Ahlsen *et al.*, 1994), it could be suggested that TSC accounts for up to 4% of all cases of autism.

Neurofibromatosis

Neurofibromatosis is also an autosomal dominant disorder with a variable severity of symptoms that primarily affect the nervous system. So far, two types of neurofibromatoses have been identified. Neurofibromatosis type 1 (NF1) has been mapped to chromosome 17 and occurs with an incidence of 1/3000 (Barker *et al.*, 1987). NF1 predisposes individuals to the development of skin spots (café au lait macules) and tumours within the peripheral nervous system (Rouleau *et al.*, 1993). Neurofibromatosis type 2 (NF2), which occurs with an incidence of 1/33000-50000 (Kanter *et al.*, 1980; Evans *et al.*, 1992), is characterised by a predisposition to the development of tumours of the central nervous system, including acoustic neuromas, meningiomas, neurofibromas and ependymomas (Martuza and Eldridge, 1988). The chromosomal region containing mutation(s) responsible for NF2 was narrowed, through a combination of family studies and deletion mapping in tumours, to chromosome 22q12 (Seizinger *et al.*, 1986; Rouleau *et al.*, 1987; Seizinger *et al.*, 1987; Couturier *et al.*, 1990; Rouleau *et al.*, 1989). Recently, two groups of investigators independently isolated a candidate gene for this disease (Rouleau *et al.*, 1993; Trofatter *et al.*, 1993). They showed that this gene, designated Schwannomin (SCH), is the site of the mutations causing NF2 by demonstrating germline and somatic SCH mutations in NF2 patients and in NF2-related tumours. Most of the variants were nonsense, frameshift, or splice mutations predicted to lead to the synthesis of a truncated SCH protein.

Gillberg and Forsell (1984) were the first to report an association between autism and neurofibromatosis. A systematic community study was carried out in individuals born in the Goteborg region of Sweden between the years of 1962-1976.

From this total population screening, 2 children out of the 26 diagnosed with infantile autism were reported to have NF1. This corresponds to a rate of 7.7%. Future studies, however, have failed to replicate these results. Fombonne *et al.* (1997) in the French epidemiological survey of autism noted that the rate of neurofibromatosis was very low. Their rate of 0.3% was similar to the median rate of four other surveys (reviewed in Fombonne, 1998) and identical with the rate found in a large sample of 341 referred autistic children in Denmark (Mouridsen *et al.*, 1992). This rate is not significantly different from what would be found by chance alone. Fombonne *et al.* (1997) concluded that, when these results are amalgamated, they strongly go against the claim of an association between autism and neurofibromatosis (Gillberg and Forsell, 1984).

Recent studies, therefore, show no specific association with neurocutaneous disorders and autism, however, they may be useful in the investigation of the neuropathology or pathophysiology of autism. Since both tuberous sclerosis and neurofibromatosis present with brain lesions that are variable in location, neuropathological studies comparing cases with and without associated autism may suggest brain areas that are focal points of the pathophysiology of autism (Folstein and Rutter, 1988; Hunt, 1995).

Metabolic Disorders

Disorders caused by a metabolic deficit have also been associated with autism at a level that is possibly greater than suggested by chance alone. These include phenylketonuria (PKU) (Friedmann, 1969), lactic acidosis (Coleman and Blass, 1985) and hypothyroidism (Gillberg *et al.*, 1992a). PKU was the first genetic condition to

be reported in association with autism. It is an inborn error of metabolism that is inherited in an autosomal recessive fashion, and is usually, but not always, associated with mental retardation (Folling, 1934). Since the initiation of programmes for the screening of newborns for PKU and the institution of a phenylalanine-restricted diet within the first weeks of life, the frequency of this disorder rapidly decreased. However, there have been reported cases of autistic children with classic untreated PKU (Miladi *et al.*, 1992; Chen and Hsiao, 1989; Pueschel *et al.*, 1985; Friedman, 1969; and others). After the initiation of a phenylalanine-restricted diet, the autistic symptoms improved (Friedman, 1969). From the above observation and after noting that autistic symptoms were present in children with PKU and normal intelligence, Friedman (1969) concluded that autism and mental retardation were separate disorders. This supports the hypothesis that, on a pathophysiological level, autism and mental retardation have different mechanisms (Folstein and Rutter, 1988). At present, PKU is the metabolic disorder most frequently reported to co-occur with autism, however this association is rare, as shown by a review of six recent epidemiological studies (Fombonne, 1998) which failed to report any single case.

1.5.2 Autism as a Genetic Disorder

It is generally considered that some cases of autism have a genetic basis (Bolton and Rutter, 1990) as shown by the data above and the publications reviewed below. The proportion of cases of autism with a genetic aetiology, however, is unknown. As mentioned above, the percentage of cases of autism associated with other disorders range from 10 to 25%, including 5 to 14% with a known genetic

disorder or chromosomal anomaly (Smalley, 1997). However, there is uncertainty as to the proportion of autism cases that have a genetic basis without demonstrable associated disorders.

Nevertheless, there is a general consensus as to the importance of genetic factors in the development of autism. Evidence that supports this view has been derived by family and twin studies, as well as the family aggregation of other conditions.

1.5.2.1 Family Studies

Publications by Smalley *et al.* (1988) and Bolton and Rutter (1990) reviewed a number of epidemiological and family studies of autism and estimated the prevalence of autism in siblings of autistic children to be about 2 - 3%. A more recent family study by Bolton *et al.* (1994) that examined 153 siblings of 99 autistic individuals found the sib recurrence risk to be 2.9%. Also, a study by Jorde *et al.* (1991), based on a large scale epidemiologic survey in Utah, reported a recurrence risk of 3.7% if the first autistic child was male and 7.0% for female probands. Studies have shown that parents with an autistic child may limit further childbearing since the disease is so severe. However, if families that have stopped further child-bearing after having a child with autism are removed from the family studies, the recurrence risk in siblings may be greater than the average noted by Smalley *et al.* (1988) and Bolton and Rutter (1990). Indeed, a family study by Ritvo *et al.* (1989) took into account the “stoppage rules” evident in families with autism and reported a risk of autism for siblings born after the proband of 8.6%. In summary, the recurrence risk in siblings is in the region of 2 - 9%. This represents a 45 – 200-fold increase in risk relative to the population prevalence (4.5 in 10,000, see section 1.4.1).

1.5.2.2 Twin Studies

Since family studies cannot separate genetic and non-genetic (environmental) influences, twin studies are also used for establishing a genetic aetiology. In twin studies, a disease that is at least partially determined genetically is expected to show a higher concordance rate among monozygotic (MZ) twins compared to dizygotic (DZ) twins. MZ twins normally have 100% of their genes in common, while DZ twins have, on average, 50% of their genes in common. This genetic relationship results in the phenotypic differences between MZ twins being attributed to environmental influences, while phenotypic differences between DZ twins are attributed to both genetic dissimilarity and environment. There have been three major twin studies of autism (Table 3). Ritvo *et al.* (1985a) found a MZ concordance of 95% (22/23) and DZ concordance of 24% (4/17), whilst Steffenburg *et al.* (1989) found a MZ concordance of 91% (10/11) and a 0% (0/10) concordance in same sex DZ twins. In a recent study that uses some of the same twins described in an earlier study by Folstein and Rutter (1977), Bailey *et al.* (1995) found a MZ concordance of 73% (27/37) and a DZ concordance of 0 (0/20). The first study by Ritvo *et al.* (1985a) was not systematic (the subjects were collected by advertisement) and there were also problems in the re-diagnosis of affected individuals (including failing to take into account the possibility of a broader phenotype). This may have inflated the concordance rates reported. Combining the data from the last two studies, the average MZ concordance rate is 77% and the average DZ concordance rate is zero. The DZ concordance rates were calculated from a small sample size and are not significantly different from the approximate sib recurrence rate of 3%. It may be concluded from these studies that the MZ concordance is much higher than the DZ

concordance rate, providing further support that genetic influences predominate in the aetiology of many cases of autism.

Table 3: Systematic Twin Studies on Autism

Study	MZ Twins	DZ Twins
Ritvo <i>et al.</i> (1985a)	95% (22/23)	24% (4/17)
Steffenburg <i>et al.</i> (1989)	91% (10/11)	0% (0/10)
Bailey <i>et al.</i> (1995)	73% (27/37)	0% (0/20)
(update of Folstein and Rutter (1977) twin study)		

Twin studies can also be utilised to estimate the heritability of liability to autism. Bailey *et al.* (1995) used their calculated MZ concordance rate of 73% and, since no DZ twins were ascertained during the course of the study, the 2.9% rate of autism among siblings of autistic individuals was applied (Bolton *et al.*, 1994). They also utilised two different prevalence estimates, 1.75 in 10,000 (Wing *et al.*, 1976) and 10 in 10,000 (Gillberg *et al.*, 1991a) in their calculations. Using structural equation modelling (Bentler, 1989), they estimated the broad heritability of liability to autism to be 93% and 91% for the base rates of 1.75 and 10 in 10,000 respectively. This again supports the role of genetic factors in the aetiology of autism.

1.5.2.3 Family Aggregation of Other Conditions

Clinical observations have noted the social peculiarities present in some parents of autistic children, as well as a history of developmental disorders of reading and spelling in parents as well as siblings of autistic probands (Folstein and Piven, 1991). Recent studies using standardised assessment procedures have reported that

cognition, language and personality deficits, and psychiatric disorders are present in families with autistic probands. These are discussed below.

Cognitive Deficits

A high rate of cognitive deficits (ie. reading, spelling and articulation disorders; language delay and mental retardation) has been noted in families of autistic children. A study by Bartak *et al.* (1975) found that reading or language disabilities were present in first-degree relatives in 26% of families. Folstein and Rutter (1977) reported similar results in their twin study. They found that the concordance rate for cognitive deficits was 82% for MZ twins and 10% for DZ pairs. A twin study by Le Couteur *et al.* (1989a) replicated Folstein and Rutter's results and suggested that cognitive abnormalities might be a milder expression of a common underlying genetic abnormality in autism.

Piven *et al.* (1989a) interviewed mothers of autistic children (N=37) about their other children (N=67). Fifteen percent of non-autistic siblings were reported to have had cognitive disorders as children. It was also noted that most of these cognitive problems were in reading and spelling, and none of these non-autistic siblings were mentally retarded.

Specific Language Abnormalities

Landa *et al.* (1989) assessed social communication in parents of autistic individuals and control subjects matched for IQ scores and educational level. They reported a significantly higher rate of deficits in both social use of language (pragmatic language) and in story telling (narrative discourse) in the parents compared to the controls.

Personality Characteristics

Particular personality characteristics have been reported in parents and siblings of autistic probands, at a rate greater than expected by chance alone. Kanner and Eisenberg (1956) noted that the parents in their study were perfectionists with interests in abstract ideas and lacked interest in developing relationships with others. Although they interpreted these findings as being evidence that “faulty” child rearing was the cause of autism, this has later been disproved (Cantwell *et al.*, 1978). Recent studies using the family history method (Macdonald *et al.*, 1989; DeLong and Dwyer, 1988; Wolff *et al.*, 1988) and a standardised interview (Piven *et al.*, 1989b), have shown an increase of incidence of social deficits in these families.

Psychiatric Disorders

Several studies have also reported an increased rate of psychiatric disorders in first-degree relatives of autistic probands (Piven *et al.*, 1991b). Major affective disorder (Piven *et al.*, 1989a; DeLong and Dwyer, 1988) and anxiety disorders (Piven *et al.*, 1989a) were found to have significantly higher rates compared with control groups. Since these disorders occurred, in most cases, prior to the birth of the proband, they could not be attributed to the stress of having an autistic child. Schizoaffective disorders have been diagnosed in a number of families in which the probands have autism and fragile X syndrome (Hagerman, 1989; Gillberg, 1992b). Finally, there have been several reports on the co-existence of autism and elective mutism within first-degree relatives (Gillberg *et al.*, 1992b).

A variety of different abnormalities aggregate in families of autistic individuals as shown by the above studies examining cognitive deficits, personality

characteristics, language, and psychiatric disorders in relatives of autistic individuals. It has been suggested (Folstein and Piven, 1991) that these deficits may be manifestations of an underlying genetic liability to autism or part of the broader phenotype, since they are often mild and share similar characteristic features of autism (eg. aloofness and pragmatic language deficits). These deficits were not explored further in this project, due to the small number of studies published on this topic and the diagnostic problems, at present, in separating individuals with these deficits from normal controls.

1.5.2.4 Summary of Genetic Involvement in Autism

In summary, there are a few specific genetic conditions that can be associated with autism. Among cases of unknown aetiology, the evidence shown above proposes a 50-200-fold increase in genetic liability to autism in siblings of autistic probands, relative to the population prevalence. Family studies have also indicated that both parents and siblings have a higher liability for social and cognitive deficits that are milder but conceptually similar to those found in autism. Whilst other factors may alter this underlying genetic liability such as sex, IQ, and prenatal and perinatal injury, the overall evidence suggests that the liability to autism is largely genetically determined. This hypothesis was examined in this study by searching for susceptibility loci involved in the aetiology of autism.

1.6 Neurobiological Factors

There is substantial evidence that autism is associated with abnormalities in the brain. However, the link between neurobiological findings and symptomatology in this genetically influenced developmental disorder has yet to be identified. The neurobiology of autism has been studied in great detail and there is an extensive literature on the subject, which is often complex and inconsistent. The poor replicability of neurobiological investigations arises for at least two reasons. First, there is lack of methodological consistency between studies. Results seem to be dependent upon the comparison group used and the variable on which the groups are matched (Dawson, 1996). Second, immense heterogeneity exists in the behavioural expression of autism, and studies differ in the subject populations they examine. Variability in clinical diagnosis and differences in the specificity of individual diagnostic instruments used in each study can also lead to heterogeneity of the sample and contribute to the differences observed (Bailey *et al.*, 1996).

Current views about the localisation of abnormalities in the brain associated with autism are the product of neuropsychological, neuroanatomical (including neuroimaging and postmortem studies), and neurochemical studies. As previously mentioned, there exists a vast amount of literature on the neurobiology of autism. The majority of this is not within the scope of this project and only areas relevant to the overall understanding of this study will be reviewed.

1.6.1 Neuropsychological

Indirect evidence of abnormal brain function has been derived from neuropsychological studies (Happe, 1994a). These investigations have been uniform in finding deficits in certain aspects of higher order cognitive functions, including abstract and pragmatic language, encoding of complex information, and executive (frontal) functions (Bristol *et al.*, 1996). Deficits in executive functioning in autism include impairments in thinking about consequences and the context of the activity, and in planning for and appropriately solving problems in a flexible, efficient manner. It has been proposed that these impairments result in the repetitive behaviours and inflexible style of thinking characteristic of autism (Ozonoff *et al.*, 1991). The universal impairment in social cognition found in neuropsychological studies of autism suggests the involvement of certain brain regions known to mediate social and emotional behaviour, namely, regions of the limbic system, such as the amygdala and orbital frontal cortex. Other studies have found an association between “theory-of-mind” impairments and autistic symptoms (McEvoy *et al.*, 1993; Ozonoff *et al.*, 1991). “Theory of mind” abnormalities refer to the inability to take notice of, and think about, the mental states of others (Frith, 1989), and may be related to the deficits in social interaction and social use of language found in autism (Lainhart and Piven, 1995). This theory has resulted in a great deal of research over the past 5-10 years to determine how it relates to autism, and has been revised recently to accommodate the large amount of neurobiological research. These studies have shown autistic children to be impaired on a variety of experimental tasks requiring theory of mind for successful solutions (Baron-Cohen *et al.*, 1985; Happe, 1994b).

1.6.2 Neuroanatomical

Mental retardation is present in approximately three-quarters of autism cases (Rutter, 1979). Mental handicap syndromes have often been associated with gross abnormalities in brain development, with a reduction in brain size frequently observed, as well as abnormalities in the cerebral hemispheres and overlying cortex (Bailey *et al.*, 1996). In turn, it has been proposed that autism may be associated with gross abnormalities in brain development and that one such abnormality may be responsible for the distinctive symptomatology present in autism. A number of structural imaging and postmortem studies have been conducted to test this theory.

1.6.2.1 Neuroimaging Studies

In vivo structural and functional imaging studies in autism have shown that the most consistent neuroanatomical lesions involve the cerebellar vermis (Courchesne *et al.*, 1994). Magnetic resonance imaging (MRI) studies, in particular, have shown a reduction in the cross-sectional area of the cerebellar vermis (hypoplasia) (Courchesne *et al.*, 1987; Gaffney *et al.*, 1987; Courchesne *et al.*, 1988; Hashimoto *et al.*, 1995). The largest of these studies was by Hashimoto *et al.* (1995), who found hypoplasia of lobules VI and VII when a group of 102 people with autism were compared to 112 control subjects. This difference was also found in children less than 2 years old who were later diagnosed with autism. Studies by other groups using similar anatomical methods have failed to replicate these findings (Garber and Ritvo, 1992; Holttum *et al.*, 1992; Kleiman *et al.*, 1992; Piven *et al.*, 1992). Reasons given for the discrepancy in the “failed” replication studies include small sample size and the use of patient controls versus healthy volunteer controls.

Another MRI study has suggested that minor malformations of the cerebral cortex may be important in a subgroup of autistic individuals (Piven *et al.*, 1990). Also, a positron emission tomography study showed that the functional relationship between the frontal and parietal cortex and subcortical structures in autistic subjects appear to differ from normal (Horwitz and Rumsey, 1994). These findings correlate with the severity of autism and specific clinical deficits and may be related to less efficient information processing (Minshew, 1994).

Abnormalities in the medial temporal lobe and related structures have also been associated with autism. Hetzler and Griffin (1981) and DeLong (1992), recognised the similarities between autistic behaviour and Kluver-Bucy syndrome, and characterised autism as the effect of developmental dysfunction in structures of the medial temporal lobe, especially the hippocampus and amygdala. This was supported by Bachevalier's (1994) finding that lesions of medial temporal lobe structures induced in newborn monkeys result in behaviour resembling autism, whereas no such changes were noted to result from similar lesions induced in adult monkeys.

A number of neuroimaging studies have also reported both indirect and direct evidence of enlarged brain size in autism. These studies have shown: (1) head circumference greater than the 97th percentile in approximately 20-40% of autistic people (Bailey *et al.*, 1996; Davidovitch *et al.*, 1996; Woodhouse *et al.*, 1996; Lainhart, 1997); (2) increased brain weight at autopsy (Bailey *et al.*, 1993b) and; (3) increased brain volume (Piven *et al.*, 1995; Filipek, 1996). It is surprising to note that a number of disorders reported to be associated with autism (eg. tuberous sclerosis, neurofibromatosis and hypomelanosis of Ito; see above), are also occasionally associated with increased brain size or head circumference (Bailey *et al.*, 1996).

1.6.2.2 Post-Mortem Studies

To date, fewer than 35 brains from autistic individuals have been examined, and none of these with current immunocytochemical techniques for synaptic, cytoskeletal, or glial markers. The early studies suggest abnormal neuronal development, evidenced by reduced neuronal cell size and increased cell packing density of limbic system structures, with decreased dendritic branching in the hippocampus, and reduced numbers of Purkinje cells in the cerebellar hemispheres (Bauman, 1996; Raymond *et al.*, 1996). An investigation of a single autistic brain by Rodier *et al.* (1996) also suggested a decrease in the number of neurons in the motor cranial nerve nuclei. These findings have not been associated with seizures, use of medication, nor the severity of clinical symptoms. Since Purkinje cell loss is a recognised complication of epilepsy (Meldrum and Bruton, 1992), and autopsy studies have lacked mental retardation controls, the degree to which any finding is specific to autism remains unknown (Bailey, 1993).

1.6.3 Neurochemistry

Neurochemical studies of neurotransmitters and immunological abnormalities can provide much information for most neurobiological studies. Interpretation of the results, however, is problematic since the biological variable being assayed is usually rather remote from either brain function or structure. Also, in the case of autism, it is difficult to know whether the difference between subjects and controls is due to aetiological factors or consequences of some aspect of autistic behaviour or cognition (Bailey *et al.*, 1996). Many changes in neurochemicals have been associated with

autism (eg. serotonin (Gillberg and Coleman, 1992), dopamine (Cook, 1990), norepinephrine (Gillberg and Svennerholm, 1987), endogenous opioids (Sahley and Panksepp, 1987) and glial proteins and gangliosides (Ahlsen *et al.*, 1993)), however, most of these findings have been inconclusive.

The most commonly replicated biological abnormality in autism is hyperserotoninaemia. The interest in serotonin is due to its effects on central nervous system (CNS) development, social behaviour, sleep, aggression, anxiety, and affective regulation (Cook and Leventhal, 1996). There is a significant amount of evidence that serotonin in the blood is raised (hyperserotoninaemia) in approximately one-third of patients with autism, and that this increase is due to increased platelet serotonin (Gillberg and Coleman, 1992). However, since increased blood levels of serotonin have been found in about one-half of non-autistic severely retarded children, it is still not clear whether this finding is only obtained in mentally retarded autistic individuals (Gillberg, 1993). Recent studies from behavioural neuroscience, pharmacological, and genetic studies have indicated that serotonin is involved in many of the symptoms of autistic disorder, such as aggression, anxiety and inflexible rituals and routines (Cook and Leventhal, 1996). These results led Cook *et al.* (1997b) to examine a possible linkage between the serotonin gene and autism (see below). Advances in the molecular biology of serotonin-related pathways are likely to provide an increased understanding about the underlying neurobiological mechanisms in autism and help in developing appropriate treatment of the social and cognitive dysfunction associated with autism (Piven, 1997).

1.6.4 Summary of the Brain Mechanisms Involved in Autism

The available evidence suggests that autism involves widespread brain dysfunction at both the cortical and subcortical levels, however the primary site of the brain injury has not as yet been identified. Investigators have been recently debating whether there is only one pathway in the brain that is not functioning normally due to one or multiple mutations and this is what is causing autism, or whether there are multiple deficits that, in combination, cause what is known as autism. Bristol *et al.* (1996), in a report to the National Institutes of Health on the state of the science in autism, stated that “the pathophysiology of autism, or the structural and functional abnormalities of the brain and how precisely they result in the abnormal behaviour of autism is far more complex than what brain structures or neurochemicals are involved. Each level of analysis is highly complex and, at present, only pieces of this puzzle in autism have been identified.”

1.7 Mode of Inheritance

Given the evidence of a genetic predisposition to autism, studies on the familial segregation of the disorder have been undertaken in order to identify a specific mode of genetic transmission. Segregation analysis can be used to estimate key parameters such as the allele frequency, penetrance, and proportion of cases explained by the proposed model of inheritance. However, problems with segregation analysis exist, especially for genetically complex disorders. A complex disorder is one that does not exhibit classical Mendelian recessive or dominant inheritance attributable

to a single gene locus. First of all, segregation analysis can be extremely sensitive to ascertainment biases (eg. the preferential inclusion of affected individuals may cause the penetrance to be greatly overestimated). Secondly, it may not be able to distinguish between the many possible modes of inheritance for complex traits. Some examples of these include: (a) a single gene disorder with reduced penetrance, (b) heterogeneity (caused by two or more independent genes), (c) oligogenic inheritance (interaction of a small number of different genes with at least one of moderate effect on the disorder), and (d) polygenic inheritance (interaction of a large number of different genes, each with a small effect on the disease). Thirdly, the number of genes influencing the trait is especially difficult to estimate (Lander and Schork, 1994).

Many genetic models have been postulated for autism. These include autosomal recessive (Ritvo *et al.*, 1985b), polygenic (Jorde *et al.*, 1991; Tsai *et al.* 1981), and X-linked (Gurling, 1986; Szatmari and Jones, 1991).

Ritvo *et al.* (1985b) applied simple segregation analysis to a sample of 46 multiplex families (with two or more autistic children). Multiplex families were selected in preference to sporadic cases, to minimise the probability of detecting single events due to environmental incidents, phenocopies, or new mutations (Smalley *et al.*, 1988). The authors found a segregation ratio of 0.19 ± 0.07 , which is consistent with that expected under autosomal recessive transmission (ie. 0.25). While one would expect the frequency of affected sib-pairs for an autosomal recessive inheritance to be around 25%, the “stoppage rules” prevalent in autistic families could account for the low sibling recurrence observed (~3%). Indeed, when Jones and Szatmari (1988) reanalysed these data, adjusting for truncated family size due to “stoppage”, the results were also consistent with an autosomal recessive model of inheritance (0.27). The results obtained by Ritvo *et al.* (1985b) are inconclusive, however, due to the

biased nature of the sample (which was ascertained in part by advertising for multiple-incidence families) and uncertainties about diagnosis (Bailey *et al.*, 1996).

A further study by the same research group (Jorde *et al.*, 1991), calculated a different result. This investigation carried out a complex segregation analysis using data collected in the UCLA-Utah Epidemiological Survey. The uniform collection of data helped to circumvent difficulties caused by selective ascertainment and thus produced a sample more appropriate for segregation analysis. In this segregation analysis they used a mixed model, compared to the simple segregation analysis employed in the previous study. The authors concluded that a single major locus model of inheritance is unlikely, but could not rule out a multifactorial threshold model (ie. a large number of genes or environmental effects contributing to an underlying liability). However, they also noted that a major gene could be segregating in a small proportion of families that could not be detected by the mixed-model segregation analysis, even with the large data set used. The revised analysis was repeated assuming that eleven parents initially considered unaffected were affected (these individuals showed clinical deficits in the three major areas characteristic for autism). This analysis resulted in a shift in the model to a dominant major single locus. It could be concluded that, to predict the exact mode of inheritance of autism, accurate diagnostic information is required on the parents. Since autistic individuals generally do not marry and there are no clear-cut diagnostic criteria for milder forms of autism, this could prove very difficult.

X-linked recessive inheritance has also been proposed (Szatmari and Jones, 1991), which could explain the increase in risk in males compared to females (3:1). Again, since it is difficult to diagnose parents, it is hard to assess directly the lack of male-to-male transmission.

The excess proportion of males affected with autism could also be accounted for by the following models (Szatmari and Jones, 1991):

1. Autosomal inheritance, with a lower penetrance for females compared to males;
2. Polygenic inheritance, with a higher genetic load required for females to express the phenotype;
3. X-linked recessive inheritance for some individuals and an environmental cause for others (mixed aetiology);
4. X-linked recessive or autosomal inheritance (genetically complex aetiology).

The results from these segregation studies on autism have been inconclusive. Even though modifying factors, such as sex-influenced inheritance, reduced penetrance, and variable expression, could account for some of the deviations present, overall a multifactorial model (with the interaction of multiple genes) is more likely than a monogenic model (Bailey *et al.*, 1996). This conclusion was drawn from two observations (Bailey *et al.*, 1996). First of all there is a significant difference between the concordance rate of MZ and DZ twins (Bailey *et al.*, 1995; Steffenburg *et al.*, 1989; Folstein and Rutter, 1977). If it is assumed that the MZ concordance rate is 77% and the DZ concordance rate is 3% (see section 1.5.2), this gives a MZ to DZ concordance ratio of $77/3 = 25$. This high ratio cannot be explained by one or two loci predisposing to the disorder: if a single recessive gene was responsible, a ratio of only 4 ($100/25 = 4$) would be calculated; and a dominant gene would contribute even less (ie. $100/50 = 2$ or $100/75 = 1.3$). There is also a larger drop in cases of autism when looking at second-degree and third-degree relatives compared to first-degree relatives (DeLong and Dwyer, 1988; Jorde *et al.*, 1991; Szatmari *et al.*, 1995). For any disorder caused by a single major gene, the risk to relatives should fall by roughly a half as genetic distance increases (Risch, 1990b). Combined data from the four

studies gives a 0.18% risk of autism for second degree relatives with a 0.12% risk for third-degree relatives. However, these data were collected from family histories which may underestimate the risk of disorder in extended relatives (Weissman *et al.*, 1986).

Secondly, Pickles *et al.* (1995) estimated that the number of genes responsible for autism, using latent class analysis, ranged between two and ten, however a three gene model was the most likely. Similar results were also obtained in simulations performed by our group, using the same approach but with individuals diagnosed with a "core", less broad definition of autism (Risch, unpublished). The population prevalence was assumed to be 1/2500, the MZ concordance rate was 77% and the sib risk value was presumed to be 3%. These figures lead to a MZ twin concordance to population prevalence ratio $\lambda_M = 1925$, and a sib risk ratio $\lambda_S = 75$ (see section 1.8.3.1). Assuming a multiplicative model and fixing the value of λ_S , the value of λ_M was predicted under a number of epistatic genetic models (Table 4). As can be seen from the table, models specifying only 1 or 2 contributing loci or infinite contributing loci give a poor fit to λ_M . Models with three or more loci begin to provide an adequate fit to λ_M , depending on the degree of recessivity. In turn, these simulations showed that numerous plausible genetic models fit the data used, however, there must be at least 3 contributing susceptibility loci and the number depends on the degree of recessivity.

In conclusion, it is clear that autism is not a single major gene disorder and the data at present point to a multifactorial complex disorder with many genes implicated.

Table 4: Predicted Values of the Population Prevalence Ratio (λ_M) (Observed = 1925), Fixing $\lambda_S = 75$ for Several Multiplicative Multilocus Models (Note: loci are referred to as additive, recessive or intermediate, depending on the effect of that particular locus/loci).

Genetic Model	Predicted λ_M
1 locus, additive ($\lambda_S(i) = 75$)	149
1 locus, recessive ($\lambda_S(i) = 75$)	299
Infinite loci, additive	5625
2 loci, additive ($\lambda_S(i) = 8.66$ each)	266
2 loci, recessive ($\lambda_S(i) = 8.66$ each)	1132
3 loci, additive ($\lambda_S(i) = 4.22$ each)	411
3 loci, recessive ($\lambda_S(i) = 4.22$ each)	3996
3 loci, intermediate ($\lambda_S(i) = 4.22$ each)	1582
6 loci, additive ($\lambda_S(i) = 2.05$ each)	900
6 loci, intermediate ($\lambda_S(i) = 2.05$ each)	2222
1 locus, additive ($\lambda_S(i) = 4$), infinite others	2461
1 locus, additive ($\lambda_S(i) = 6$), infinite others	1719
2 loci, additive ($\lambda_S(i) = 4$), infinite others	1846

1.8 Linkage Analysis

The convincing evidence of a genetic component in the liability to autism makes the disease an obvious candidate for molecular genetic strategies (Rutter *et al.*, 1994). These strategies search for susceptibility genes to a disorder, which could be particularly helpful in elucidating the biological basis of autism since biochemical and

neuroanatomical studies have been inconclusive. Presently, there are two major approaches that use markers to identify genes underlying human diseases. The first is “genetic linkage analysis”. This method is most commonly used for determining the chromosomal location of genes for inherited disorders, where the causative genes are unknown. The second is the “candidate gene” method, which presupposes that there is prior information on pathogenesis pointing to specific genes where variation is likely to occur. The frequency of different alleles in the candidate gene is in turn compared between affecteds and controls using statistical tests for association. There are some diseases, mutations and specific alleles that have been shown to co-occur with autism (eg. Fragile X, chromosome 15 deletions, HLA alleles). Although the traditional candidate gene approach was not used in this study, these “candidate regions” were investigated in greater detail in this project, using genetic linkage analysis to identify susceptibility loci.

1.8.1 Principles of Genetic Linkage Analysis in Mendelian Inheritance

In principle, genetic mapping of disease loci involves locating the relative position of a gene or genetic marker within the genome, by measuring its tendency to segregate together with some other measurable trait through several generations. The method of mapping genes by virtue of their co-segregation with other loci on a chromosome through meiosis is termed genetic linkage analysis. Thus, genetic mapping is a description of a gene’s meiotic behaviour rather than its physical location (Thompson *et al.*, 1991). Genetic linkage analysis is a tremendously important and powerful approach in medical genetics because it allows mapping of genes, such as

some disease genes, in which neither their biochemical nor their molecular basis has yet been detected (Thompson *et al.*, 1991).

The production of accurate and high-resolution genetic linkage maps is necessary for positional, as opposed to functional, cloning of genes. “Functional cloning” refers to the identification of a disease gene based on fundamental information about the basic biochemical defect, without reference to its chromosomal map position. Pure “positional cloning” assumes no functional information, and the responsible gene is located solely on the basis of genetic map position with the use of linkage analysis (Collins, 1995). Even before the identification of a disease gene, its placement on a genetic map is useful for identifying susceptible individuals and for genetic counselling (Cox Matise *et al.*, 1994). The unit of measurement for genetic linkage is the genetic length of a chromosome over which, on average, one recombination event is observed (“recombination” referring to the process of crossing over between homologous chromosomes during meiosis). This unit is called a Morgan (in honour of Thomas Morgan who discovered the process of crossing over in 1910). A centiMorgan (cM) is the genetic length over which one observes recombination in one percent of recombination events or meioses. In the human genome, one cM translates roughly into one million base pairs (bp), when taking into account the physical length of the genome (3×10^9 bp) (Thompson *et al.*, 1991).

In principle, the smaller the genetic distance between two loci, the greater the probability that they will be linked and transmitted together to the next generation. This goes against Mendel’s law of independent assortment which states that an individual’s genes will be transmitted to the next generation independently of one another. Mendel’s law holds true for most pairs of loci, except those that occupy the same region of a chromosome. Under the process of independent assortment,

recombinations in a particular chromosome occur 50% of the time. A recombination frequency less than 50% or a genetic distance less than 50cM between two loci on the same chromosome, infers that these two loci are linked.

Linkage studies look at the transmission of alleles in a family to estimate the recombination frequency between known loci and the disease gene. The direct method of linkage analysis involves observing and counting recombinants and nonrecombinants. The recombination fraction is then calculated by dividing the number of observed recombinants (k) by the number of opportunities for recombination (n):

$$\theta = k/n$$

To do this, it is important to know the linkage phase, ie. the chromosome on which each allele is located. Generally, this can only be known in three or greater generation families. For example, Figure 2 shows a three generation family with individuals affected with the dominant disorder NF1 (shaded symbol). The genotypes for a two-allele marker locus (1F10) are shown below each individual in the pedigree. Since the genotypes of the grandparents are known, the linkage phase of the grandchildren is known. Allele 1 seems to co-segregate with the disease phenotype (ie. this marker is linked to the disease), therefore children in this family who are affected with NF1 usually have allele 1 at 1F10. In seven out of eight children in generation three this seems to be correct, however in one child a recombination has occurred (individual III.6). The recombination frequency for this family is, therefore, $1/8 = 12.5\%$, ie. the marker and the disease locus are approximately 12.5cM apart.

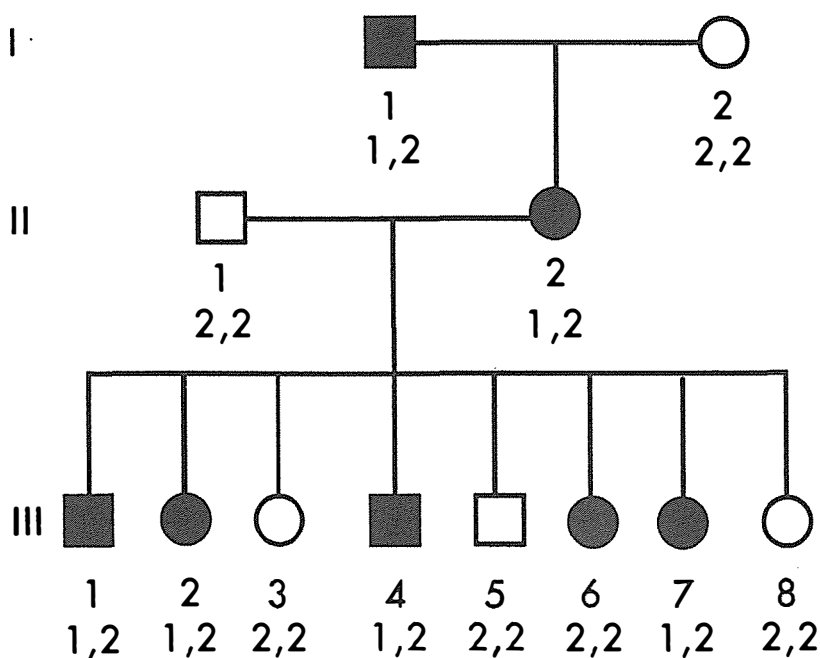


Figure 2: An NF1 pedigree in which each member has been typed for the 1F10 polymorphism (A). A recombination has occurred at individual III-6. (from Jorde, Carey and White, 1995)

When the parental phase is unknown the recombination fraction is estimated using the method of maximum likelihood (Fisher, 1921). In this method, the likelihood (ie. the probability of occurrence of the data) is calculated for a variety of assumed values of θ , and the θ value associated with the highest likelihood is selected as the best estimate (θ_{\max}). A lod score (Z ; Bernard, 1949) is the logarithm (\log_{10}) of the likelihood that two loci are linked at a given recombination frequency (θ), versus the likelihood that the two loci are not linked ($\theta = 0.5$). The likelihood principle is defined as:

$$Z(\theta) = \log_{10} (L(\theta) / L(\theta=0.5))$$

The lod scores are then graphed against the θ values, as shown in Figure 3. The estimate of the recombination fraction is, in turn, that value of θ at which $Z(\theta)$ is highest (Z_{\max}).

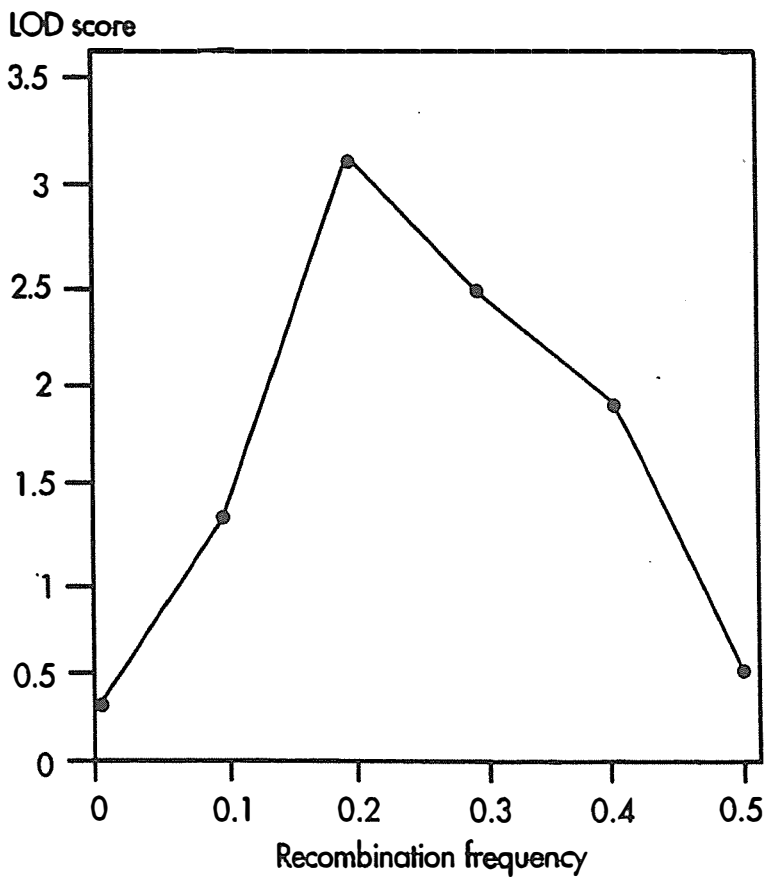


Figure 3: The lod score is plotted against the recombination frequency to determine the most likely recombination frequency for a pair of loci (from Jorde, Carey and White, 1995).

Positive lod scores indicate evidence for linkage, and negative lods indicate absence of linkage. If linkage exists between the two loci, the lod score is expected to become larger with the addition of more families. Conventionally, a lod score of 3.0 is the threshold used to declare linkage, since a score of 3.0 indicates that the likelihood in favour of linkage is 1000 times greater than the likelihood against linkage. Conversely, a lod score less than -2 (odds of 100 to 1 against linkage) is considered to be evidence that two loci are not linked. Morton (1955) originally proposed these particular values. In calculating these values, Morton used a power (to detect linkage if present) of 0.99 at the chosen recombination fraction ($\theta < 1/2$) and a significance level (probability of a false-positive result) of 0.001. His choice of power and significance level was determined to ensure that the rate of false positive results stay below 5 percent. The significance level of the test for linkage (0.001) is much smaller than the significance levels of 0.05 or 0.01 customarily used in statistical tests. The reason for this is that a large proportion of statistically significant false positives are calculated using these values. When using a significance level of 0.001, all but a few percent of results are real. Whilst this holds true in a sequential test (the sampling of a number of small families in a sequential manner), in large-sample theory (use of a large amount of informative data in the analysis) the rate of false positive results increases and the lod score drops from 3 to 2.07. To obtain a lod score of 3 in large sample theory, a significance level of $\alpha = 0.0001$ is required (Morton, 1998). This relates to a reliability of 0.996. Overall, in genetic linkage analysis a lod score of 3 is commonly used to declare a significant result of linkage between two loci.

A linkage map consisting of polymorphic DNA markers is the most appropriate map, as it allows the inheritance of different alleles of each marker to be traced in the families (informative genetic markers are ones that are highly

polymorphic and are heterozygous in a large proportion of individuals). The closer the genetic distance between the marker and the trait, the lower the probability that a meiotic crossover between the two loci will occur to prevent co-inheritance of the genetic marker (Thompson *et al.*, 1991). Linkage maps at 5-10 cM resolution are sufficient to identify the map location of diseases with Mendelian patterns of segregation, but for polygenic disorders which are characterised by multiple loci, higher resolution maps may be necessary (Cox Matise *et al.*, 1994). When linkage analysis is carried out as a sequence of pairwise comparisons between a trait locus and each of a number of marker loci, this is referred to as “two-point” analysis. For this type of analysis, lod scores are computed for each comparison (ie. trait vs. *i*th marker or marker vs. marker) and combined with the results of the same comparison from other families and/or investigators. With the production of dense genetic maps, it has become possible to analyse several linked loci simultaneously. This is referred to as “multipoint” linkage analysis (Lathrop *et al.*, 1984) and it is shown to be at least 5 times as efficient as two-point analysis in estimating the recombination value (Lathrop *et al.*, 1985) (see section 1.8.3.2). Once the allele(s) of genetic markers are shown to co-segregate with the disease phenotype, these markers provide information on the position of the gene in the genome. With refined experimentation, the disease gene can be located more precisely.

The prerequisites for linkage analysis in mapping disease genes responsible for Mendelian disorders are that the locus linking to the disease must be present in different forms in the population (ie. polymorphic genetic markers), and the availability of a sufficient number of informative meioses where the same disorder segregates. The following sections explore these prerequisites of linkage analysis for a Mendelian trait.

1.8.1.1 DNA Polymorphism

The term genetic polymorphism is used to describe variation in DNA sequence. Polymorphic loci are characterised by a number of different alleles, which allows members of a population to be categorised into distinct genotypes. Polymorphic loci occur mainly throughout regions of the genome that do not contain genes. Genetic polymorphism is, in turn, defined as the occurrence of multiple alleles at a locus, where at least two alleles appear with frequencies greater than one percent. By convention, then, polymorphic loci are those at which at least two percent of the population are heterozygous (Ott, 1992). Polymorphic markers with a high heterozygosity index (ie. the percentage of individuals that are expected to be heterozygous at that locus) are desirable, as the presence of multiple alleles increases the chance that the segregation of normal and disease-carrying chromosomes can be traced within the pedigree.

1.8.1.2 Polymorphic DNA Markers

Several types of DNA polymorphism have been utilised as genetic markers. These include (i) single nucleotide substitutions and (ii) repetitive DNA sequences. Single nucleotide polymorphisms (SNPs) have recently been systematically identified within the mapped sequence tagged sites (STS) distributed over the whole genome. One type of SNP are restriction fragment length polymorphisms (RFLPs) and these are detectable by restriction enzyme analysis (Kan and Dozy, 1978). The variations within these single nucleotides are usually biallelic (2 common alleles) and they have the potential for a +/- assay. These polymorphisms are also particularly suitable for genomic analysis using microarrays, due to their simple structure. Microarrays, or DNA chips, can contain as many as 3000 SNPs (Sapolsky *et al.*, 1999) and would

therefore potentially allow whole genome analysis of many DNA samples in a relatively fast and economic way. SNPs, however, are only about a third as informative as microsatellites (see below) and a complete map will probably not be available until the entire genome is sequenced (Tynan, 1998).

Tandem repeat sequences are sites where the same short DNA sequence is repeated a number of times. These repeats are thought to arise by slippage during replication and unequal crossing over (Efstratiadis *et al.*, 1980; Jeffreys *et al.*, 1986; Levinson and Gutman, 1987). Replication errors during cell division can cause the number of repeats to change. The frequency of this kind of replication error is high enough to make alternative repeat lengths at the site common, but the rate of change in the length of repeats is low enough that the size of the repeat sequence at the site serves as a stable polymorphic trait in family studies (Housman, 1995).

Tandem repeat DNA elements include satellite DNA, minisatellite DNA and microsatellite DNA. The polymorphism present in satellite DNA (which is present in heterochromatin) is in the length of the clusters of tandem arrays, which can be up to several megabases in length (Cooper and Krawczak, 1993). A number of different families of satellite DNA tandem arrays have been identified, including a simple sequence of 5-25 bp repeats, alphoid (171 bp repeat) and Sau3A (~68 bp repeat) (reviewed by Vogt, 1990). Minisatellite sequences, which are also known as variable number of tandem repeat (VNTR) polymorphisms, are variations in the length of stretches of DNA containing a variable number of units of the same short sequence. This core consensus sequence is GGTGGGCAGARG, where R is equal to a purine, and there is thought to be approximately 10^4 copies of VNTRs in the human genome. Finally, microsatellites are short (less than 500bp) tandem repeats, interspersed throughout the genome with sequence motifs of one to five nucleotides (Hearne *et al.*,

1992). The most abundant of these are the (CA/GT)_n dinucleotide repeats, with approximately 50,000 copies per haploid genome (Weber and May, 1989).

Microsatellites are the most useful of these DNA polymorphisms as they are easily isolated and characterised, and they display considerable polymorphism due to the large variation in the number of repeat units. Microsatellites are therefore the most ideal markers, at present, for the construction of high-resolution genetic maps, which can then be used to identify susceptibility loci involved in common genetic diseases. RFLPs, on the other hand, exhibit a low degree of polymorphism, and VNTRs are not evenly distributed throughout the genome, with clustering in the telomeric and centromeric regions (Royle *et al.*, 1988), making them less useful in the construction of genetic maps.

The first genetic linkage map in man was constructed using RFLPs (Donis-Keller *et al.*, 1986). Botstein *et al.* (1980) were the first to propose that differences in the DNA sequence should be treated like allelic variants of a gene, allowing their use as genetic markers for gene mapping. With the rapid rate of increase in the number of polymorphic markers identified (Figure 4), many laboratories have used these markers to construct genetic maps of whole chromosomes or particular portions of a chromosome (NIH/CEPH, 1992; Weissenbach *et al.*, 1992). Due to the number of CA repeats dispersed throughout the genome (~50,000; see above), these new genetic maps consist mainly of these types of microsatellites. The Human Genome Project (HGP) has completed mapping the entire human genome with polymorphic markers at 1-cM intervals. This project, which officially began in 1990, also aims at locating all the 60,000 to 70,000 genes of the haploid genome and sequencing the entire 3 billion or more nucleotides within it.

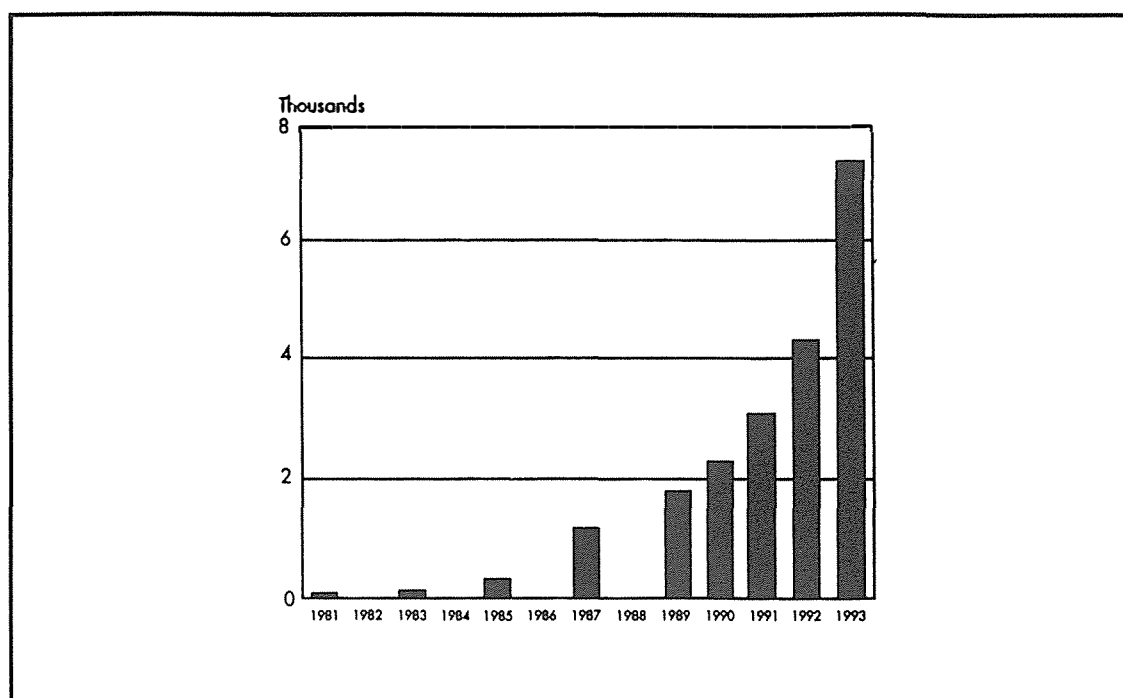


Figure 4: The number of known polymorphic markers from 1981 till 1993 (from Jorde, Carey and White, 1995).

1.8.1.3 Family Material

As mentioned previously, the possibility of obtaining a significant result in genetic linkage studies relies on the availability of enough polymorphic markers and the number of informative meioses. With recent progress in recombinant DNA technologies and the completion of stage I of the Human Genome Project, which aimed at constructing the genetic map, the availability of polymorphic markers poses no significant problem. Genome-wide scans of the entire human genome with microsatellite markers in informative families have been widely performed. Indeed, a large number of important disease genes have been localised using this approach and many have been cloned (eg. Early Onset Breast Cancer [Hall *et al.*, 1990], neurofibromatosis type II [Rouleau *et al.*, 1993]). The availability of sufficient family material is, however, a crucial point, and it is important to know how many informative meioses are required before attempting a genome search by linkage

analysis. This can be calculated by means of computer simulations, in which segregation of a potentially informative marker is simulated in the available family material. The simulation also estimates the expected and maximum lod-score which could be attained in the selected number of families if the marker was linked to the disease gene at a certain recombination fraction (estimate of the power of the proposed study), and if the marker was unlinked to the disease gene (estimate of the false positive rate). Traditionally, simulation analysis also requires the genetic model (ie. the mode of inheritance) to be specified, and the results obtained are only valid within the context of this model. It is important to specify an accurate genetic model, since mis-specifying genetic parameters may cause a loss of power in localising disease genes using the traditional lod-score method of genetic linkage analysis. The sib-pair method of genetic linkage analysis was used in this study (see below in 1.8.3) and does not rely on the specification of a genetic model, so the results are not reliant on this parameter. However, once a reasonable genetic model can be postulated for the transmission of the disease in the affected families, and there is enough family material available, genetic linkage analysis using the traditional lod-score method is known to work effectively to localise susceptibility loci.

Our group, assuming multilocus inheritance, completed computer simulations on the first 100 families collected, to determine the expected maximum lod score (EMLS) that could be obtained from this number of informative meioses. From these simulations it is plausible to expect that we can identify any loci ($EMLS > 3.0$) with $\lambda_S = 3$ or more with a 10cM marker map, if all markers have an average heterozygosity of 0.67 (Table 5). With the addition of 100 more families we can expect to detect any susceptibility loci with $\lambda_S \geq 2$ using a 10cM genetic map (het. = 0.67) (Table 5). As described in section 1.5.2.1, our estimate of λ_S for autism is

between 50-200, thus our sample size should give enough power to discover at least several gene effects.

Table 5: EMLS for Multipoint Analysis, 10cM map, Het. = 0.67.

	λ_s		
<u>N</u>	<u>2.0</u>	<u>3.0</u>	<u>4.0</u>
100	1.91	3.43	4.36
200	3.82	6.85	8.76

1.8.2 Complex Disorders

The use of linkage analysis in mapping complex traits is problematic since it is difficult to find a precise model that adequately explains the inheritance pattern. As previously defined, the term “complex trait” refers to any disorder that does not exhibit classic Mendelian recessive or dominant inheritance attributable to a single gene locus. Cystic fibrosis and Huntington’s disease are classic examples of Mendelian disorders. In both cases, genetic linkage analysis was used to localise the disease gene. Finding a genetic marker that co-segregates with a complex trait, however, may be complicated by many factors including reduced penetrance, presence of phenocopies (non-genetic cases), genetic heterogeneity, polygenic inheritance and high frequency of disease-causing alleles in the population (Lander and Schork, 1994). Reduced penetrance, phenocopies and heterogeneity can be problems relevant to some Mendelian disorders as well.

- (I) **Incomplete Penetrance and Phenocopy:** Incomplete penetrance can be seen when individuals carry the predisposing allele but do not show phenotypic signs of the disease. Phenocopy, on the other hand, is seen when affected individuals do not carry the predisposing allele but develop the disease due to other causes (eg. environmental agents). These factors may affect genetic mapping studies, since the predisposing allele may either be present in some unaffected individuals or absent in some affected individuals.
- (II) **Genetic Heterogeneity:** Genetic heterogeneity is present in those disorders that are caused by multiple mutations at several genes. This can result in identical phenotypes, especially when the products of the affected genes are part of a common biochemical pathway. An example is retinitis pigmentosa, which involves retinal degeneration, and can result from mutations in any of at least 14 different loci (Bleeker-Wagemakers *et al.*, 1992; Kumar-Singh *et al.*, 1993). Genetic heterogeneity affects genetic mapping because a certain chromosome region will co-segregate with the disease in some families but not others. Allelic heterogeneity (the presence of multiple disease-causing mutations in a single gene), however, does not affect genetic mapping since the disease locus co-segregates with the same chromosomal region in all affected families.

The effect of genetic heterogeneity and incomplete penetrance was examined in our sample of families using computer simulations. Assuming all markers had an average heterozygosity of 0.67, we concluded that 200 families would detect linkage under the lowest penetrance model at a recombination fraction of 0.05, if at least 50% of the families are linked to this gene. Approximately 33% of linked families would be sufficient in models

with higher penetrance or at lower recombination fractions. In the case that only 25% of the families are linked, a sample size of 290 would be needed.

- (III) Polygenic Inheritance: Some complex traits require the simultaneous presence of mutations in multiple genes and this is known as polygenic inheritance.

These multiple interacting genes may consist of either oligogenes

(incompletely penetrant genes with a relatively large and therefore potentially

recognisable effect on the trait) and/or polygenes (genes with a minor

individual effect on the trait that may make their isolation by linkage analysis

difficult) (Morton, 1998). Polygenic (complex) traits may be classified as

distinct traits, measured by a specific outcome, or quantitative traits, measured

by a continuous variable (eg. diastolic blood pressure) whose level may be

determined by the combined action of individual quantitative trait loci.

- (IV) High Frequency of Disease-Causing Alleles: If the disease-causing allele

occurs at a high frequency in the population, it could be difficult to localise.

An example of this is late-onset Alzheimer's disease. Linkage studies showed

weak evidence of linkage to chromosome 19q. However, due to the low lod

scores, these results were dismissed. It was not until the discovery that the

apolipoprotein E type 4 allele on chromosome 19 was a predisposing factor

that it was realised that these linkage results were correct. The reason behind

the calculation of a low lod score was found when the frequency of this allele

was shown to be ~16% in most populations, thus interfering with the

traditional linkage analysis (Corder *et al.*, 1993).

1.8.2.1 Defining the Disease Phenotype

Narrowing the definition of a disease, or restricting the patient population so that only individuals with the “core” form of the disease are diagnosed, can assist in overcoming some of the difficulties mentioned above in mapping complex diseases. The selection of families is based on the presence of one or more of the following characteristics:

- (I) Unusual phenotypic features: Identifying a subset of families with an unusual phenotypic feature may reduce the number of genes involved in the disorder. For example, it was found that colon cancer with extreme polyposis was caused by a single dominant gene, and that other forms of colon cancer can be distinguished by the phenotype of replication errors in tumours (Aaltonen *et al.*, 1993).
- (II) Age of Onset: Strictly genetic forms of disease often exhibit an earlier onset compared to complex or sporadic forms. For example, genetic homogeneity was found in breast cancer when only families with disease onset below 47 years were included in the analysis (Hall *et al.*, 1990).
- (III) Extreme Severity: Particularly in the case of quantitative traits, it is often shown that inclusion of individuals with phenotypic values in the tails of the distribution increases the power to detect linkage (Risch and Zhang, 1995).

In this study, only individuals diagnosed with the “core” syndrome of autism were included (ie. individuals diagnosed with Asperger syndrome and/or autistic spectrum disorders were excluded). All individuals with an IQ below 30 and/or associated with another disorder were also excluded from the analysis. By creating a homogeneous set of families with a diagnosis that is replicable between clinicians, our chance of detecting susceptibility loci should be increased.

Age of onset was also examined as a basis of discriminating distinct subgroups of families in our study. Results from the analysis of the first 37 families showed that there was great variability in the ages of onset, ranging from briefly after birth to 78 months and that the ages of onset between siblings were not significantly different. However, when both parents' reports of age of onset were examined separately, there was high variability between the ages specified. From these data, it was concluded that the use of retrospective parent reports of age of onset may be too unreliable and we could find no evidence to support the value of dividing autistic families on the basis of age of onset.

1.8.2.2 Linkage Analysis Methods Used for the Study of Complex Disorders

Genetic linkage analysis relies on the specification of an accurate genetic model. However, it is difficult to use classical segregation analysis to accurately model a complex trait since it requires a significant amount of pedigree data to estimate the large number of unknown parameters required. In light of this problem, many linkage analysis methods that differ from the traditional lod score method have been proposed for complex traits. One involves using parametric linkage analysis under several simple modes of inheritance (Risch *et al.*, 1989). This type of analysis is called parametric because parameters such as penetrances and allele frequencies must be specified for the disease gene. However, analysing data under a clearly "wrong" model is expected to result in a loss of power and in biased estimation (overestimation of the recombination fraction). For some two-locus models, the degree of information loss has been shown to vary depending on the mode of inheritance (Schork *et al.*, 1990). Other methods involve non-parametric linkage analysis. These methods do not assume that a single major disease locus (with a specific mode of

inheritance) accounts for the majority of the genetic variance, and are best suited for searching for susceptibility loci involved in genetically complex disease. One non-parametric method is the affected sib pair approach, and this is the method of choice in this study.

1.8.3 Sib-Pair Methodology

Affected sib-pair analysis examines siblings (both with the disease) for the sharing of marker alleles at multiple sites in the genome. The more often the affected siblings share the same allele at a particular site, the more likely it is that the site is close to the disease gene (Penrose, 1935). Although allele-sharing methods are often less powerful (in terms of the chance to detect linkage) than parametric analysis with a correctly specified model, they tend to be more robust. Affected siblings show an excess of allele sharing even in the presence of incomplete penetrance, phenocopy, genetic heterogeneity and high-frequency disease alleles (Lander and Schork, 1994). This method is very useful in studying autism, since the unknown parameters such as affection status of the parents and the mode of inheritance do not need to be specified.

Sib-pair analysis involves testing whether affected siblings inherit a region that is identical-by-descent (IBD) more often than expected under random Mendelian segregation. Under random segregation the probabilities of sharing zero, one or two alleles IBD are 25%, 50% and 25% respectively. If both parents are available, the data can be partitioned into separate IBD sharing for the maternal and paternal chromosomes. In this case, each affected sibling could share zero or one allele with the other sibling, with a 50%-50% distribution expected under random segregation.

A deviation from the expected 50% IBD allele sharing suggests the presence of a trait-causing gene (Lander and Schork, 1994).

1.8.3.1 Power to Detect Linkage

The power of sib-pair methods to detect linkage to a disease-susceptibility locus for a given sample size, depends only on the risk ratio λ_s (ie. increase in risk to siblings compared with the population prevalence) and on no other genetic model parameters (such as number of alleles at a given locus, gene frequencies and penetrance; Risch, 1990b). In autism, the increase in risk to siblings is equal to an overall lambda (λ_s) value of 50-200. This lambda value (λ_s) is an overall risk ratio that summarises the collective effect of all the disease loci. The higher the λ_s value, the stronger the genetic effect. A locus with a small genetic influence on the disease corresponds to a $\lambda_s < 2$ (ie. it determines <2-10% of the total genetic variance), whereas a moderate to large gene effect coincides with a $\lambda_s > 4$ (ie. it determines >4-20% of the total genetic variance). There is a high observed λ_s value for autism, suggesting that affected sib-pair analysis may be a powerful approach for mapping susceptibility genes in this disorder. However, this still depends on the number of genes involved and on their interaction with each other.

Genes for a specific disease can interact in either a multiplicative or additive manner depending on the penetrances of the individual loci (ie. the probability of being affected given a certain genotype). The genes responsible for autism are generally thought to interact in a multiplicative manner. This means that if there are N loci involved, the overall λ_s is the product of the locus-specific risk ratios, ie. $\lambda_s = \lambda_1\lambda_2\lambda_3\dots\lambda_{N-1}\lambda_N$ (Risch, 1990b). Since the locus-specific risk ratios are not known at

the beginning of a study, exclusion mapping is generally conducted over a range of lambda values. The range used in this study was between 1.5 and 10, representing weak to strong gene effects.

The maximum likelihood statistic (MLS) T , is used to detect evidence for linkage of a specific marker in a set of affected sib pairs. MLS is analogous to the lod score statistic (Z) in that its value is the \log_{10} of the odds in favour of linkage (Morton, 1955). A T value of 3.0 corresponds to odds of 1000:1 in favour of linkage and is considered to provide evidence of linkage. The MLS statistic T can be shown mathematically as follows:

$$T = n_o \log_{10} (n_o / N_{\alpha s_0}) + (N - n_o) \log_{10} (N - n_o / N - N_{\alpha s_0})$$

Where N is the number of affected sib-pairs, n_o is the number of sib-pairs that share 0 marker alleles IBD, $N - n_o$ is the number that share 1 or 2 marker alleles IBD, and $N_{\alpha s_0}$ denotes the probability that a sib-pair share 0 alleles at a random locus IBD.

As evident from the above formula, the power to detect linkage at a given value of λ_s is directly proportional to the number of sib-pairs studied. Figure 5 shows this relationship diagrammatically.

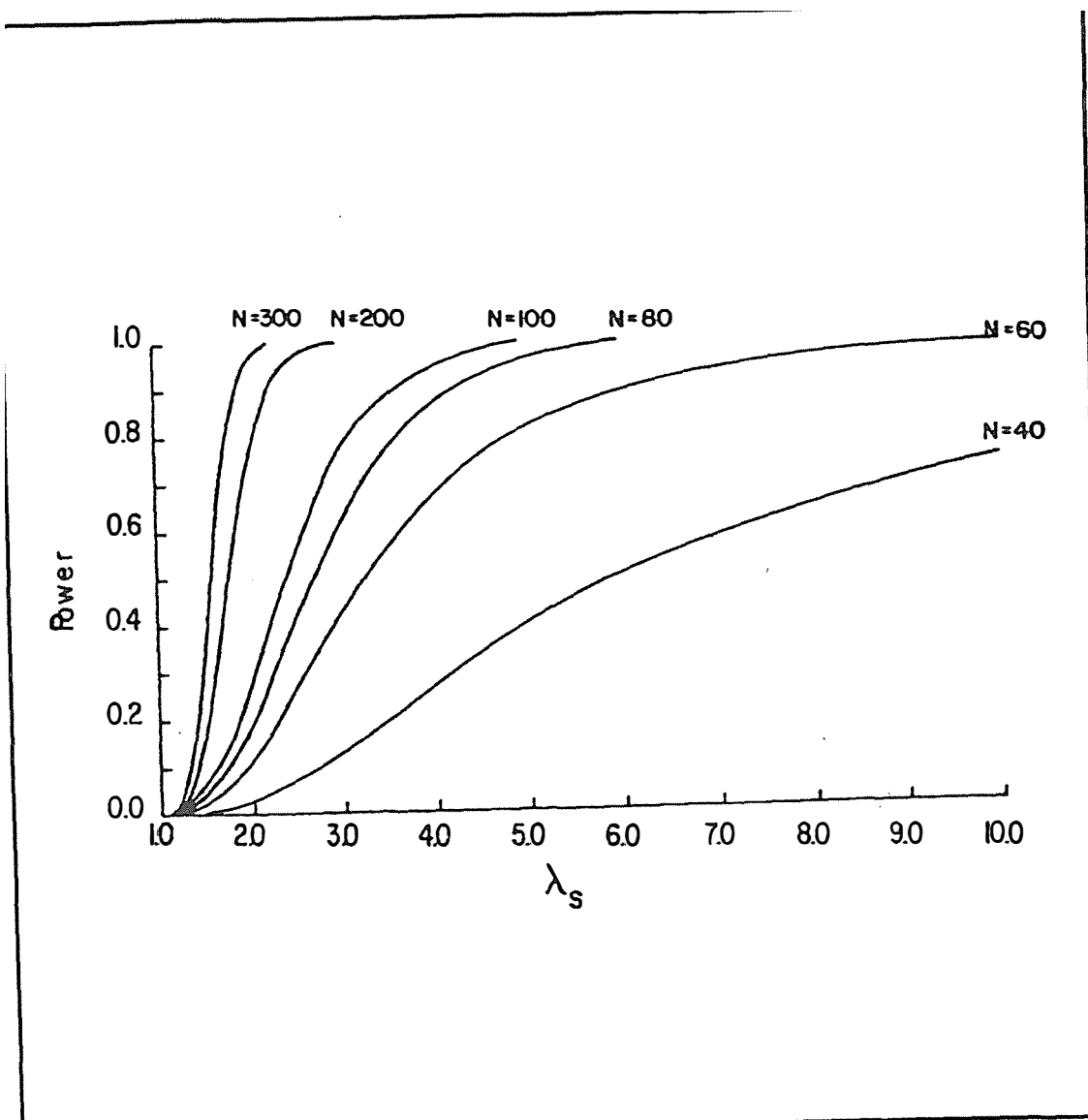


Figure 5: Power to detect linkage at differing lambda values (λ_s), using N affected sib pairs and assuming zero recombination and a completely informative marker (From Risch, 1990b)

Figure 5 shows that if the relative risk for a locus is 2, a sample size greater than 200 will be required to detect linkage with 80% power. However if $\lambda_s = 5$, a sample size of 60 will be sufficient to detect linkage with 80% power. This value represents the minimal sample size required, since recombination (θ) is assumed to be zero, and the marker is assumed to be completely informative (100% polymorphic). With increasing θ , the power to detect linkage decreases drastically. A study by Risch

(1990b), found that the power to detect linkage using the MLS approach on a sample of 300 sib-pairs dropped by approximately 50% at a recombination fraction of about 0.10. However, the effects of recombination can be reduced by using multi-marker analysis (Goldgar, 1990; Fulker *et al.*, 1995; Olson, 1995; Kruglyak and Lander, 1995).

1.8.3.2 Multipoint Sib-Pair Analysis

Multipoint linkage analysis examines several linked loci simultaneously to locate disease susceptibility loci. Not only is this method able to detect linkage when recombination between the marker and the disease locus is not equal to zero, but it can increase the power to detect linkage (Risch, 1990b) and also can be used to exclude regions of the genome from linkage to a disease gene (Risch, 1993). Multipoint sib-pair analysis also allows for maximal use of marker information when assessing linkage, and determines a confidence region likely to contain a susceptibility gene. In this method of analysis, a lod score is calculated for each sib-pair at each marker locus and at a number of intervals between these markers, under the assumption of a specific lambda value. The lod scores at each interval are then added up over all sib-pairs, which gives the total lod score. As in the lod score method, a lod score of < -2 suggests lack of linkage and a lod score > 3 is suggestive of linkage.

1.8.3.3 Genes Localised in Complex Disorders Using Multipoint Sib-Pair Analysis

Multipoint sib-pair analysis has been used to localise susceptibility loci for some complex disorders. Three examples are insulin-dependent diabetes mellitus (IDDM), multiple sclerosis (MS) and ankylosing spondylitis (AS).

Insulin-Dependent Diabetes

Insulin-dependent or type 1 diabetes mellitus (IDDM) is an autoimmune disease associated with loss of tolerance to several pancreatic islet cell molecules (Krause *et al.*, 1979). It is a complex disorder determined both genetically and environmentally, and frequently occurs in childhood. Mouse model studies have shown that this severe form of diabetes is caused by a combination of a major effect at the major histocompatibility complex (MHC) and at least ten other susceptibility loci elsewhere in the genome (Risch *et al.*, 1993; Podolin *et al.*, 1998). A genome-wide scan of 356 affected sib-pair families from the United Kingdom indicated a similar genetic basis for human IDDM. A major gene component was located at the MHC locus (IDDM1) and can explain 34% of the familial clustering of the disease (MLS = 34, $\lambda_s = 2.5$) (Davies *et al.*, 1994; Cucca *et al.*, 1998). Two other loci showed evidence of linkage including chromosome 10p13-p11 (MLS = 4.7, $\lambda_s = 1.56$) and chromosome 16q22-16q24 (MLS = 3.4, $\lambda_s = 1.6$) (Mein *et al.*, 1998). The programme used to calculate the multipoint MLS values and the information content and exclusion maps was MAPMAKER/SIBS.

Multiple Sclerosis

Multiple sclerosis (MS) is a chronic neurological disorder which presents with progressive sensory or motor dysfunction that can ultimately lead to disability and death (Cumings *et al.*, 1965). Strong evidence indicates that MS is a complex autoimmune disorder, with its definition based solely on clinical criteria and its mode of inheritance non-Mendelian. Genetic influence in the disorder has been shown through population and twin studies, with a 20-fold increase in risk for first degree relatives over the general population (Sadovnick *et al.*, 1988), and MZ concordance

rates of 25-30% compared to 4% for DZ twins and siblings (Ebers *et al.*, 1986; Mumford *et al.*, 1994). A genome-wide search using 257 microsatellite markers in 100 Canadian sib-pairs showed evidence of linkage at chromosome 5p (D5S406) with a maximum multipoint lod score of 3.1 (maximum pointwise lod score = 4.24). This value increased to 3.6 with the addition of 44 extra sib-pairs, but diminished to 1.6 when an extra 78 sib-pairs were analysed. It was suggested that there was significant heterogeneity in the third data set compared to the first two, and evidence for a susceptibility locus in that region of chromosome 5p could not be considered confirmed (Ebers *et al.*, 1996).

Ankylosing Spondylitis

Ankylosing spondylitis (AS) or spondylarthritis is a common inflammatory condition that is strongly associated with the HLA region (Brown *et al.*, 1998). The reduction in recurrence risk among increasingly distant relatives of AS patients suggested that the genetic component of AS involved multiple interacting genes (Lawrence, 1977). A genome-wide linkage screen was undertaken using 254 microsatellite markers in 121 affected sibling pairs with AS from Britain. Multipoint analysis using the programme GeneHunter showed significant linkage to the MHC region (MLS (D6S273) = 3.8; MLS (82-2) = 8.1), and suggested that 31% of the susceptibility to AS is coded by genes linked to this region. A MLS of 2.6 was also obtained on chromosome 16 at the locus D16S422, suggesting that non-MHC genes may also influence susceptibility to the disease (Brown *et al.*, 1998).

1.8.4 Association Analysis

Association studies can be used in conjunction with linkage analysis to assist in the detection of disease-susceptibility genes. This is particularly the case for genetic effects too weak to be identified by current methods of linkage analysis (Risch and Merikangas, 1996). Association studies examine the difference in frequencies of a marker allele between patients and controls. An association will be detected if the marker investigated is a susceptibility gene, or if it is in linkage disequilibrium (LD) with the susceptibility gene locus.

Linkage disequilibrium is a term used to describe a situation in which an arrangement of alleles at different loci on the same chromosome (the haplotype) occurs at a frequency greater than would be expected by chance. This can be caused by a variety of factors including: (1) recent admixture of populations with different allele frequencies, (2) selection in favour of a specific allele (which increases the frequency of haplotypes containing this allele), and (3) genetic drift or population bottlenecks and new mutation (creates disequilibrium with closely linked markers) (Weeks and Lathrop, 1995). Since LD decreases with time in proportion to the recombination fraction between the loci, disequilibrium between closely linked loci will decay less rapidly. If these linked loci in LD are in the vicinity of a gene, this may lead to the identification of a disease-susceptibility gene in the affected individuals.

Using LD to map disease loci is more efficient in genetically isolated populations. This is because there is a greater chance that there is only one ancestral haplotype on which the disease mutation arose and, if the population is relatively young, LD could extend over substantial distances increasing the probability of detection. An example of an isolated population is the Finnish population which is

believed to have settled in the region over 1,000 years ago, with little immigration until at least the middle of the 14th century (Virtaranta-Knowles *et al.*, 1991).

A significant problem with association studies, however, is that the results are highly sensitive to mis-specification of allele frequencies in the study population. This can be due to recent admixture of different populations (with each sub-population contributing different allele frequencies) (Knowler *et al.*, 1988), or to inappropriate matching of patients and controls. The haplotype-relative-risk (HRR) approach overcomes this problem by constructing an “artificial” control population from the set of alleles that were not transmitted to affected offspring from their parents (Falk and Rubinstein, 1987). The estimation of allele frequencies in this case is only conditional on the transmitted versus non-transmitted alleles. A similar analysis is termed the transmission-disequilibrium-test (TDT) (Spielman *et al.*, 1993). This method compares the observed transmission of an allele of interest from heterozygous parents with the expected rate of transmission (ie. 50%). The additional cost of typing more individuals in these methods compared to traditional association analysis is well justified, since population stratification problems leading to false positive results are overcome. These family-based association methods have been used to detect disease genes (Thomson, 1995), and since they do not require any additional genotyping to sib-pair linkage analysis, the two methods can be used in parallel to assist in the detection of disease susceptibility loci. To ensure, however, that LD is independent of any linkage present when multiple sibs from the same family are tested, a Monte Carlo simulation is used (see Risch *et al.*, 1999; Appendix 6.3).

1.8.5 Linkage and Association Studies in Autism

Due to the sufficiently large familial relative risks and the data mentioned above supporting autism as a genetic disorder, mapping studies can be undertaken to identify the genetic loci that contribute to the disorder (Bristol *et al.*, 1996). To date there have been only three systematic genome scans in autism research and the number of association studies in this disorder has been relatively limited. The results from these molecular genetic studies are outlined below, in order of the area of the genome they covered and the year in which they were completed.

1.8.5.1 HLA and 19 Autosomal Markers (Spence *et al.*, 1985)

The first linkage study in autism was accomplished by Spence *et al.* (1985). They studied 34 multiplex families (two or more autistic children) from the UCLA Registry for Genetic Studies of Autism, for linkage to the HLA region (chromosome 6) and 19 other autosomal markers. No evidence of linkage was obtained between the disease and the HLA region using the lod score method. When an association test was performed by comparing the number of shared HLA haplotypes in autistic siblings to what would be expected by chance, a significant result was still not found. The largest lod score calculated from these 34 families was 1.04 for haptoglobin, a marker whose gene lies on chromosome 16. When the families were divided into two groups, ie. families with only affected males and families with at least one affected female, a marker suggestive of linkage was identified in the second group of families. This marker was linked to the ABO blood group locus on chromosome 9 and the calculated lod score was 1.54.

1.8.5.2 HLA (Warren *et al.*, 1991/1992/1996a; Daniels *et al.*, 1995)

Investigators have reported associations between alleles of the HLA complex and autism when compared to normal controls. These autism-associated alleles include the null allele of the C4B gene (Warren *et al.*, 1991), the extended haplotype B44-S30-DR4 (Warren *et al.*, 1992; Daniels *et al.*, 1995), the third hypervariable region of HLA-DRB1 (Warren *et al.*, 1996a) and two alleles of the DR beta 1 gene (Warren *et al.*, 1996b). The third hypervariable region of HLA-DRB1 has shown the strongest association with autism. Warren *et al.* (1996a) calculated that the relative risk of developing the disorder if you have this marker was 19.8 for allele 1 of this marker and 4.2 for allele 3. This region is essential for the presentation of peptides to T lymphocytes by antigen-presenting cells (APC), and is therefore necessary for the development of cell-mediated and antibody immune responses to pathogens and other foreign antigens (Warren *et al.*, 1996b).

1.8.5.3 Chromosome 15 (Gillberg *et al.*, 1991b; etc.)

A number of case studies have reported the association of autism with partial trisomy / tetrasomy and duplication of chromosome 15 (see section 1.5.1.2). Hotopf and Bolton (1995) offered a number of explanations for these findings, hypothesising that autism has a heterogeneous genetic basis or, alternatively, that autism may have a consistent genetic basis in an aberration of chromosome 15.

1.8.5.4 Fragile X (Hallmayer *et al.*, 1994)

The association of the Fragile X syndrome (FraX) with autism is well known (see section 1.5.1.2). The gene responsible for FraX is FMR-1 (Oberle *et al.*, 1991; Verkerk *et al.*, 1991) and the syndrome is usually caused by an amplification of the

CGG trinucleotide repeat region of this gene (Rousseau *et al.*, 1991). Hallmayer *et al.* (1994) examined the possibility that a mutation elsewhere in the FMR-1 gene was involved in the aetiology of autism. This was achieved by carrying out linkage analysis on 35 multiplex families using microsatellite markers that were tightly linked to the FMR-1 gene (DXS548, FRAXA1, FRAXA2, CGG repeat). Using four different dominant models of varying penetrance values and a recessive model, linkage to the FMR-1 gene was excluded (lod score between -24 and -62) in all models.

1.8.5.5 X-Chromosome (Hallmayer *et al.*, 1996)

Due to the increased ratio of autistic males to females, even in multiplex families (see section 1.4.3), a search for an autism susceptibility gene on the X-chromosome has been conducted. This was the first reported application of the sib-pair approach to the search for genetic influences in this disorder. Thirty-eight multiplex families were tested for linkage to 35 microsatellite markers using multipoint sib-pair analysis, and linkage to the entire X-chromosome could be excluded for $\lambda_s = 4$. A maximum lod score of 1.24 was obtained between DXS424 and DXS1001 for a $\lambda_s = 1.5$. Another lower peak (0.93) was observed at DXSYX1. Overall, a moderate to major gene effect for autism was excluded on the X-chromosome. A susceptibility gene increasing the sibling risk between 1.5 to 3 fold could, however, not be excluded, in particular a gene of small effect located between DXS424 and DXS1001 (Hallmayer *et al.*, 1996).

1.8.5.6 X-Chromosome (Petit *et al.*, 1996)

An association study was completed on 90 autistic children from a Child Psychiatry Day Care Unit in France and unrelated controls. Eight markers on the X-

chromosome were tested using χ^2 analysis. DXS287 was the only marker that showed a different allele distribution between control and patient groups. This difference was enhanced in the more severely autistic and least retarded children. Investigators noted that if this locus on the X-chromosome is specific for autism, the result would have to be replicated on another group of autistic children, as well as a non-retarded autistic population. Also, the accuracy of the result could be increased if the parents were typed and the haplotype relative risk method (HRR) was employed.

1.8.5.7 Harvey-Ras Factors (Herault *et al.*, 1995)

Herault *et al.* (1995) compared five polymorphic markers in a sample of 50 autistic children and 50 controls, and found that the HRAS marker B3/B3 homozygous genotype was significantly less frequent in autistic children than in controls. The HRAS gene (chromosome 11p15.5) codes for the *ras* protein, which is active in neuron development, and the researchers hypothesise that it may be the basis of autism. However, this has been only a “preliminary study and a further study with a larger population and a wider IQ distribution must be performed”.

1.8.5.8 Serotonin (Cook *et al.*, 1997b)

Since serotonin transporter inhibitors have been shown to be partially efficient in the treatment of autistic symptoms such as restricted and repetitive behaviours, language use and aggression (Gordon *et al.*, 1993; McDougle *et al.*, 1996), Cook *et al.* (1997b) considered the serotonin transporter (HTT) as a candidate gene in autism. They examined possible linkage and linkage disequilibrium between two polymorphisms reported for the serotonin transporter gene (a VNTR in the second intron of HTT and a short variant of the HTT promoter) with autism, using the TDT

test. TDT analysis of 86 US families showed no evidence for linkage or linkage disequilibrium between autism and the VNTR, however, preferential transmission of the short variant of the HTT promoter was found (TDT $\chi^2 = 4.69$, 1 d.f., $P = 0.030$) (Cook *et al.*, 1998). It should also be noted that another haplotype at HTT was found to be preferentially transmitted in a sample of German subjects with autistic disorder (Klauck *et al.*, 1997). These results need clarification by further studies with different populations and larger groups of autistic individuals.

1.8.5.9 Whole Genome Scan (International Molecular Genetic Study of Autism Consortium, 1998)

A two-stage genome search for susceptibility loci in autism was performed by the International Molecular Genetic Study of Autism Consortium (IMGSAC; 1998). They used 99 affected relative pairs, which included 87 affected sib-pairs and 12 affected non-sib relative pairs. Fifty-six of the sib-pair families were from the United Kingdom (UK). A least one individual in each pair had a clinical diagnosis of autism, satisfying the ADI criteria (see section 2.1.2) for autism and showing abnormalities before the age of 3 years. The other individual had to have either a clinical diagnosis of Asperger syndrome, Pervasive Developmental Disorder (PDD), autism unaccompanied by language delay, or had one of these clinical diagnoses but fell 1 point below the threshold on one of the behavioural domains of the ADI algorithm (IMGSAC, 1998).

In stage one of the genome screen, 354 microsatellite markers were typed in 39 families. Using ASPEX (for the sib-pair families; Hinds and Risch, 1996) and GeneHunter (for the relative-pair families; Kruglyak *et al.*, 1996), pairwise and multipoint MLS were calculated. Sixty-two markers had an MLS > 0.5. In stage

two, 60 additional families were genotyped using a subset of 175 markers that focused on the regions of interest identified in stage 1. Based on all 87 sib-pair families, approximately 32% of the genome was excluded for $\lambda_s = 2.5$, which included the entire X-chromosome. In these sib-pair families, regions on six chromosomes (4, 7, 10, 16, 19 and 22) were identified that had lod scores (MLS) > 1 . The most significant result was found on chromosome 7q between the markers D7S530 and D7S684, with a MLS of 3.55 in the subset of UK families and an MLS of 2.53 in all families. The next most significant result was near the telomere on chromosome 16p, with an MLS of 1.97 in the UK families and 1.51 in all families. Linkage disequilibrium was tested for 13 markers in the regions of interest on chromosome 7 and 16. This was achieved using the transmission disequilibrium test (TDT), as implemented in the ASPEX programme. There was no evidence of linkage disequilibrium in either of these regions. The investigators explained that the large distance between these markers (5-10 cM) could have affected these results.

1.8.5.10 Whole Genome Scan (Bass *et al.*, 1998)

A genome screen has also been completed by Bass *et al.* (1998) and reported at The American Society of Human Genetics 48th Annual Meeting in Denver, Colorado in 1998. They reported analysing 49 affected sib-pairs and 3 affected cousin pairs using 430 markers that were around 7cM apart. Results from this analysis identified a number of regions with a p -value < 0.05 or a MLS ≥ 1 , including sections on chromosomes 1, 2, 3, 7, 18 and 20. The exact chromosomal location of these findings was not explained in this report.

1.8.5.11 Whole Genome Scan (Philippe *et al.*, 1999)

The third published genome-wide screen for autism susceptibility loci was performed with 264 microsatellite markers in 51 multiplex families, using non-parametric linkage analysis. Families with at least two siblings fulfilling the DSM-IV criteria for autistic disorder and the ADI algorithm for ICD-10 childhood autism were recruited from Sweden, France, Norway, USA, Italy, Austria and Belgium. Using two-point and multipoint affected sib-pair analysis they revealed 11 chromosomal regions with a maximum lod score (MLS) greater than 0.6 ($P < 0.05$). Four of these regions overlapped with regions on chromosomes 2q, 7q, 16p and 19p identified in the IMGSAC genome scan. Another overlapped with the 15q11-q13 region identified in previous candidate gene studies. The remaining six regions were on chromosomes 4q, 5p, 6q, 10q, 18q and Xp. The most significant result overall was obtained for a region close to marker D6S283 (MLS=2.23), however this result was not statistically significant for linkage (MLS>3).

1.9 Conclusion

In summary, autism is described as a chronic developmental disorder characterised by behavioral, communication and social abnormalities that manifest in the first years of life. While the aetiology of autism is unknown, there is general agreement that it results from early neurodevelopmental dysfunction.

Family studies demonstrate that the prevalence for autism among siblings of autistic children vary between 2% and 9% (Jorde *et al.*, 1991). Since the prevalence of autism in the general population is about 4-5 in 10,000 births, the recurrence risk

for autism among siblings is 50-200 times greater than in the general population (Bolton *et al.*, 1994).

Studies have also reported an association with autism of a number of other genetic conditions. These include: fragile X syndrome; chromosome 15 anomalies; tuberose sclerosis; and phenylketonuria, just to name a few. These associations though, are not sufficient to explain the strong familial clustering of autism.

It has been suggested that the increase in prevalence of autism in families is due to genetic factors and a number of twin studies have supported this theory. These studies estimate that the concordance rate in monozygotic twins is approximately 77%, while the concordance rate reported for dizygotic twins is about 0% (Folstein and Rutter, 1977; Ritvo *et al.*, 1985; Steffenburg *et al.*, 1989; Bailey *et al.*, 1995). The dizygotic concordance rate is not significantly different from the approximate sibling recurrence rate of 3%.

This evidence of a genetic component in the aetiology of autism allows molecular genetic strategies to be used to determine the biological basis of autism. Since autism is most likely a complex disorder involving several interacting genes of unknown mode of inheritance, non-parametric linkage analysis utilising the multipoint sib-pair approach is, at present, one of the best methods for identifying susceptibility loci in this disorder. This method detects how frequently two affected sibs share the same chromosomal region identical-by-descent (IBD). Sib-pair analysis is very useful in studying autism because it is independent of both the affection status of parents and the mode of inheritance. The power of sib-pair method to detect linkage depends on the risk ratio. The risk ratio λ_s is the increase in the risk of the disorder in sibs of an affected individual, compared with the general population prevalence.

Assumptions about genetic model parameters, such as number of alleles at a given locus, gene frequencies, and penetrances are not required in sib-pair analysis (Risch, 1990b). In the case of autism, the overall λ_s (increase in risk to siblings) is approximately 50-200. Due to the high λ_s value in autism, affected sib-pair analysis may be a powerful tool for mapping susceptibility genes in this disorder, depending on how many individual genes determine the total value of λ_s .

1.10 Rationale of the Study

In this study, multipoint sib-pair analysis was employed to identify susceptibility genes involved in the aetiology of autism. From our simulations we concluded that at least 3 loci are involved in the disorder and, assuming $\lambda_s = 75$, the study has the power to detect up to 5 loci of equal effect on autism using this method of analysis (section 1.7).

From additional computer simulations (see Table 5), we showed that it is plausible to expect that we can identify any loci ($MLS > 3$) with $\lambda_s \geq 3$, using a microsatellite map consisting of markers spaced approximately 10cM apart and at least 100 multiplex families. If 200 families are used in the genome scan, susceptibility loci with $\lambda_s \geq 2$ should be detectable using the sib-pair method.

Also, to ensure a more reliable diagnosis of autism and a smaller chance of obtaining false positive and negative results, we only used individuals diagnosed with the “core” syndrome of autism (ie. no individuals with Asperger syndrome or

pervasive developmental disorder-not otherwise specified, were included). This homogeneous set of families should enhance our chances of detecting linkage.

1.11 Research Aims and Significance

The objective of the study is to search for DNA markers linked to autism in a large sample of clinically well-characterised families. In the longer term, such markers will allow the identification of a gene or genes that are responsible for familial forms of autism.

1.11.1 Specific Aims

This linkage study on autism is a large-scale project that began in 1992 and is run in collaboration with universities in Australia and the United States of America. This PhD project addresses certain goals set out in the overall study design. The goal of the large-scale project is to complete the ascertainment and collection of samples in 180 families, create an exclusion map of the entire genome based on these families, identify regions not excluded and investigate them in greater detail by saturating them with markers. More specifically the aims of the large-scale project are:

- I. Recruit and evaluate approximately 140 U.S.A. and 40 Australian families with at least two autistic sibs, using strict diagnostic criteria.
A research diagnosis is made using the Autism Diagnostic Interview (ADI) and with the Autism Diagnosis Observation Schedule (ADOS).

Only those children satisfying the pre-specified cut-off scores in all three-symptom areas of the ADI and with an age-of-onset prior to three years are considered affected for the purposes of the project.

- II. Immortalise cell lines from the affected sibs and their parents in these 180 families. For families with one parent missing, unaffected sibs are included, as available, to increase genetic information.
- III. Genotype at least 350 highly polymorphic ($PIC > 0.70$) microsatellite loci at approximately 10 cM intervals in 90 sib-pair families, including affected sibs and parents, and unaffected sibs when a parent is missing.
- IV. Construct an exclusion map using multipoint sib-pair analysis assuming λ_S values between 1.5 and 10 for the 90 families of the initial screen. Through this, regions that cannot be excluded will be identified.
- V. For those regions not excluded in (IV), the remaining 50 U.S.A. and 40 Australian families will be genotyped. Additional microsatellite markers in these regions will be identified and all previously recruited families will be genotyped with these markers. Thus, the regions not initially excluded will be either excluded or confirmed as candidate regions containing autism susceptibility gene(s). Confidence intervals for any susceptibility loci identified will be constructed.
- VI. Candidate regions for susceptibility genes will be saturated with microsatellites. The confidence regions will be further narrowed using multipoint linkage analysis. Final candidate regions will be examined for linkage disequilibrium with these markers using the transmission disequilibrium test with the 180 sib-pairs and their parents.

The aims specific to the present PhD project are divided into two parts depending on the families analysed:

A. U.S.A. Families

- I. Isolation of DNA from cell-lines of some of the 90 U.S.A. multiplex families.
- II. A systematic scan of chromosomes 19-22, as part of the first stage screening on 90 multiplex families from the U.S.A.
- III. Refined study of candidate regions identified previously, such as the HLA region on chromosome 6 and the long arm of chromosome 15.
- IV. Construction of exclusion maps.
- V. Replication of the data obtained by us and the Stanford group, of regions showing positive lod scores (chromosome 1, 10, 15, 17 and 18).

B. Australian Families

- I. Establishment of immortalised cell lines for some of the 40 Australian families with 2 or more children diagnosed with autism and extraction of DNA from whole blood and white blood cell (WBC).
- II. Duplication of positive findings (lod score > 1) that were obtained in the initial genome scan on the U.S.A. families and replicated by the Perth group (see part V. above).
- III. Duplication of the analysis of chromosomal regions showing positive lod scores in other studies, eg. chromosome 7q (IMGSAC, 1998), using the Australian group of families.

1.11.2 Significance

The identification of genes responsible for autism may lead to an accurate and reliable biological marker being discovered. Since autism cannot be diagnosed confidently before the age of 4 – 5 years with current techniques, a biological marker could help identify children at risk for autism at a younger age if it was used in high-risk families. This would help in treatment and/or intervention strategies, which are most effective with early diagnosis and intervention.

A marker would also help in the determination of the boundaries of the autistic phenotype. At present there is much debate as to whether siblings of autistic children who have communication or cognitive disabilities suffer from a mild form of autism, and if these are part of an autistic spectrum.

Thirdly, genetic research will provide essential information about the regulation of brain development. This will allow the development of a more effective treatment for autism and its spectrum disorders.

CHAPTER 2

SUBJECTS AND METHODS

2.1 Clinical Methods

2.1.1 Ascertainment of Families

Families from the U.S.A. were recruited nationwide by advertisement targeted at local and national parent support groups and referrals from clinicians. The Australian families were recruited through peak agencies, clinic/assessment centres/hospitals, private agencies/self help or parent groups, special needs schools and prior autism research teams. Once families were identified, they were contacted via “mail-outs” by the above mentioned organisations. These "mail-outs" included information on the research project and asked whether or not they would like to participate in the study.

Families willing to partake were asked to contact the research group and a preliminary screen was performed over the telephone to indicate whether or not the specific family was suitable for the study. If this telephone-screen indicated that a family may be suitable, the aims of the project were explained and they were asked to sign a consent form allowing the clinicians on the project to view all participating family members clinical records. If these records verified that a family had a minimum of two children with a preliminary diagnosis of autism, they were

investigated further using standardised diagnostic procedures (see below; Le Couteur *et al.*, 1989b; Lord *et al.*, 1994; Lord *et al.*, 1989).

2.1.2 Diagnosis

Uniform, reliable and well-documented diagnostic tools are important to allow legitimate comparisons between studies. For this study, two standardised instruments were used for gathering the diagnostic information: the Autism Diagnostic Interview (ADI; Le Couteur *et al.*, 1989b and Lord *et al.*, 1994) and the Autism Diagnostic Observation Schedule (ADOS; Lord *et al.*, 1989). The ADI is an investigator-based interview, which reviews four specific areas relevant to autism: social interaction and behaviour, language and communication, interests and daily routines and age of onset. It is designed to provide a definition of autism for research and can not be used to disprove a clinical diagnosis. The ADOS is a behaviour observation instrument, which involves particular tasks that elicit specific behaviours in the child over a 20-45 minute time frame. ADI and ADOS were obtained for all children within a family regardless of the clinical history.

The ADI was used in this research project for the following reasons:

- (I) The interviewer asks the parent or caregiver to elicit certain examples of the child's behaviour, social activity etc. These can be used to evaluate particular signs of autism and allowed the interviewer to interpret the quality of the child's behaviour, rather than relying on the parent's description.
- (II) The interview has been specifically structured to include questions about all the behaviours used to diagnose autism.

- (III) Information on past and present signs and symptoms is collected in a systematic way. This is important, since the application of the diagnostic criteria requires data only from the year period between the ages of 4 and 5. This in turn allowed children who have been evaluated at different ages to be compared on a common scale.
- (IV) The ADI includes all the information on the individual needed to apply the ICD-10 criteria.
- (V) Both the ADI and the ADOS were videotaped, allowing other diagnosticians to independently verify each assessment. This permitted ongoing quality control of diagnosis.

A recent study on the reliability of administering and scoring ADIs between two diagnostic groups has shown a good interrater reliability, with a kappa coefficient for autistic verses non-autistic groups of 0.90 (Spiker *et al.*, 1994). Also, research by Lord *et al.* (1994) and Fombonne (1992) has indicated that the ADI has good reliability and validity for diagnosing autism and distinguishing this diagnosis from mental retardation.

Certified interviewers, who had completed the ADI training course given by Catherine Lord and Michael Rutter, conducted the ADI and ADOS evaluations. Donna Spiker, PhD, Linda Lotspeich, MD, and Patty McCague evaluated the majority of the U.S.A. families, whereas Amelia Nikolova, Wayne Hill, Robyn Young and Patty McCague interviewed the Australian families. To achieve the greatest possible blindness of evaluators, different interviewers were used for each child in a family whenever possible. Also, to check the reliability of the ADI administration and scoring, all interviews were videotaped and some randomly selected and independently scored by a second blinded, trained diagnostician.

An autistic individual was defined as one that had scores above the pre-specified cut-off points in the four areas in the ICD-10 system (ie. social impairment; language and communication impairment; unusual and restricted interests and routines; age of onset prior to three years). Multiplex families with at least one parent and two or more siblings who met full ADI definitions of autism were required in this study. Children were classified as unaffected if their scores were below the cut-off points in all four areas. Children falling into the ambiguous or uncertain diagnostic category (ie. with scores above the cut-off point in one, two, or three areas) were excluded from the linkage analysis, even though they met clinical definitions of syndromes related to autism (eg. pervasive developmental disorder). The unaffected and uncertain children were only used to reconstruct parental genotypes if parents were missing from the pedigree, and to check the accuracy of the data. Since there were no reliable diagnostic criteria for diagnosing autism in adults, the affection status of all parents was considered unknown. This conservative strategy tried to minimise the misclassification errors that have proved to be a major obstacle in psychiatric genetics. The ADI area scores for children in the first 37 multiplex families collected are shown in Table 6.

2.1.3 Intelligence Test Scores

Intelligence test scores (IQ) and mental age data were obtained from existing school or clinical records for the U.S.A. affected children. When available, nonverbal IQ scores were obtained from the performance subtests of the Wechsler Scales or Stanford-Binet Intelligence - 4th Edition, or the full Leiter International

Table 6: ADI Area Scores for Children in Autism Multiplex Sibling Families

(N = 37 Families, 117 Children)*

	Autistic (N = 83) (In 37 families)	Not Autistic (N = 26) (In 22 families)	Uncertain (N = 8) (In 6 families)
Social total			
Range	17-44	0-9	0-14
Median	36	1	1.5
Language total (V)^a			
No. of children	58	26	8
Range	11-29	0-8	0-7
Median	21	0.5	0
Language total (NV)^a			
No. of children	25	0	0
Range	9-16	-	-
Median	15	-	-
Ritual total			
Range	3-12	0-2	0-4
Median	8	0	1.5
Onset present?^c			
N (% Yes)	83 (100)	0 (0)	6 (75)

*** Social cut-off point** = 12, range = 0-42; language cutpoint = 10 for verbal subjects, range = 0-30, and cut-off point for nonverbal subjects is 8, range = 0-16; rituals cutpoint = 3, range = 0-12.

^aV = verbal: having phrase speech by age 5 years.

^bNV = non-verbal: not having phrase speech by age 5 years.

^cMeets onset criteria; developmental delays or deviance < 3 years

Performance Scale or Merrill-Palmer Scale. When these scores were unavailable, scores from such tests as the Stanford-Binet Intelligence Scale – 3rd Edition, Slossen Intelligence Scale, or McCarthy Scales were used. For a small number of children, mental age estimates and ratio IQ scores were based on nonverbal scores from

various developmental inventories (eg. Daily Living Domain of the Vineland Adaptive Behavior Scale, the full Bayley Scales, or Developmental Profile II).

If both individuals in the affected sib-pair had a mental age (MA) less than 18 months and/or a nonverbal IQ below 30, families were removed from the study because of the difficulties in diagnosing individuals in this group.

In the Australian families, affected individuals were classified under different categories of mental retardation as outlined in ICD-10 (World Health Organisation, 1993). Those individuals classified as mildly mentally retarded generally displayed delay in motor skills and had reasonable school achievement but still required support. Moderately mentally retarded children had difficulties in understanding instructions and displayed academic problems. Severely mentally retarded autistic children generally exhibited a lack of speech with no participation in the ADOS, whereas profoundly mentally retarded individuals displayed the same characteristics as those with severe mental retardation but with passiveness.

2.1.4 Ethical Considerations

This study has met the requirements of the Ethics Committees at Edith Cowan University, WA; Graylands Hospital, WA; The University of Western Australia; and Stanford University, U.S.A. Protocols and justification for the use of human subjects were renewed annually through the Human Subjects Committee at Stanford University. The subjects used in this project were from families in which there were two or more siblings with autism. Families were identified as potentially eligible based on diagnostic information provided by the parents. There were no

gender or ethnic restrictions on recruitment. The research staff who interviewed the parents about their children and videotaped a structured behaviour observation of each child collected the relevant information required for diagnosis. Additional records (medical and diagnostic evaluations) were collected from the parents or from appropriate medical or educational personnel, after the parent(s) signed a “release of information form”. To ensure confidentiality, all data were coded and stored in the computer by subject number; records containing names were stored in securely locked cabinets, available only to trained project staff. Project personnel could access computerised data with a password only. All staff on the project had been given specific training about issues of confidentiality in clinical research.

2.2 Clinical Characteristics of Autism Multiplex Families

2.2.1 U.S.A. Families

A total of 98 affected sib-pairs from the sub-group of 90 U.S.A. families were analysed in the first part of the study. These families encompassed a total of 187 affected children, 30 unaffected sibs, and 157 parents. The affected siblings in these families ranged in age at the time of the ADI administration from 2.9 to 40.9 years (mean = 9.8 yrs, S.D. = 7.3, median = 7.3, 25th - 75th percentiles = 4.9 - 11.3 years). The male to female ratio for this sub-group was 3.6:1 for the affected individuals and 0.8:1 for the unaffected individuals. Of these 90 families; two have four affected children, three families have three and 85 families have two affected children. For

families with n affected children, where n was greater than 2, $n-1$ sib-pairs were formed by matching the first child in the pedigree with each of the subsequent children; these pairs were fully independent. Both parents were available for analysis in 67 families. In those families where one parent was absent, unaffected siblings were included when possible (8 families) to reconstruct missing parental genotypes for identity-by-descent (IBD) tallies.

The mean nonverbal IQ was 66 (S.D. = 28, range = 16-160) for the affected children in this group of 90 U.S.A. autistic families. One hundred and eight children (56%) had scores below 70 in this same group of families, and were classified as mentally retarded. The mean mental age was 68 months (sd = 55 months, range = 13-373 months). All families had at least one affected sibling with an IQ above 30 and a mental age greater than 18 months.

The second set of U.S.A. families used to conduct genotyping studies of regions not excluded by the first 90 families, consisted of 49 multiplex families that were ascertained in a manner similar to the first set. These families were not used as part of this PhD project. For more information on these families refer to Risch *et al* (1999).

2.2.2 Australian Families

Forty-one Australian families were recruited for analysis. These families consisted of a total of 86 affected children, and 73 parents. Thirty-seven of these families had two affected children and four families had three affected siblings. Both parents were available for analysis in 32 families. The ratio of males to females in

the affected individuals was 3.0:1. When the ADI was administered to the affected siblings in these families they ranged in age from 3 to 46 years (mean = 12 yrs, S.D. = 8, median = 10, 25th - 75th percentiles = 7 - 13 years).

In this group of 45 ASPs from Australia, the majority of children were classified as being either mildly (35%) or moderately (32%) mentally retarded (see section 2.1.3). Nine percent of these children with autism were severely mentally retarded and 24% were classified as normal. No individuals were classified as profoundly mentally retarded and therefore all families were kept in the study.

2.3 Laboratory Methods

The laboratory methods used in this PhD project are stated below. Whilst some cell lines were established and DNA extracted by myself, the majority were prepared by collaborators in Stanford University and by co-worker Boyd Salmon at the Medical Research Institute in Royal Perth Hospital. I did the majority of genotypes, stated as part of this project.

2.3.1 DNA Isolation

Blood samples were obtained from all available family members at the time of evaluation. Blood was collected by venipuncture, with 20ml obtained from most patients. These samples were immediately transported to the laboratory. Cell cultures were started upon arrival and transformed with Epstein-Barr virus (EBV)

into lymphoblastoid cell lines, providing a renewable source of DNA. Genomic DNA was extracted either from whole blood (if an adequate supply had been drawn from the individual) and/or from immortalised cell-lines, by using standard proteinase K digestion and salting out procedures (Miller *et al.*, 1988).

2.3.1.1 Cell Lines

Epstein-Barr virus (EBV) transformed lymphoblast cell lines were established using standard procedures (Anderson and Gusella, 1984). B-lymphocytes were isolated from peripheral blood in cell culture by using EBV to immortalise cells. The peripheral blood cells may then be cryogenically stored for subsequent DNA extraction. The procedure of Anderson and Gusella (1984) differs slightly from the original procedures. The method was greatly improved by the use of cyclosporine A, a drug that proved very efficient in preventing the regression of newly established lymphoblastoid cell lines (Ventura *et al.*, 1988).

Lymphoblastoid cell lines were established from 10ml of whole blood that had been collected in a 10ml venipuncture vial containing heparin. The first step was to isolate the peripheral blood lymphocytes (a class of white blood cells) from the whole blood and place these in liquid nitrogen (lqN) for future use. Blood was centrifuged at 2000 rpm for 7min, collecting the buffy coat in a Nunc tube under sterile conditions using a transfer pipette, cooling on ice, and then adding dimethylsulphoxide (DMSO; ACS grade, ICN Biomedicals) at a concentration of 100µl/ml. The cell lines were established by defrosting the aliquoted buffy coat in warm water and adding 7ml of 12% FCS/RPMI (Fetal Calf Serum / 16.36g/L RPMI 1640; 2g/L sodium bicarbonate; 20ml/L penicillin / streptomycin; pH 7.2; stored at -20°C; Trace Biosciences [see Appendix 6.2 for solution protocol]) in a small 25cm²

tissue culture flask. Infection was achieved by mixing cells with 0.1-1ml of active EBV supernatant. The virus activates the enzyme telomerase, preventing the telomeres from shortening during each cell replication. This stabilises the telomeres and allows the cells to continue to proliferate, making them immortal. Cell lines were incubated at 37°C in a 5% CO₂ atmosphere and fed every third day with 12% FCS/RPM1. When a sufficient number of transformed cells (~50-100 million cells) were obtained, they were either transferred to a 50ml Falcon tube containing fresh culture medium (5ml of 12% FCS/RPM1), aliquoted out into 5 Nunc tubes and frozen in lqN with 10% DMSO, or placed in a 50ml Falcon tube containing no medium for DNA extraction (see below).

2.3.1.2 Salting-Out Method of DNA Extraction

DNA from lymphoblastoid cell lines was isolated and purified using standard procedures (Miller *et al.*, 1988). DNA is purified from whole blood or cell suspensions following cell membrane lysis and precipitation in the presence of high salt concentrations. Subsequent alcohol washes further purify the DNA, making it suitable for use in PCR. The salting-out method is simple, rapid and inexpensive and provides high, reproducible yields of pure genomic DNA.

If sufficient blood was drawn from the individual, DNA could also be extracted from whole blood. The procedure involved centrifuging approximately 10ml of whole blood that had been collected in a venipuncture vial with the anticoagulant ethylenediaminetetraacetic acid (EDTA), in a 50ml Falcon tube containing ~40ml of red cell lysis buffer (RCLB: 10mM Tris, pH 7.6; 10mM NaCl; 5mM of MgCl₂; stored at 25°C) at 2000 rpm for 7min. The supernatant was removed and the above step repeated 2 to 3 times to remove all red blood cells. The pellet was

then incubated overnight at 37°C with 30µl of proteinase K (20mg/ml of dH₂O, stored at -20°C; ProkayTM: Biotech Laboratories), 2ml of red cell lysis buffer (RCLB: 10mM Tris, pH 7.6; 10mM NaCl; 5mM of MgCl₂; stored at 25°C) and 13ml of white cell lysis buffer (WCLB: 10mM Tris-HCl, pH 7.6; 10mM EDTA, di-Sodium salt; 50mM NaCl; 0.2% SDS; stored at 25°C). If cell lines were used the DNA isolation procedure began at this incubation step, without the addition of RCLB. On the next day, 7ml of 6M NaCl was added to the proteinase K mixture with vigorous mixing and centrifuged at 5000 rpm for 45min. The DNA was then precipitated out of the resulting supernatant with an equal volume of cold isopropanol (AR grade, stored at -20°C) and placed in a 1.5ml microcentrifuge tube using a transfer pipette. The DNA was washed once with 70% ethanol (AR grade, stored at -20°C) and then again with 100% ethanol (AR grade, stored at -20°C). The DNA pellet was dried, then resuspended in 1ml of sterile water and incubated overnight at 37°C to dissolve the DNA.

2.3.1.3 Measurement of DNA Concentration

The optical density (OD) of the DNA solution was measured at 260nm and 280nm using a DU^R 640 Spectrophotometer (Beckman), which was blanked at both wavelengths using distilled water. A 260/280 ratio greater than 1.5 indicated the presence of relatively pure DNA. Readings at 260nm allowed the calculation of the concentration of the nucleic acid in the sample, since an OD of 1 corresponds to approximately 50µg/ml of double-stranded DNA (Sambrook *et al.*, 1989). Using these calculated concentration values, DNA was diluted to 5ng/µl using dH₂O and this “working” DNA solution was stored at 4°C. DNA samples were labelled with the family ID number, the individuals’ sample numbers and the concentration of the

sample. The original “stock” DNA solutions were aliquoted into 3 separate Nunc tubes and stored at -20°C .

2.3.1.4 Evaluation of DNA Samples

Agarose gel electrophoresis was used to evaluate the quality of the DNA by viewing the amount of shearing present in the sample. Pure DNA should show a discrete single band without smearing (see Figure 6). Preparation and examination of agarose gels were carried out according to Sambrook *et al.* (1989). In brief, a 0.8% agarose gel (4g of agarose powder [Promega] in 500ml of 1xTAE buffer [0.04M Tris-acetate; 0.001M EDTA], with 0.05 $\mu\text{l}/\text{ml}$ of ethidium bromide [Sigma Molecular Biology]) was poured onto an 80ml mini-gel tray, small toothed combs were inserted at the top and middle of the gel, and the gel allowed to set for approximately 40 minutes (due precautions were taken at all times in regard to the chemical hazard generated by ethidium bromide). The combs were removed and the gel placed in a mini-sub electrophoresis tank (BIORAD) with 1xTAE buffer. The sample consisting of 10 μl of 5ng/ μl DNA and 1 μl of 6 x Ficoll loading buffer (0.25% bromophenol blue; 0.25% xylene cyanol FF; 15% Ficoll [Type 400; Pharmacia] in water) was in turn loaded on to the gel. DNA samples of known concentrations and a Lambda size control (λPstI 9: 0.5 $\mu\text{g}/\mu\text{l}$, HpaII cut [Fisher Biotec]) were also loaded as standards. The gel was run at a constant 50 Volts for 1 hour, then viewed under UV light on a Hoefer Mighty Bright ultraviolet transilluminator (Hoefer Scientific Instruments). The gel was then photographed with a Polaroid camera.

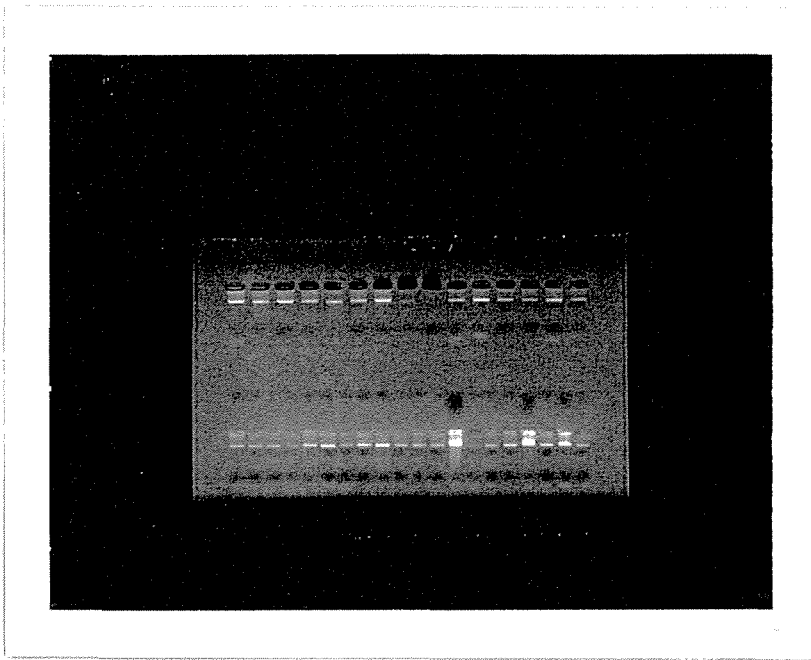


Figure 6: Agarose gel of genomic DNA

2.3.2 Microsatellite Analysis

2.3.2.1 Microsatellites

The human genome contains many polymorphic microsatellite markers. Microsatellites are short (less than 500 bp) tandem repeats interspersed throughout the genome at about 30 kb distances, with repeat sequence motifs of one to five nucleotides (Weber, 1990a). The high degree of polymorphism displayed is due to the variability in the number of tandem repeats and it is noted that the greater the

number of repeats at a locus, the more likely that the locus is polymorphic (Weber, 1990b). The number of times a sequence is reiterated at a particular locus has been shown to be sufficiently stable within pedigrees, but polymorphic among members of a population (Weber and May, 1989). Hence, microsatellites are very useful genetic markers (see section 1.8.1.2).

Microsatellites can be analysed efficiently using the PCR technique due to the small size of the microsatellite amplification products (100 to 500 bp; Edwards *et al.*, 1992). These amplification products can also be easily detected on polyacrylamide gels.

2.3.2.2 Selection of Microsatellites and Synthesis of Primers

In recent years, microsatellite maps covering the whole human genome have been developed. For the majority of the genotyping experiments in this study, existing polymorphic microsatellites that have been mapped meiotically by Genethon (Gyapay *et al.*, 1994), the Cooperative Human Linkage Centre (CHLC; Murray *et al.*, 1994), and other groups (Buetow *et al.*, 1994; Collins *et al.*, 1996) have been used. The majority of the marker distances used in the analysis were obtained from the Genetic Location Database (LDB; <http://cedar.genetics.soton.au.uk>), generated by the Wessex Human Genetics Institute and the Genetic Epidemiology Research Group. These map locations have been obtained by integrating data of different types (genetic linkage maps, physical maps, cytogenetic data and mouse homology) and constructing a single “summary” map (Collins *et al.*, 1996). A total of 362 microsatellites were chosen for stage one of this study and were genotyped in 90 U.S.A. families by two collaborative groups (Stanford University in the U.S.A. and the Edith Cowan University/Graylands Hospital in Perth, Australia). One hundred

and fifty seven additional markers were typed in these 90 families in stage two of the project. These additional markers were chosen to (1) saturate genomic regions where evidence of linkage (lod score > 1) had been obtained during stage one; (2) follow-up specific candidate regions, such as chromosome regions 6p and 15q, as well as some reported by the IMGSAC (1998) (chromosome region 7q); (3) fill in regions with large gaps in the initial screen that had not been excluded in the preliminary analysis.

A total of 50 stage one microsatellites and 50 stage two microsatellites were run on the 90 U.S.A. families as part of this PhD project. A further 41 Australian families were genotyped using a subset of 24 markers. The microsatellites analysed in this PhD study are listed in Appendix 6.1, together with the heterozygosity values (amount of polymorphism) and the number of nucleotides in the repeat sequence for each marker. Assuming a sex-averaged total autosomal map length of 3500cM, the average distance between all markers used in stage one of the study was 10cM (see Appendix 6.1).

The use of these markers was facilitated by commercially available pairs of oligonucleotide primers, with one of each pair (usually the forward primer) labelled with a fluorescent dye on its 5' end. In those cases where the markers were analysed using radioactive methods (see Appendix 6.1) the primers were unlabelled. The fluorescent primers were available as sets that allowed multiplexing (analysis of more than one primer at a time) to occur both in the PCR reaction and during electrophoretic separation (see ABI PRISM™ Linkage Mapping Set [PE Applied Biosystems, 1994] for more detail). Forward and reverse primer pairs were supplied, or custom synthesised if not commercially available, by commercial suppliers (Bresatec, Australia; Applied Biosystems (ABI), Australia; Research Genetics, USA) and couriered in either a lyophilised form at ambient temperature or diluted to 8-

10 μ M with dH₂O. Upon arrival, lyophilised primers were resuspended in dH₂O to a stock concentration of 1 μ g/ μ l. These primers were further diluted to a working concentration of 100ng/ μ l, in most cases, by adding 50 μ l of forward primer to 50 μ l of reverse primer and 400 μ l dH₂O. Primers supplied at 8-10 μ M were generally diluted to 4-5 μ M by adding equal volumes of forward and reverse primers (see Appendix 6.1 for dilution protocol of each microsatellite used in this PhD project). Both the “stock” and the “working” solution of primers were stored at -20°C.

2.3.2.3 PCR Amplification and Different Methods of Labelling the PCR

Products for Subsequent Detection

Microsatellites were amplified using the polymerase chain reaction (PCR). PCR is an *in vitro* method for enzymatically synthesising defined sequences of DNA. The reaction uses two oligonucleotide primers that hybridise to opposite strands and flank the target DNA sequence that is to be amplified. The elongation of the primers is catalysed by Taq DNA Polymerase, a heat-stable DNA polymerase (Saiki *et al.*, 1988) that has been isolated from the thermophilic eubacterium *Thermus aquaticus*. A repetitive series of cycles involving template denaturation, primer annealing, and extension of the annealed primers by Taq DNA Polymerase results in exponential accumulation of a specific DNA fragment, where the ends of the fragment are defined by the 5' ends of the primers (Ausubel *et al.*, 1991).

There are a number of different methods for detecting the target DNA sequence. Examples include; silver staining of conventional PCR products, incorporation of radioactive (α -P³²) labelled deoxycytosine triphosphate (dCTP) nucleotides into the PCR reaction, incorporation of fluorescently-labelled deoxyuridine triphosphate (dUTP) into the PCR products and labelling the 5' end of

the primer with a fluorescent tag. Most primers used in this study were labelled with a fluorescent tag at the 5' end, and the PCR products were identified using the ABI PRISM 310 DNA Analyser (capillary electrophoresis) or the ABI PRISM 373 DNA Analyser (gel electrophoresis). In the cases where the primers were not fluorescently labelled, radioactive nucleotides were incorporated into the PCR products and products were separated on polyacrylamide gels and detected by autoradiography. For a diagrammatic representation of the overall genotyping method see Figure 7.

The PCR procedure used in this study was a slight modification from the original (Mullis and Falcon, 1987). Polymerase chain reactions (Erlich *et al.*, 1991) were carried out using different protocols depending on the thermocycler utilised and the condition optimised for each specific microsatellite. Generally, PCR reactions carried out in a MJ Research thermocycler were performed in 5µl total volume and contained 10ng DNA, 200µM dNTPs, 0.13-2.0µM or 1.0-20ng/µl of each primer (the forward primer had been fluorescently labelled using either HEX, TET and FAM dyes), 50mM KCl, 10µM Tris-HCl, 1.5mM MgCl₂, and 1.25U Taq-Polymerase (Perkin Elmer) (refer to Appendix 6.1 for the reaction mixture of each microsatellite).

The solutions were pipetted into a capillary tube (Biotech) using a programmable pipettor (Corbett Research), and processed in a thermocycler (MJ Research). After denaturation for 5-10min at 94°C (5min for standard AmpliTaq and 10min for AmpliTaq Gold (Perkin Elmer); see Appendix 6.1), 27 cycles were carried out, each cycle consisting of 30sec at 94°C, 75sec at 55°C, and 15sec at 72°C. A final extension step was performed for 5min at 72°C. The DNA polymerase AmpliTaq Gold is activated by heat and allows "hot start" PCR to be introduced into

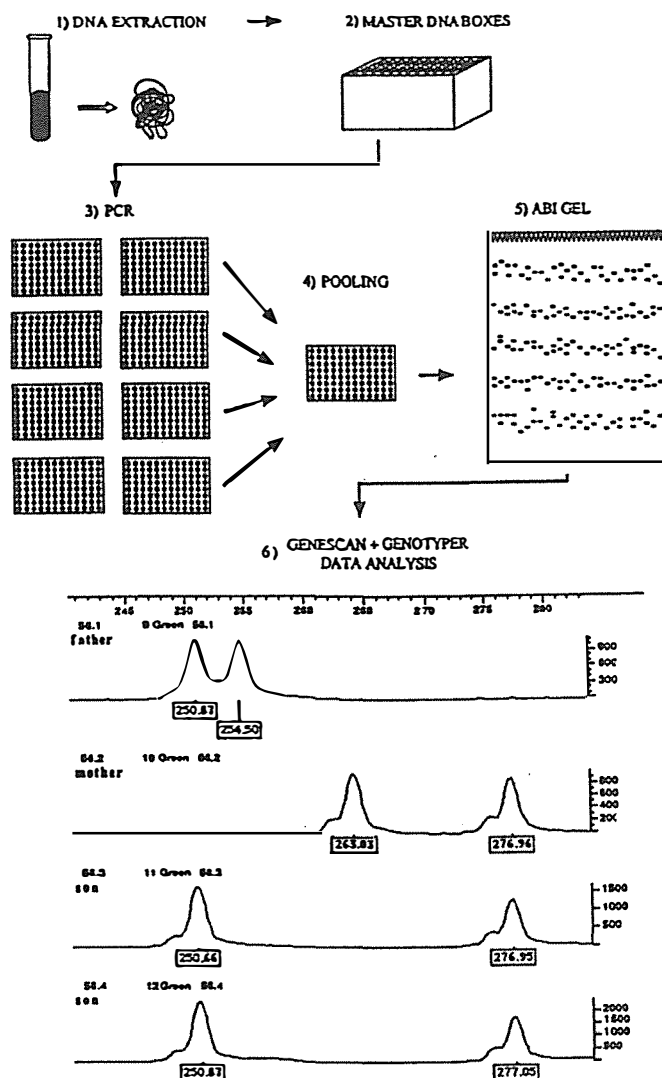


Figure 7: Outline of the genotyping method. The electropherograms in the bottom section show the results from one family with two affected sibs that were produced by the analysis programme Genotyper (see 2.3.2.6) (diagram from Maestrini *et al.*, 1998).

existing amplification systems with little to no modification of cycling parameters or reaction conditions. Hot start PCR can improve amplification of most templates by

lowering “background” and increasing the amplification of specific products. This is achieved by the covalently bound moiety to the polymerase. When the enzyme is heated to approximately 94°C, the moiety is detached from the polymerase and the polymerase becomes active. Since this temperature is above the minimum required for the primers to anneal, it prevents the primers annealing to non-specific regions (mis-priming) and, in turn, prevents the formation of artifact peaks, which can complicate the analysis.

The majority of microsatellites were amplified using the 2400 or 9600 Perkin Elmer Thermocyclers. PCR conditions differed slightly between markers (see Appendix 6.1), however they basically involved a final reaction volume of 5µl, in which 2µl was DNA (5µg/ml). The remaining volume contained 250µM dNTPs, 0.04-1.92µM or 13-25ng/µl of each primer (the forward primer had been fluorescently labelled using HEX, TET and FAM dyes), 50mM KCl, 10µM Tris-HCl, 1.56-3.12mM MgCl₂, and 1.25U Taq-Polymerase or Amplitaq Gold (Perkin Elmer). All PCR reactions were performed in 200µl tubes on a 24 or a 96 well plate, under either a standard or a touchdown cycling programme (see Appendix 6.1). The touchdown cycling programme was used when amplification resulted in poor yields of a specific product using the standard programme. The touchdown programme differs from the standard programme by reducing the annealing temperature in the second group of cycles. By reducing the annealing temperature, the primers are prevented from annealing to non-specific regions. The lower temperature also prolongs the activity of the DNA polymerase, increasing the total yield of product. The standard programme involved an initial denaturation time of 5min at 94°C for standard Taq polymerase and 15min at 94°C if Amplitaq Gold was used in the solution. This was followed by 10 cycles of 15sec at 94°C, 15sec at 55°C and 30sec

at 72°C. Then 20 cycles of 15sec at 89°C, 15sec at 55°C, 30sec at 72°C were performed with a final elongation step of 10min at 72°C. The touchdown programme also entailed an initial denaturation at 94°C for 10min or 5min depending on the type of Taq polymerase used. The first 15 cycles of this programme comprised 20sec at 94°C, 60sec at 63°C - 55°C (-0.5°C annealing temp. /cycle) and 30sec at 72°C. The second 15 cycles involved 20sec at 94°C, 60sec at 55°C and 30sec at 72°C, with a final extension at 72°C for 5min. For microsatellites where radioactive labelling was used, the reaction mix was a slight modification of the original procedure by Saiki *et al.* (1988) and included 0.05µl of 10mCi/ml α -P³² dCTP nucleotides per reaction (see Appendix 6.1 for PCR conditions).

After completion of PCR reactions, products analysed on the Perkin Elmer 310 DNA Analyser were diluted with water and an aliquot (see Appendix 6.1) added to 12µl of formamide and 0.8µl of GS500 Tamra standard dye (Figure 8). Samples loaded on to the Perkin Elmer 373 DNA Analyser were multiplexed to a total volume of 1.5µl and combined with 0.5µl of the internal lane size standard (Tamra GS500; Applied Biosystems), 2.5µl of deionised formamide and 0.5µl of Blue Dextran EDTA (1mg Blue Dextran per 1ml EDTA [20-50mM, pH 8.0]). Formamide was added to disrupt the hydrogen bonds within the PCR products and prevent the products from re-annealing at a lower temperature once they were denatured. At least 50% of the final sample loaded on to the gel was formamide. The samples loaded on to the 310 DNA Analyser contained an excess of formamide to allow up to 10µl of PCR products to be run in each injection. PCR products loaded on to the SE1600 Poker Face II sequencer were combined with 8µl of formamide loading buffer in a 0.65ml microcentrifuge tube. All samples were denatured for 5min at 94°C before loading on to the detection apparatus.

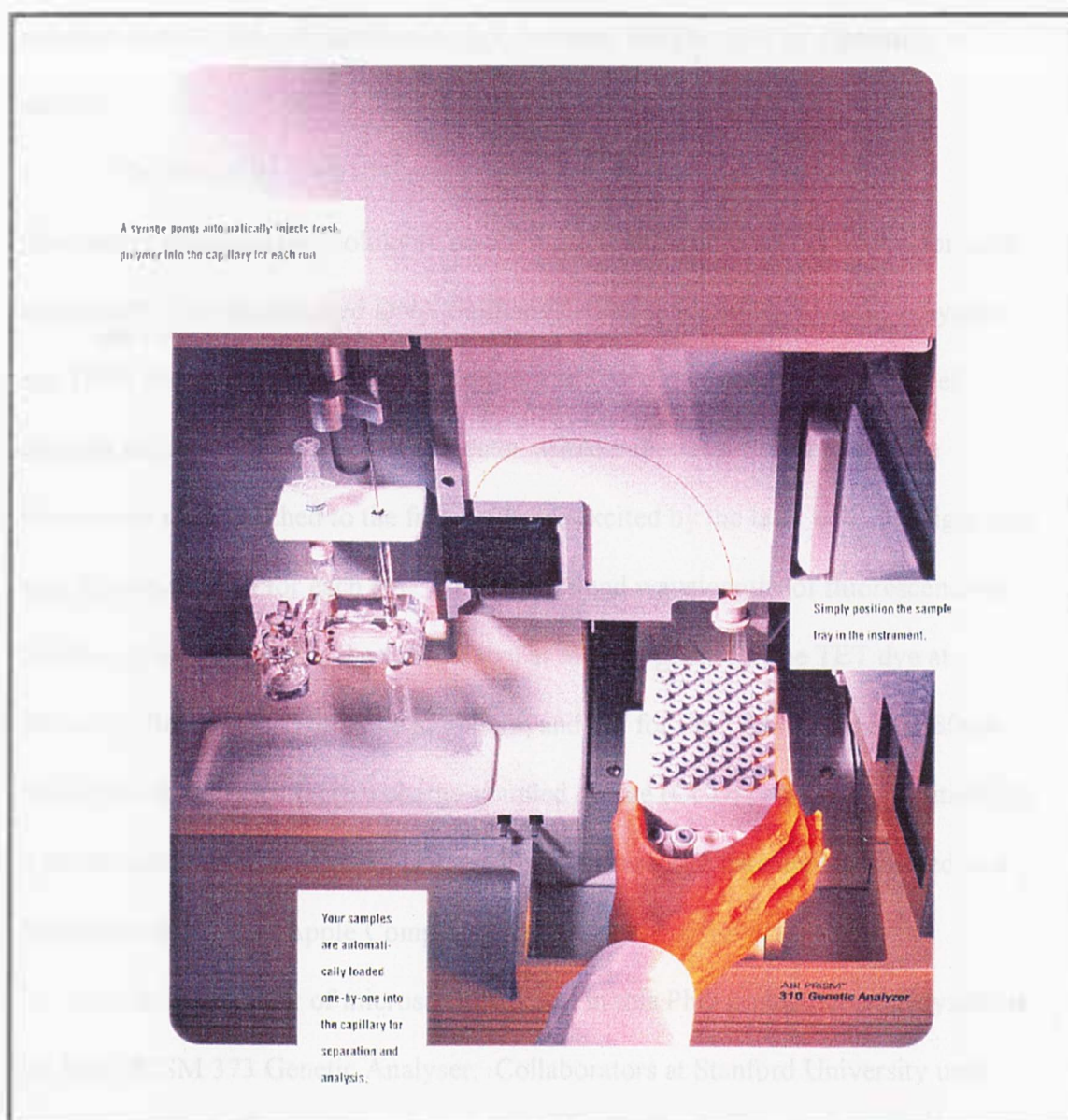


Figure 8: Loading samples onto the Perkin Elmer 310 DNA analyzer

2.3.2.4 Electrophoretic Separation and Detection of the PCR Products of Different Length

The PCR products for each microsatellite (ie. the alleles) were discriminated by length fractionation using polyacrylamide gel electrophoresis (PAGE). PAGE operates via the movement of negatively charged DNA towards the positive electrode, under the influence of an electrical current. In the case of capillary electrophoresis (see below) the alleles separate according to size in a polymer

solution instead of a polyacrylamide gel, however the principle of separation is similar.

The Perkin Elmer 310 Genetic Analyser uses real-time four-colour fluorescent detection technology (Connell *et al.*, 1987) to detect the alleles for each individual. The dye-labelled DNA fragments electrophorese through the polymer, and DNA fragments separate according to size. As the labelled samples travel through the capillary and into the detector window they are illuminated. The fluorescent dyes attached to the fragments are excited by the laser and emit light at a specific wavelength for each dye. The colours and wavelengths of fluorescence of the four dyes are; blue for the 6-FAM dye at 531nm, green for the TET dye at 545nm, yellow for the HEX dye at 560nm, and red for the TAMRA dye at 580nm. The signal is collected on to a charge-coupled device (CCD) camera and digitised by a photomultiplier tube (PMT). The resulting fluorescence data are transmitted to a Macintosh computer (Apple Computer) for analysis.

A small number of microsatellites used in this PhD project were analysed on an ABI PRISM 373 Genetic Analyser. Collaborators at Stanford University used a similar machine to visualise and size all of their PCR products (Perkin Elmer 377 DNA Analyser). The Perkin Elmer 373 system operates in a similar way to the Perkin Elmer 377 system, except that it contains different filters and the products are run for a longer period of time. The ABI Prism 373 system also uses real time, four-dye, one-lane technology to detect the DNA fragments for each microsatellite.

2.3.2.5 Preparation of Detection Systems for Analysis

The ABI Prism™ 310 DNA Analyzer (Figure 9)

A detailed description of the running of the GeneScan application for this apparatus can be found in the ABI Prism™ 310 Genetic Analyzer User's Manual (1995)[Chapter 4, Running GeneScan], distributed by PE Applied Biosystems Inc.

In brief, the polymer (Performance Optimized Polymer 4, PE Applied Biosystems) was added to a 1ml glass syringe and set into position, being sure no air bubbles were present in either the syringe or the block. The buffer (10x Genetic Analyzer Buffer with added EDTA, PE Applied Biosystems) was diluted to 1x with distilled water and poured into the buffer vial located on the autosampler, as well as the buffer reservoir attached to the pump block. Distilled water was also added to a buffer vial and an Eppendorf tube located on the autosampler. GeneScan capillaries were cut down to 30cm length and the laser detection window on the capillary washed with 70% methanol to remove any particulate or fluorescent matter that may also be detected by the laser. The autosampler was then re-calibrated and the samples (PCR products with formamide and standard added) placed into the autosampler tray. Finally a GeneScan Injection List was prepared which states the sample names and the assignment to an autosampler position. The Injection List also specifies the order of samples and the modules containing the parameters to be used for the run. In this project most samples were injected into the capillary for 5-10sec and run through the capillary and past the laser detection window for 17min. This gave enough time for microsatellites up to the size of 350bp to be accurately separated and analysed. After 96 samples have been analysed, the polymer and buffer were exchanged for fresh media. After 200 runs the GeneScan capillary was also replaced.



Figure 9: ABI Prism™ 310 DNA analyzer

The ABI Prism™ 373 DNA Analyzer (Figure 10)

The protocol for the preparation of gels generally followed the instructions of the manufacturer (GeneScan 672 Software User's Manual, 1993). A detailed description of the procedure can be found in the 373 DNA Sequencer User's Manual (1996)[Chapter 2, Preparing Gels], distributed by PE Applied Biosystems Inc.

Special care was taken to eliminate any particulate or fluorescent matter from the acrylamide gel solutions and to keep the glass plates extremely clean and free of dust, as the laser system detected any fluorescence present interrupting the reading of the fluorescent DNA fragments. All solutions were filtered and only ultrapure, high-grade reagents and distilled, deionised water were used to prepare the solutions.

In brief, 75ml of fresh 6% acrylamide-urea gel solution (6% acrylamide, 8.3M urea [Molecular Sigma Biology], 1xTBE; see Appendix 6.2 for protocol) was poured into a 150ml beaker, and to this, 400µl of 10% ammonium persulphate (Molecular Sigma Biology) and 45µl of TEMED (N,N,N',N'-tetramethylethylenediamine; Molecular Sigma Biology) were added. The acrylamide solution was carefully poured between the two plates by holding them in an upright position and resting the lip of the solution vessel against the edge of the notched plate. The plates were then placed in a horizontal position and a 48-tooth comb secured with two large binder clamps. Gels were left to set for a minimum of two hours to ensure complete polymerisation prior to use.

Before the gel was placed in the electrophoresis chamber, the comb and the clamps were removed and both sides of the plates were rinsed with water (to clean away the excess acrylamide and urea), and then air-dried. To ensure that no peaks were produced by fluorescent particles either on the plates or in the gel, the gel and plates were scanned before loading the samples. 1xTBE buffer (89mM Tris-base, 89mM Boric acid, 2mM EDTA; pH~8.3 at ambient temp.) was then placed in the upper and lower buffer chambers and the sample wells were flushed with this running buffer using a pipette. The run parameters were set at Filter set B, field 2500 volts, a current of 40mAmp, power at 30 watts, a laser setting of 40mW, and a PMT voltage between 600-800 Volts. Once the wells were loaded with the samples, the gel was run for 10 hours.

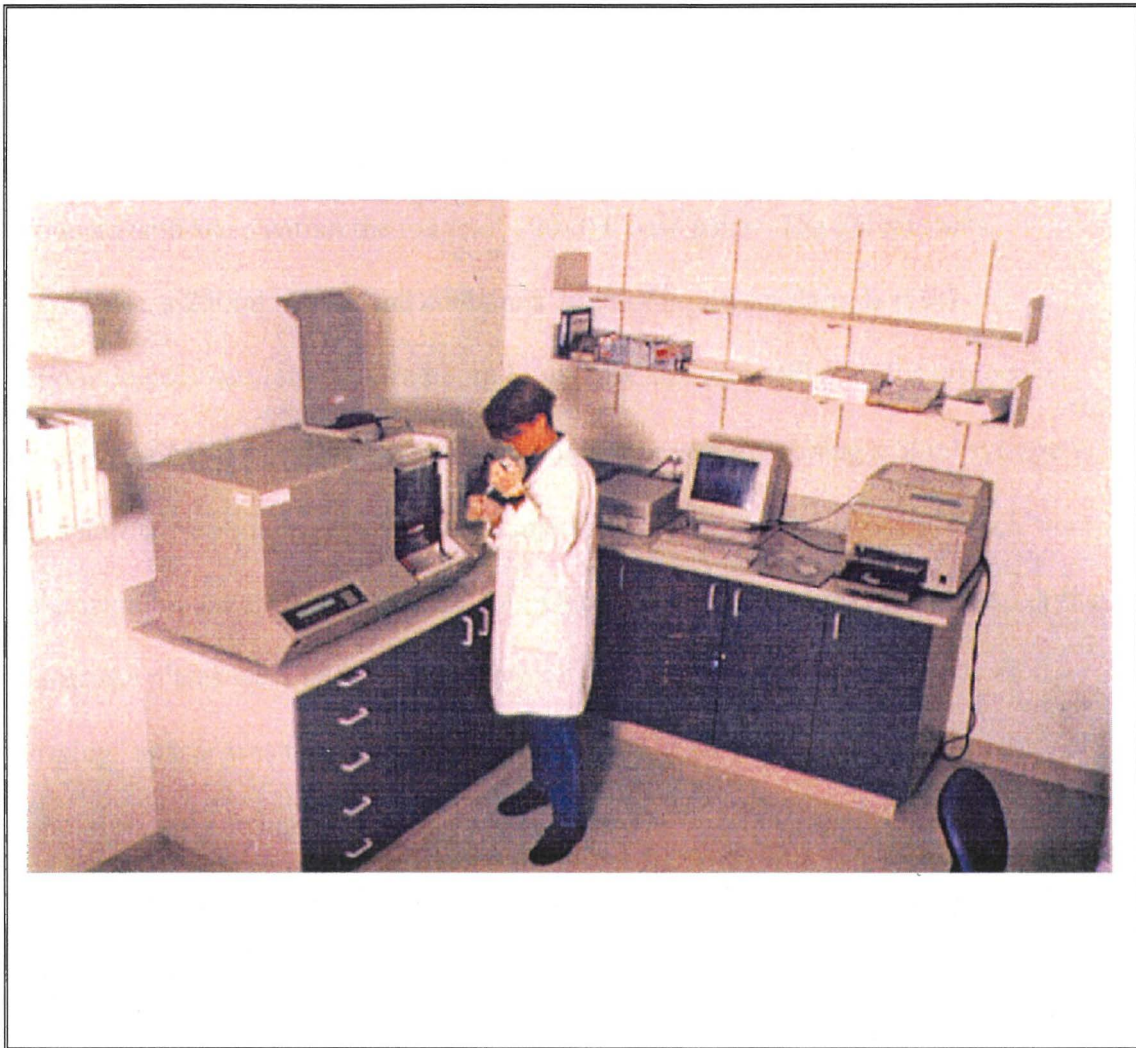


Figure 10: ABI Prism™ 373 DNA analyzer

SE1600 Poker Face II

The protocol for preparation of these sequencing gels was similar to that used for the 373 DNA Sequencer (see above). Since radioactive products were detected using this system, care taken to prevent any fluorescent particles in the acrylamide gel solutions and on the glass plates was not required. Hot tap water and liquid Pyroneg (Diversey) detergent was used to wash the plates before use and 70% ethanol assisted in drying. The smaller of the two plates was treated with Sigmacote (Molecular Sigma Biology). This silicone solution prevented the gel from sticking

tightly to both plates, and reduced the possibility that the gel would tear when removed from the mould after electrophoresis.

A 6% polyacrylamide gel was used since it provides best resolution for products with sizes within the range of 100-350 base pairs. The gel mix was prepared in a 250ml beaker and contained 70ml of liquid acrylamide (19:1 acrylamide:bis-acrylamide; Gradipore Liqui-Mix; see Appendix 6.2 for protocol), 170 μ l of TEMED and 170 μ l of 25% ammonium persulfate. Using a 50ml syringe, the solution was quickly drawn up and injected between the glass plates, which were held on an incline of approximately 20°. Once the gel mix had reached the top of the plates, the plates were laid down horizontally and the wedge removed. A “well-forming” spacer was then inserted and the gel left to set at room temperature for approximately one hour. If the gel was to be used the following day, the top and bottom of the gel were kept moist with paper towels and sealed in plastic wrap. It was then stored flat at 4°C.

The electrophoresis tank used was a SE1600 Poker Face II sequencer (Hoefer Scientific Instruments). This was attached to a 3000Xi-electrophoresis power supply (Bio-Rad). The first step was to fill the lower buffer chamber with 1xTBE buffer (see Appendix 6.2 for protocol) to a depth of approximately 2cm. The masking tape on the glass plates was then removed and an aluminium and perspex backing plate, which provides even heat distribution across the gel, was clamped against the smaller glass plate. This was done with two side clamps of the same length as the larger glass plate. This apparatus was placed in the lower buffer tank and secured, and the upper buffer tank was filled with 1xTBE buffer. The “well-forming” spacer was removed and the well flushed with 1xTBE buffer to remove excess urea and prevent distorted migration of DNA due to unpolymerised acrylamide. The top “safety lid”

and top and bottom connectors were then plugged into the unit, and the electrodes were connected to the power pack.

Before loading the PCR products, each gel was pre-warmed for 20 minutes at 1400 volts. A 66-well comb or a 44-well comb was placed in the well formed by the “well-forming” spacer, and 1 to 2µl of each sample was loaded into each well (1µl was loaded if the 66-well comb was used, 2µl for the 44-well comb). The top “safety lid” was replaced and the gels were run for 1.5-3.5 hours at 1400 volts depending on the size range of the alleles.

After gel electrophoresis, the glass plates were detached using a wedge and the upper plate (without the gel), as well as the spacers, were removed. Each gel was fixed by gently flooding with at least 200ml of 10% methanol (AnalaR, BDH LabSupplies) and 10% acetic acid (AnalaR, BOH LabSupplies) for 15 minutes. After removing any excess moisture by placing the plate on a slight incline, a piece of blotting paper, slightly larger than the gel, was placed over it. The adhering gel was peeled off the glass plate, covered with plastic wrap, and dried under vacuum on a slab gel dryer (Savant SGD4050) for 1 hour at 80°C. The dry gel was placed in an X-ray cassette (Amersham Hypercassette™) with a sheet of X-ray film and held overnight at room temperature. The X-ray film was developed using a Curix benchtop automatic processor (Agfa).

2.3.2.6 Allele Assignment

The size of the microsatellite PCR products was determined by reference to the internal lane size standards by applying the method of Carrano *et al.* (1989). The 310 GeneScan™ Analysis 2.1 (PE Applied Biosystems, 1996) software automatically recognises the fragment size distribution of the size standards labelled

with the TAMRA dye. The size of the unknown DNA fragments labelled with the other three dyes (HEX, TET and FAM) is determined by correlation with the internal GeneScan size standard Tamra (GS500; see Figure 11). The GeneScan Analysis software generates a calibration or sizing curve based upon the migration times of the known sized fragments in the standard, relative to which the unknown fragments are mapped and converted from migration times to sizes. Since the DNA fragments of the standard are being electrophoresed under the same conditions as the sample, this is a very precise way to size unknown DNA fragments. It is also very replicable and allows genotypes run in different lanes and on different gels to be compared.

The data collected by the GeneScanTM Analysis 2.1 software were analysed by the Genotyper^R 2.0 software (PE Applied Biosystems, 1993-6) to determine the microsatellite allele sizes for each individual at each locus. Products detected radioactively were sized relative to a control PCR product run on the same gel whose alleles had already been assigned a number. This was achieved by numbering each allele in a “top-down” fashion, such that the largest fragment was numbered one, the second largest numbered two, and so on. This procedure allowed easier identification of the allele number for the newly run products.

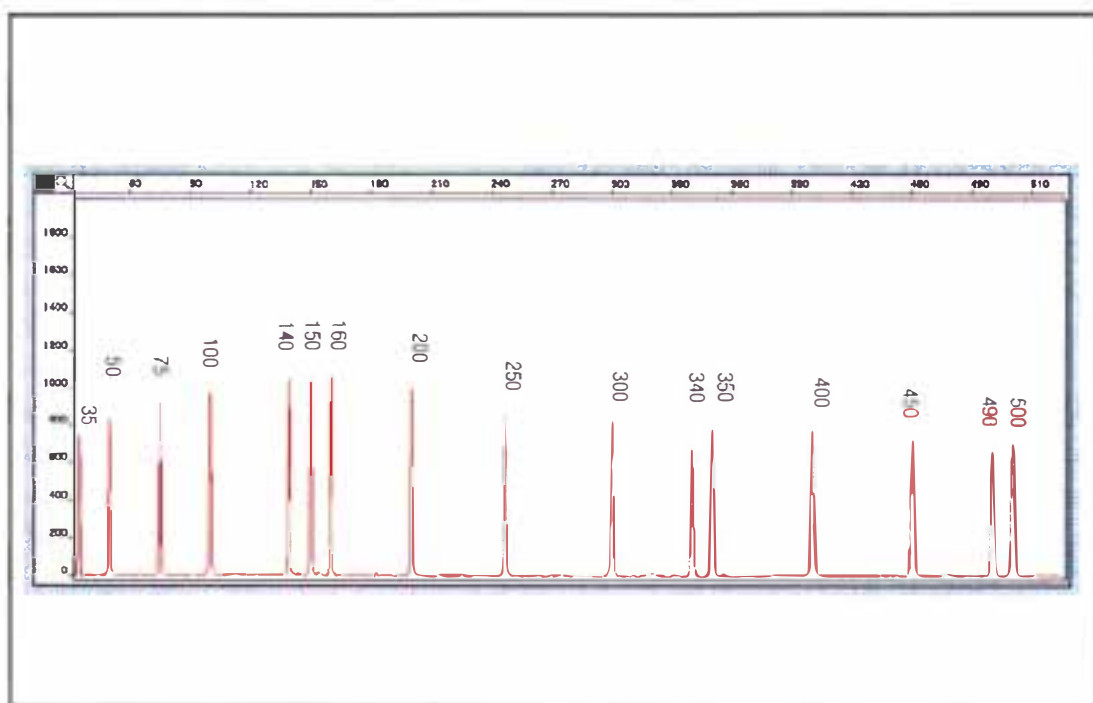


Figure 11: Tamra GeneScan-500 size standard

2.3.3 Verification and Processing of Genotyping Data

Most data in this study (except those obtained using the radioactive procedure) were collected using an automated procedure. The data for the sizes of each allele were produced by the GeneScan^R Analysis 2.1 software (PE Applied Biosystems, 1996) and analysed by the Genotyper^R 2.0 software (PE Applied Biosystems, 1993-96). The analysed data were then transferred into “Cyrillic for Pedigree Drawing” programme (Chapman, 1993-1996) where the haplotypes were checked and then automatically moved into MLINK (Chapman, 1993-1996) where it could be directly analysed by the linkage programme ASPEX (see section 2.4). This automatic data collection, transfer and storage system produced reliable scoring of allele sizes, and human error caused by manually entering the data was minimised.

To ensure the accuracy of results, an independent party manually checked random genotypes. Two monozygotic twin pairs were also included in the analyses as a positive control and to estimate the rate of genotyping error. Chromosomal regions showing positive lod scores were also replicated in either the Perth or Stanford laboratory depending on where the results originated, using the same 90 U.S.A. families. As an added quality control on the accuracy of marker typing, the order and distances of microsatellites were calculated using the data collected from our sample of individuals. Since the genetic map positions for those markers have already been determined in other large collections of families (eg. the CEPH families), the correlation between these maps showed the accuracy of our data. The programme used to construct these genetic maps is a component of the linkage programme ASPEX (see below).

To ensure that the methodology was sufficiently sensitive to detect a real excess of sharing when such excess was present, the pseudoautosomal marker DXYS154 was included in the analysis. This marker does not recombine between the X and Y chromosome, making it completely sex-linked. Since 73% of the U.S.A. ASPs are concordant for sex (63% males and 10% females), we would expect, at locus DXYS154, excess sharing of alleles inherited from the fathers, but not of those inherited from the mothers. This provided an additional positive control.

2.4 Statistical Analysis

2.4.1 Statistical Model

The power of the overall study to detect susceptibility loci to autism has been determined under an oligogenic model, with the assumption that autism is caused by at least three genes. Using multipoint analysis formulae described by Risch (1990b), the expected maximum lod scores (EMLS) for various values of λ_s and numbers of sib pairs (N) were calculated (Table 5). Based on the assumptions that all parents are typed and the polymorphic information content (PIC) of all 350 markers (10cM map) is 0.70, it is reasonable to expect that we can identify loci with $\lambda_s=2.0$ or more in a sample of 100 affected sib pairs. With these values we can also expect to exclude 87% of the genome for a locus with $\lambda_s=2.0$. Calculations show that with the typing of extra markers and the addition of 100 more multiplex families, we should be able to exclude the entire genome for a λ_s greater than or equal to 2.0 (ie. a gene of moderate effect on the disease), if a gene of this influence is not present in the disorder (Table 7).

Table 7: Expected Proportion of Genome Excluded, PIC=0.70

N	$\lambda_s=2.0$	$\lambda_s=2.0$
	10cM map	5cM map
100	0.87	0.92
200	0.99	0.99

2.4.2 Linkage Analysis

Statistical analyses were based entirely on identity-by-descent (IBD) data since at least one parent was available for all sibling pairs. These data were calculated by identifying the parental origin of each sibling allele, and comparing the corresponding alleles for each member of an affected sibling pair, at a given locus, for each parent. If both alleles could be uniquely identified, the pair was scored as either identical or non-identical by descent from that parent at that position. If the match was ambiguous, the position was scored as uninformative.

In a family with more than two affected siblings, the first affected sibling was paired with each remaining affected child, allowing the contribution of several independent sib-pairs within the one family. While this did not bias the observed sharing results, it did mean that data from large families carried more weight than data from small families.

Multipoint sib-pair analysis (see section 1.8.3) was employed (Risch, 1990b; Risch, 1993) in this study to test for a susceptibility gene(s) to autism. This method of analysis has already been used to construct an exclusion map of chromosome 17 for type 1 diabetes (Williams *et al.*, 1994), to suggest evidence of a susceptibility locus in chromosome 6 for schizophrenia (Schwab *et al.*, 1995), and to construct an exclusion map of the X chromosome for autism (Hallmayer *et al.*, 1996). The multipoint sib-pair method of linkage analysis has also suggested evidence for susceptibility loci in autism, with a lod score of 2.53 calculated in one region on chromosome 7 assuming $\lambda_s = 5$ (IMGSAC, 1998).

The multipoint lod score at a particular locus was determined by finding the nearest informative markers flanking the locus on either side, for both the maternal

and paternal alleles. The probability of the given marker data (ie. the probability of sharing vs. non-sharing of alleles), assuming the locus was in a particular IBD state, was calculated from the IBD states of these flanking alleles, and the recombination distances to the locus. These probabilities were then used to determine the odds of the marker data (lod score) given a disease gene of known penetrance at the locus, versus the odds of the marker data if the disease gene had zero penetrance. The calculation of marker allele frequencies was not required in this method of analysis, since all statistical analyses were based entirely on IBD data.

Multipoint lod scores were calculated for a fixed model (ie. fixed value of λ_S), allowing an exclusion map to be formulated. For a fixed value of λ_S , negative lod scores can be obtained if the model gives a poorer representation of the marker data than does no gene effect. The conservative lod score criterion of -2 was used in this study for exclusion of a locus with a given λ_S value (Hauser *et al.*, 1996). Maximum lod scores at each locus were also obtained by maximising the likelihood of the marker data at each point, as a function of the genetic model (λ_S value). These results never drop below zero, the value obtained when $\lambda_S = 1$.

Regions identified as possibly containing susceptibility loci (Lod Score > 1) were analysed further using two-point linkage analysis. Two-point linkage analysis calculates the probability that a marker and a disease locus are linked and does not consider the adjacent markers on the genetic map. This was evaluated by calculating the number of alleles shared versus the number of alleles unshared at each marker across sib-pairs and comparing with a chi-square (χ^2) distribution.

2.4.3 Linkage Disequilibrium Analysis (LD)

Linkage disequilibrium was examined in chromosomal regions showing suggestive evidence of linkage ($\text{lod} > 1$) using the transmission/disequilibrium test (TDT) of Spielman *et al.* (1993; see section 1.8.4). This programme identified all parents who were heterozygous at the marker locus, and defined two alleles for each affected child, the one transmitted to the affected child and the one that was not. All transmitted alleles from these heterozygous parents were then tallied with all non-transmitted alleles. A significant difference (by the χ^2 test) in the frequencies of alleles in these two groups indicated that an allele (or alleles) was (were) preferentially transmitted to affected children, and hence associated with disease susceptibility.

2.4.4 Computer Programme

All statistical analyses, including the multipoint and two-point sib-pair analyses and the TDT analyses, were performed with the ASPEX computer package (Hinds and Risch, 1996). It was also used to confirm the relationship between members of each family, and to search for and eliminate unlikely double crossovers in the multipoint analyses in regions that had dense markers.

The programme used in the ASPEX computer package to perform multipoint exclusion mapping using affected sibling pair data for discrete traits (Hinds and Risch, 1996) was SIB IBD. This programme produces a default output file that summarises the sharing at each locus, broken down by parent. The markers were

sorted into order along the chromosome, as specified in the loc (location) parameter. A loc file was generated at the beginning of each analysis of a chromosome and states the markers used, the distances between each marker in Morgans, as well as the lambda value specified in the analysis.

A multipoint exclusion map was also generated by the programme and the genetic recombination distance (offset) and the lod scores were specified for each marker at each lambda value. Multipoint lod scores calculated at specified points between and at each marker for different lambda values ($1.5 < \lambda_s < 10$) were graphed using the “Microsoft Excel for Windows 95” computer programme (Microsoft, 1985-1996), to view areas of linkage and/or exclusion. Lambda values or the sibling risk ratio (λ_s) were used to quantify the effect of a locus. For this study, loci with a small genetic influence on the disease corresponded to a $\lambda_s \leq 2$, whereas a large gene effect coincided with a $\lambda_s \geq 5$. A moderate influence on the disease gene corresponded to a λ_s value between those stated for a small and large gene effect ($3 \leq \lambda_s \leq 4$).

CHAPTER 3

RESULTS

Two collaborative groups (Stanford University in U.S.A. and Graylands Hospital/Edith Cowan University in Perth, Australia) performed a hierarchical search of the genome for autism susceptibility genes on 90 U.S.A. multiplex families. A 10cM map containing 362 markers was used in the initial screen. Genomic regions of interest were investigated further during the second stage, with a denser map containing 157 additional markers. Using the ASPEX computer programme, multipoint sib-pair analysis was performed on the genotype data to detect linkage. Exclusion maps were generated under different lambda values to show the regions of the genome which show no evidence of linkage. Regions indicating suggestive linkage ($\text{lod} > 1$) were analysed further using two-point linkage analysis.

A total of 50 microsatellites in the initial screen and 50 additional markers in the follow-up screen were analysed as part of this PhD project on the 90 U.S.A. families. An additional group of 41 Australian multiplex families were analysed using a subset of 24 microsatellites. The results below are presented in the order in which they were obtained (see Figure 12). The analysis of the U.S.A. families is shown first by displaying the results of the systematic scan of part of the genome, followed by the refined study of candidate regions and concluding with the results from the duplication of the positive regions identified by us and the Stanford group.

The section on the study of the Australian families presents our findings for the regions identified as possibly significant during the initial scan on the U.S.A. families and by the IMGSAC (1998). Finally, the results from the linkage

disequilibrium (LD) analysis are presented on regions where results suggestive of linkage have been obtained in the combined U.S.A. and Australian group of families.

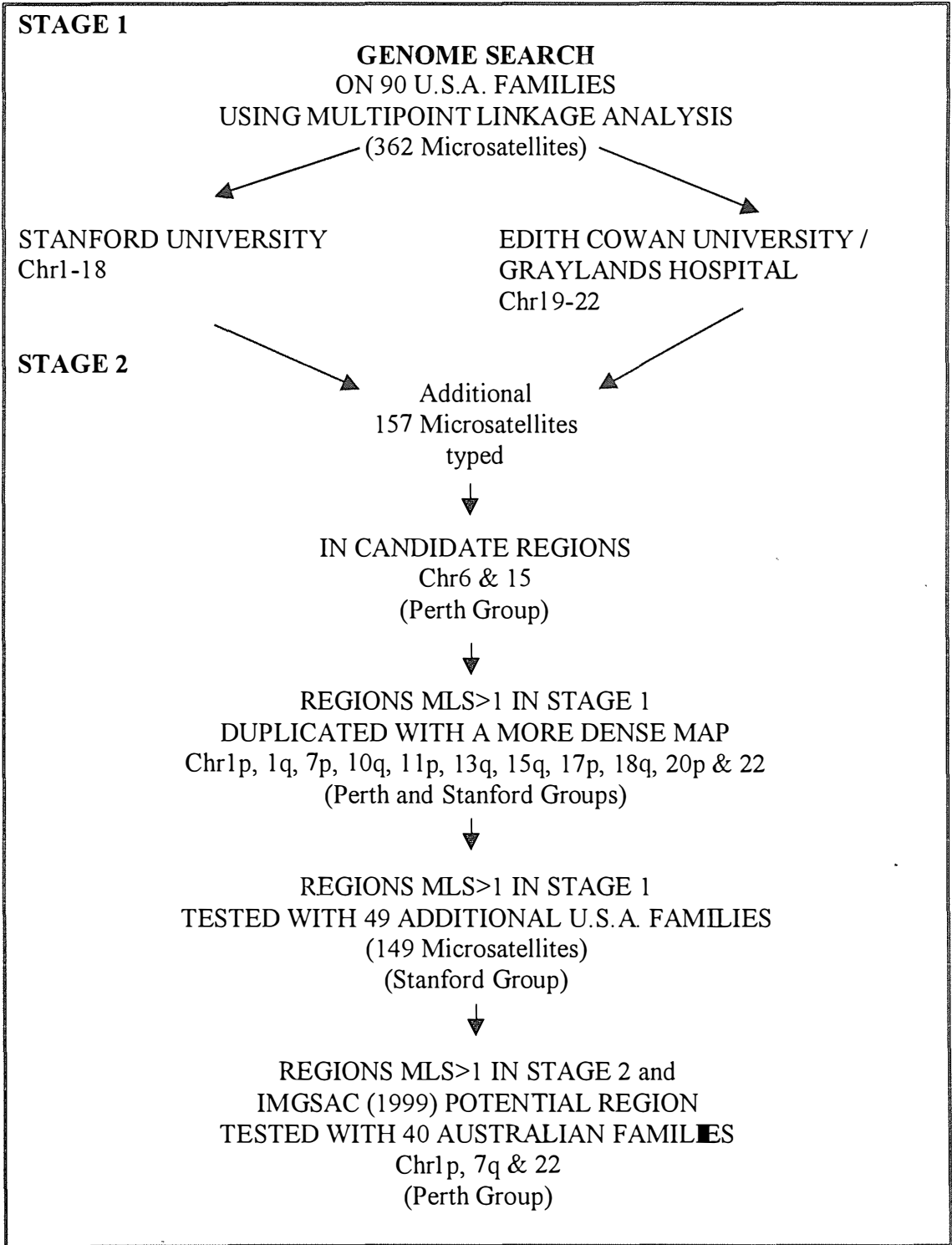


Figure 12: Flow-chart of the linkage study on autism (Note: Perth Group includes collaborators from E.C.U. and Graylands Hospital)

3.1 Linkage Results from the U.S.A. Families

3.1.1 Genotyping Error Rate

To ascertain the accuracy of the results that follow as well as the degree to which potential biases might be confounding the results, the genotyping error rate for this study was determined. A marker completely linked to sex was also run to ensure that the method was sensitive enough to determine excess in sharing at a specific locus if indeed present for this disease.

In the analysis a number of controls were employed to identify potential typing errors in the results. Firstly a pair of monozygotic twins were genotyped with the other sib-pair families to determine the genotyping error rate in the entire genome scan. Additionally, random genotypes, as well as regions showing positive lod scores ($MLS > 1.0$), were manually checked by an independent party using the same protocol. For the entire genome scan, a genotyping error rate of 1% was calculated. This result was derived from the analysis of the monozygotic twins where 10 loci out of a total of 946 were found to be genotypically discordant (Risch *et al.*, 1999).

The error rate for the genotypes included in this Ph.D. study was calculated by the analysis of random genotypes by an independent party. Five hundred and sixty-eight genotypes in total were re-analysed and 7 errors in total were found. This corresponds to a genotyping error rate of approximately 1.2%.

To ensure this method of analysis was sufficiently sensitive to detect susceptible genes, if present, DXYS154 was analysed in the U.S.A. sib-pair families (see 2.3.3) to see if linkage to sex was detected. Since the majority of ASPs are

brother pairs we would expect to an excess of sharing of alleles inherited from the fathers, but not of those inherited from the mothers. In total, there were 58 paternal alleles shared and 22 paternal alleles not shared, giving a χ^2 of 16.2. This corresponded to a lod score of 3.52, displaying significant linkage of this marker to sex. By contrast, there was no difference in the sharing of maternal alleles (42 alleles were shared and 45 alleles were not shared). Since 73% of the U.S.A. ASPs are concordant for sex, these results indicate that the methodology was sufficiently sensitive to detect an excess of sharing of this magnitude when present.

3.1.2 Systematic Scan of Chromosomes 19-22

A systematic scan of part of the genome, namely chromosomes 19-22, was completed as part of the first stage screening on 90 multiplex families from the United States of America. Multipoint linkage analysis on the 50 microsatellites genotyped was performed under different lambda values (ie. $\lambda_S = 2, 3, 4, 5$ & 10) using the ASPEX programme. The results of these analyses are depicted graphically. The figures show the lod scores calculated at each marker typed and at a number of intervals between these markers. The order of the microsatellites on the figures follows that given in the Genetic Location Database (LDB; see section 2.3.2.2) and the distance between the markers is stated in cM. A dividing line at lod score -2 shows the values of λ_S (below this line) that could be excluded from linkage to an autism susceptibility gene in that interval.

The allele sharing data at each locus are shown in table format. The "Paternal" and "Maternal" tables show the number of shared alleles that were

inherited from each parent, the number of alleles inherited from the respective parent that are not shared, the percentage of alleles shared at this locus, and whether or not there is a significant difference between shared and non-shared alleles at the designated 5% level of significance (χ^2 test). The "Combined" table shows the summary of allele sharing at each locus irrespective of parent of origin. In the cases where there was no way of differentiating whether the allele was maternal or paternal in origin, these results were included in the "Combined" sharing table. The "Most-Likely Model" table has five columns tabulating the percentage sharing and the lod score obtained by the maximum likelihood calculation (MLS). The middle three columns show the ratio of sharing of zero, one and two alleles. The results for each chromosome are described below.

3.1.1.1 Chromosome 19

Genes of Large Effect

As depicted in Figure 13, genes with effects corresponding to a $\lambda_s \geq 5$ could be excluded from linkage to autism for over 90% of chromosome 19. The region that could not be excluded was between the markers D19S420 and D19S418.

Genes of Moderate Effect

Twenty seven percent of chromosome 19 could be excluded from linkage to a gene of moderate effect ($3 \leq \lambda_s \leq 4$). This region included markers D19S221, D19S226 and D19S414.

Genes of Small Effect

No region on chromosome 19 could be excluded from the presence of genes of small effect on autism ($\lambda_s \leq 2$).

Most Likely Lod Scores

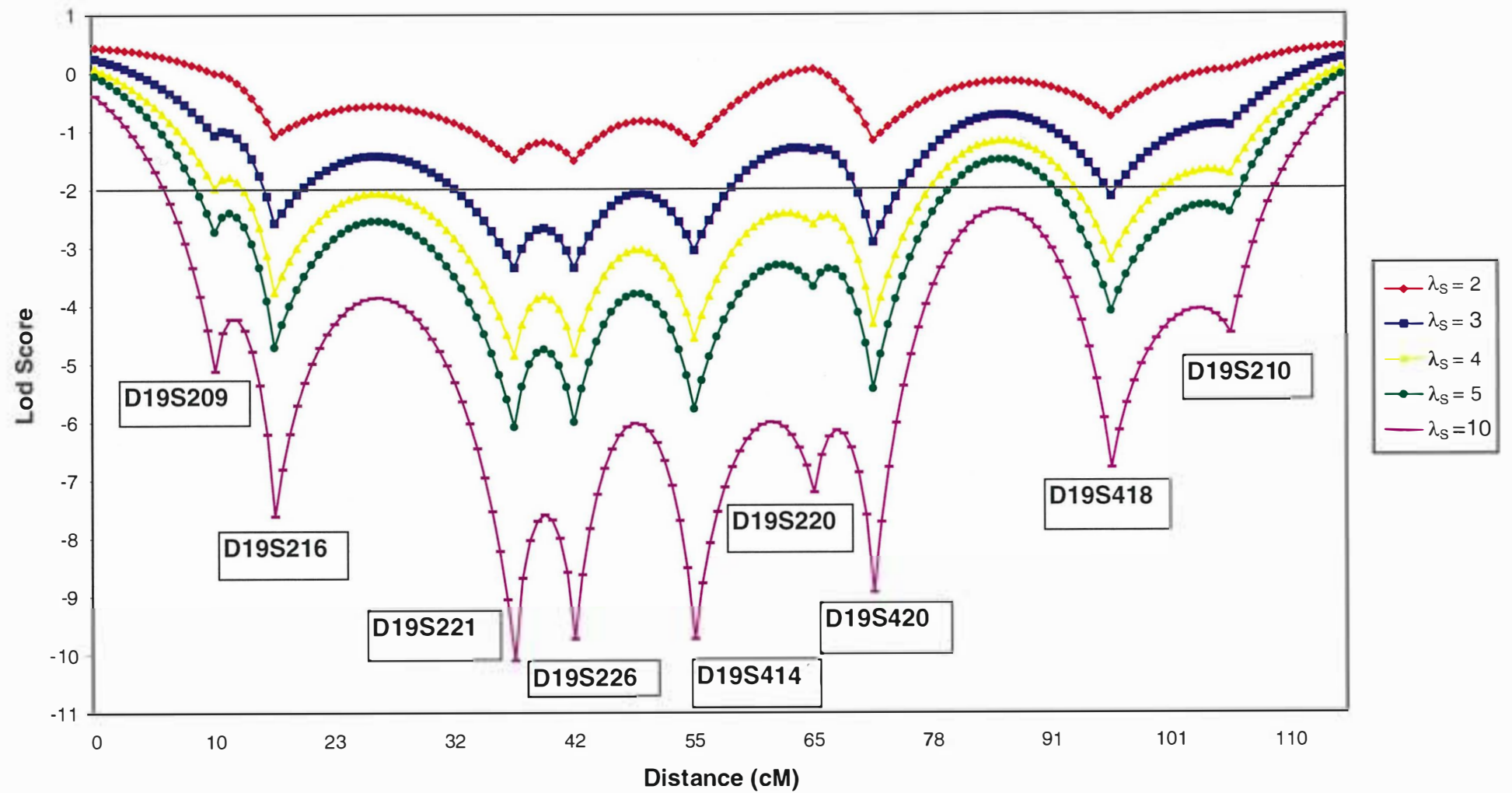
The paternal and maternal alleles that were shared between sib-pairs at all markers on chromosome 19 showed a significant difference only for the paternally inherited alleles at marker D19S210 when assessed by the χ^2 test (Table 8). The percent sharing at this marker was 65.9% ($\chi^2 = 4.45$, $p = 0.035$). When the percentage of alleles shared was calculated without taking in account parental origin of the alleles, the results did not deviate significantly from what was expected under the null hypothesis of no linkage (50%).

The percentage sharing and the lod score (MLS) obtained by the maximum likelihood calculation for each locus, also did not show any significant linkage on this chromosome. The highest lod score calculated was 0.73 at D19S220.

Table 8: Most Likely Lod Scores for Chromosome 19 (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D19S209	26	22	54.2%	0.33	31	24	56.4%	0.89	
D19S216	23	24	48.9%	0.02	33	21	61.1%	2.67	
D19S221	30	31	49.2%	0.02	26	29	47.3%	0.16	
D19S226	29	28	50.9%	0.02	32	30	51.6%	0.06	
D19S414	17	27	38.6%	2.27	26	22	54.2%	0.33	
D19S220	29	25	53.7%	0.30	35	23	60.3%	2.48	
D19S420	26	27	49.1%	0.02	32	28	53.3%	0.27	
D19S418	24	18	57.1%	0.86	22	24	47.8%	0.09	
D19S210	29	15	65.9%	4.45	27	27	50.0%	0.00	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D19S209	58	47	55.2%	1.15	56.9	0.19	0.49	0.32	0.50
D19S216	56	45	55.4%	1.20	52.7	0.22	0.5	0.28	0.08
D19S221	56	60	48.3%	0.14	51.5	0.24	0.5	0.26	0.03
D19S226	61	58	51.3%	0.08	51.6	0.23	0.5	0.27	0.03
D19S414	46	52	46.9%	0.37	53.4	0.22	0.5	0.29	0.13
D19S220	67	51	56.8%	2.17	57.8	0.18	0.49	0.33	0.73
D19S420	59	56	51.3%	0.08	53.4	0.22	0.5	0.28	0.13
D19S418	47	43	52.2%	0.18	54.2	0.21	0.5	0.29	0.17
D19S210	58	44	56.9%	1.92	57.1	0.18	0.49	0.33	0.49

Figure 13: Chromosome 19 multipoint results using USA families



3.1.2.2 Chromosome 20

Genes of Large Effect

Genes of large effect on chromosome 20 could be excluded from linkage to autism in over 86% of the chromosome (Figure 14). Two areas could not be excluded, namely an 8.8cM region between D20S117 and D20S95, as well as a 6.2cM region between D20S115 and D20S189.

Genes of Moderate Effect

A high percentage of chromosome 20 (69%) could also be excluded as containing a gene of moderate effect. The region not excluded spans all markers between D20S117 and D20S186.

Genes of Small Effect

Forty four percent of chromosome 20 could be excluded from the presence of genes of small effect on autism. This includes a 22.1cM region involving D20S118 and D20S195, a 22.3cM region between D20S178, D20S196 and D20S100, and a 16.1cM region surrounding D20S173 and D20S171.

Most Likely Lod Scores

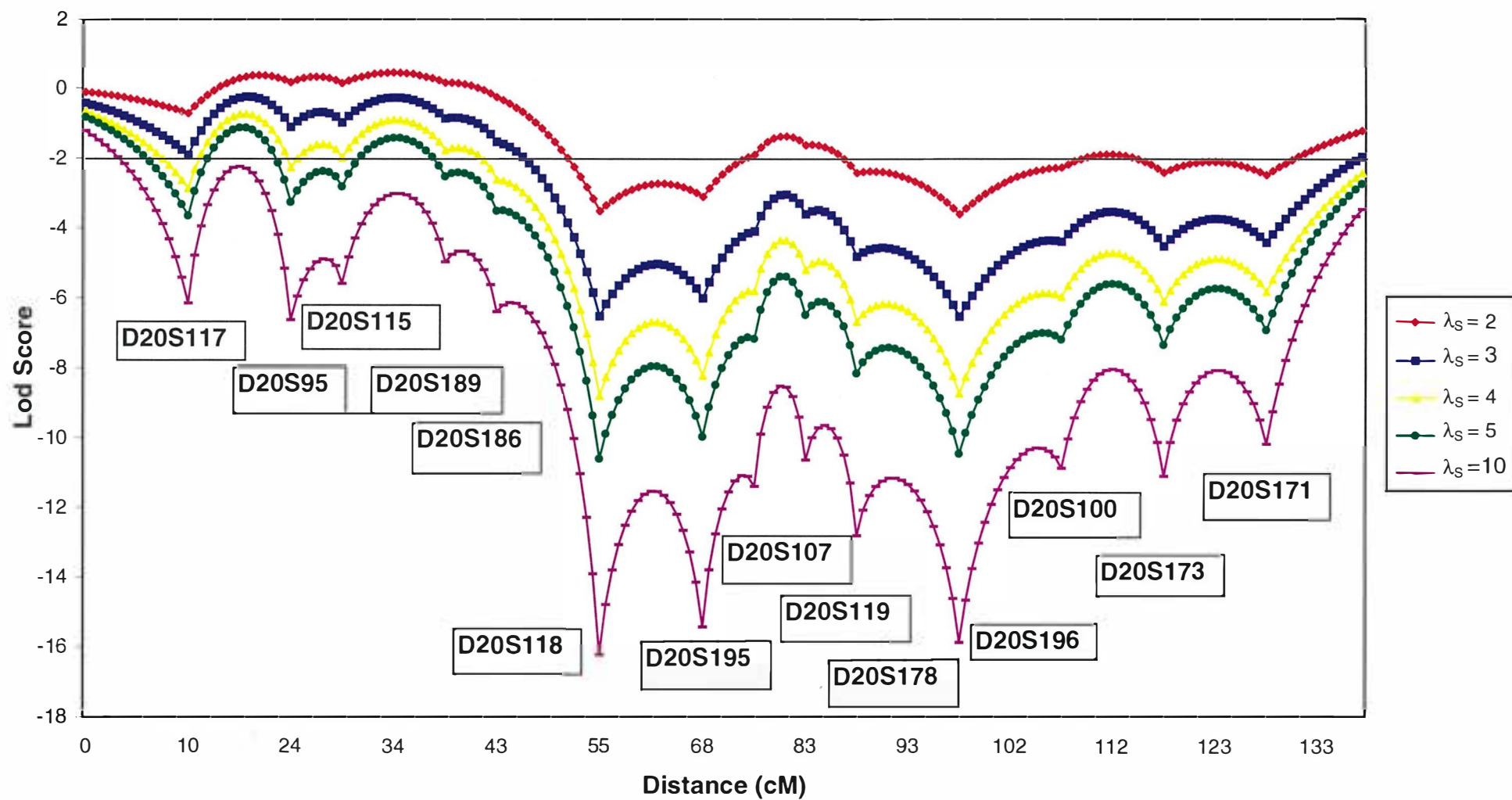
No significant deviation from the null hypothesis of 50% sharing was noted in the percentage of paternal alleles shared between sib-pairs at all markers on chromosome 20 (Table 9). Significant sharing, however, was noted for the maternally inherited alleles at both markers D20S189 ($\chi^2 = 4.57$; $p = 0.033$) and D20S100 ($\chi^2 = 3.92$; $p = 0.048$). The percentages of alleles shared at these markers, however, were not significant at the 5% level of significance in the combined data set.

Maximum likelihood calculations also showed no susceptibility region on this chromosome. Linkage results for 62% of the chromosome (between markers D20S118 to D20S171) did not deviate significantly from the distribution that would be expected by chance (50% sharing). The amount of sharing identical-by-descent for the remaining 38% of the chromosome was also non-significant, with the greatest amount of sharing calculated at D20S95 (57.3%), corresponding to a MLS of 0.67.

Table 9: Most Likely Lod Scores for Chromosome 20 (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D20S117	30	27	52.6%	0.16	32	32	50.0%	0.00	
D20S95	37	23	61.7%	3.27	34	32	51.5%	0.06	
D20S115	25	19	56.8%	0.82	24	22	52.2%	0.09	
D20S189	24	27	47.1%	0.18	36	20	64.3%	4.57	
D20S186	31	28	52.5%	0.15	38	28	57.6%	1.52	
D20S118	21	32	39.6%	2.28	26	29	47.3%	0.16	
D20S195	26	29	47.3%	0.16	28	31	47.5%	0.15	
D20S107	29	28	50.9%	0.02	32	24	57.1%	1.14	
D20S119	22	30	42.3%	1.23	36	22	62.1%	3.38	
D20S178	23	31	42.6%	1.19	26	25	51.0%	0.02	
D20S196	22	37	37.3%	3.81	36	32	52.9%	0.24	
D20S100	22	26	45.8%	0.33	32	18	64.0%	3.92	
D20S173	22	24	47.8%	0.09	22	21	51.2%	0.02	
D20S171	20	24	45.5%	0.36	25	27	48.1%	0.08	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D20S117	63	60	51.2%	0.07	52.8	0.22	0.50	0.28	0.09
D20S95	72	56	56.3%	2.00	57.3	0.18	0.49	0.33	0.67
D20S115	50	42	54.3%	0.70	57.2	0.18	0.49	0.33	0.60
D20S189	60	47	56.1%	1.58	56.4	0.19	0.49	0.32	0.50
D20S186	69	56	55.2%	1.35	55.4	0.2	0.49	0.31	0.37
D20S118	47	61	43.5%	1.81	50.0	0.25	0.50	0.25	0.00
D20S195	55	61	47.4%	0.31	50.0	0.25	0.50	0.25	0.00
D20S107	61	52	54.0%	0.72	51.8	0.23	0.50	0.27	0.04
D20S119	58	52	52.7%	0.33	51.6	0.23	0.50	0.27	0.03
D20S178	49	56	46.7%	0.47	50.0	0.25	0.50	0.25	0.00
D20S196	59	70	45.7%	0.94	50.0	0.25	0.50	0.25	0.00
D20S100	54	44	55.1%	1.02	50.0	0.25	0.50	0.25	0.00
D20S173	45	46	49.5%	0.01	50.0	0.25	0.50	0.25	0.00
D20S171	48	54	47.1%	0.35	50.0	0.25	0.50	0.25	0.00

Figure 14: Chromosome 20 multipoint results using USA families



3.1.2.3 Chromosome 21

Genes of Large Effect

Genes that have a large effect on autism could be excluded from the entire length of chromosome 21 using microsatellite markers spaced approximately 10cM apart (Figure 15).

Genes of Moderate Effect

A gene of moderate effect could also be excluded from 84% of chromosome 21. This includes the region between D21S1256 to D21S266, excluding the first 8cM between D21S1256 and D21S1253.

Genes of Small Effect

However, a gene a small effect on the disorder could only be excluded from 30% of chromosome 21. The exclusion region was 13cM in total and was defined by markers D21S1253 and D21S263.

Most Likely Lod Scores

The percentage sharing of alleles at all markers on chromosome 21 did not differ significantly from the null hypothesis (50% sharing), even when the paternal and maternal inherited alleles were viewed separately (Table 10). The only significant result was at locus D21S1253 which showed the percent of paternal alleles shared to be lower than what would be expected by chance (25% sharing, $\chi^2 = 9.00, p = 0.0027$).

The maximum likelihood calculations showed no significant linkage of chromosome 21 to an autism gene, with the MLS scores over the entire chromosome infrequently differing from zero. This is the value obtained when $\lambda_S = 1$, i.e., the increase in risk to siblings approaches the population prevalence of the disorder.

Figure 15: Chromosome 21 multipoint results using USA families

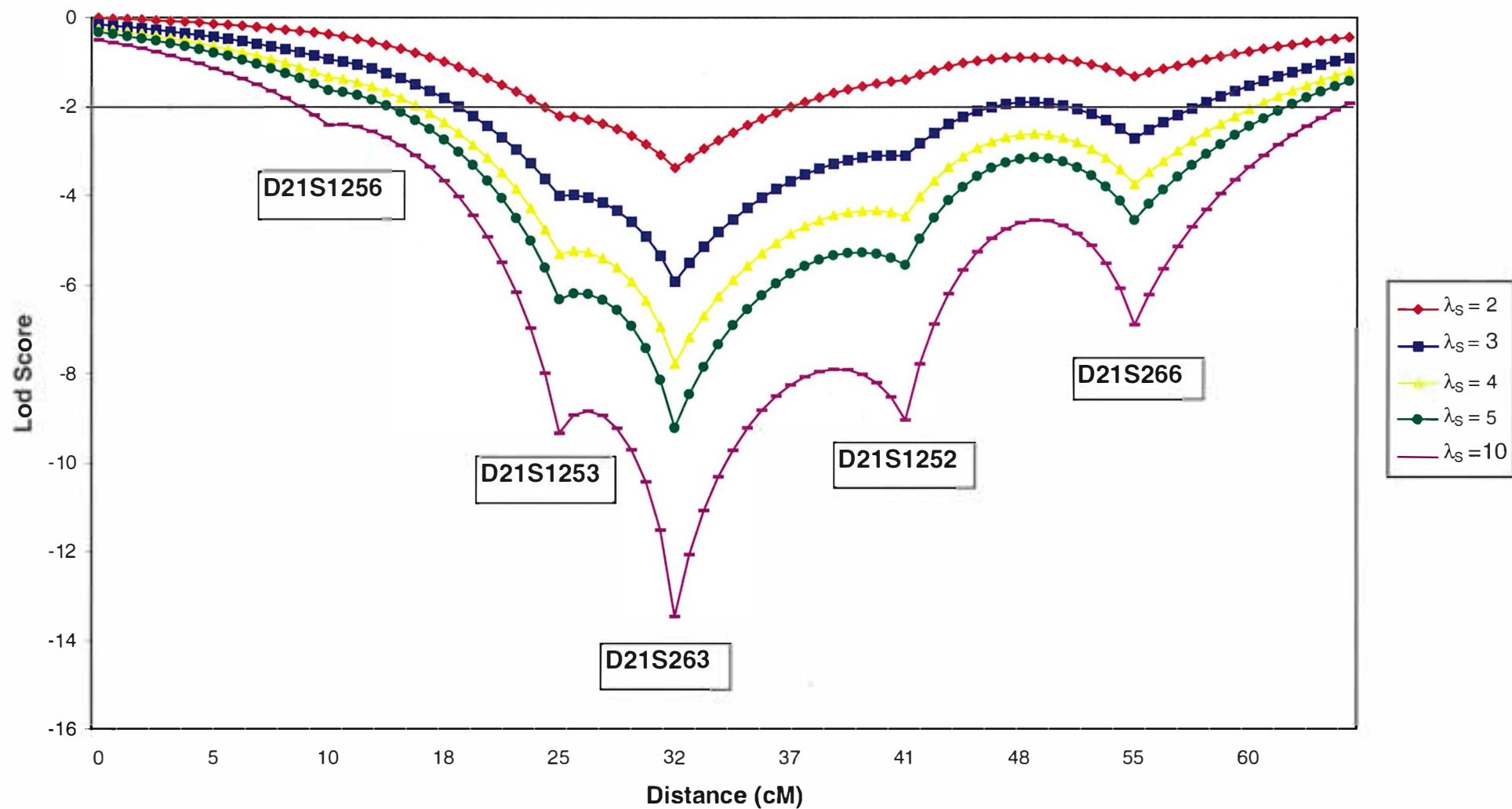


Table 10: Most Likely Lod Scores for Chromosome 21 (U.S.A. Families)

Locus	Paternal				Maternal			
	+	-	%	χ^2	+	-	%	χ^2
D21S1256	11	12	47.8%	0.04	15	7	68.2%	2.91
D21S1253	9	27	25.0%	9.00	27	15	64.3%	3.43
D21S263	15	23	39.5%	1.68	22	23	48.9%	0.02
D21S1252	22	30	42.3%	1.23	33	24	57.9%	1.42
D21S266	19	18	51.4%	0.03	22	17	56.4%	0.64

Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D21S1256	26	19	57.8%	1.09	53.4	0.22	0.5	0.29	0.06
D21S1253	36	42	46.2%	0.46	50.0	0.25	0.5	0.25	0.00
D21S263	38	47	44.7%	0.95	50.0	0.25	0.5	0.25	0.00
D21S1252	55	54	50.5%	0.01	51.8	0.23	0.5	0.27	0.03
D21S266	41	35	53.9%	0.47	51.3	0.24	0.5	0.26	0.01

3.1.2.4 Chromosome 22

Genes of Large Effect

Only seventeen percent of chromosome 22 could be excluded from containing a gene of large effect on autism. As illustrated in Figure 16, the exclusion region is located from approximately 2cM centromeric of D22S272 to D22S274.

Genes of Moderate Effect

Most of chromosome 22 (96%) could not be excluded from linkage to a gene of moderate effect on autism. The area that could be excluded was ~3cM telomeric of D22S423 to D22S274.

Genes of Small Effect

No region on chromosome 22 could be excluded from the presence of genes of small effect on autism.

Most Likely Lod Scores

The percentage of alleles shared by the affected sib-pairs at each locus on chromosome 22 is shown in Table 11. Whilst the percentage of shared alleles inherited from the mother did not deviate significantly from the expected value of 50%, the null hypothesis was rejected at a number of loci on chromosome 22 when the percentage of shared paternal inherited alleles was calculated. These loci include all markers between D22S315 and D22S272. The most significant result was at D22S278 with 82.1% of alleles shared IBD in the 90 U.S.A. sib-pairs analysed ($\chi^2 = 11.57, p = 0.00067$). When the results of the allele sharing from both parents were combined, significant deviation from the null hypothesis was still calculated at eight loci. The three most significant results were at D22S273 (65.9%, $\chi^2 = 8.58, p = 0.0034$), D22S280 (62.9%, $\chi^2 = 7.76, p = 0.0053$) and D22S278 (71.2%, $\chi^2 = 10.59, p = 0.00114$).

The maximum likelihood results for chromosome 22 are shown in Table 11. Ten markers had MLS scores greater than 1 on this chromosome, with D22S929 exceeding the genome wide significance level of 3 (3.08). The percentage of allele sharing at this marker was 65.4%. This marker was linked to the disorder Neurofibromatosis 2 in 1987 by Seizinger *et al.*

Two-Point Lod Scores

Two-point lod scores at this locus, however, showed a lower probability of linkage with a MLS of 1.52 calculated at D22S929 (Table 12). The highest two-point lod score was 2.46 at D22S278 (71.9% sharing). This marker is located 12.3cM telomeric from D22S929, the marker showing significant linkage to autism in the multipoint sib-pair analysis.

Figure 16: Chromosome 22 multipoint results using USA families

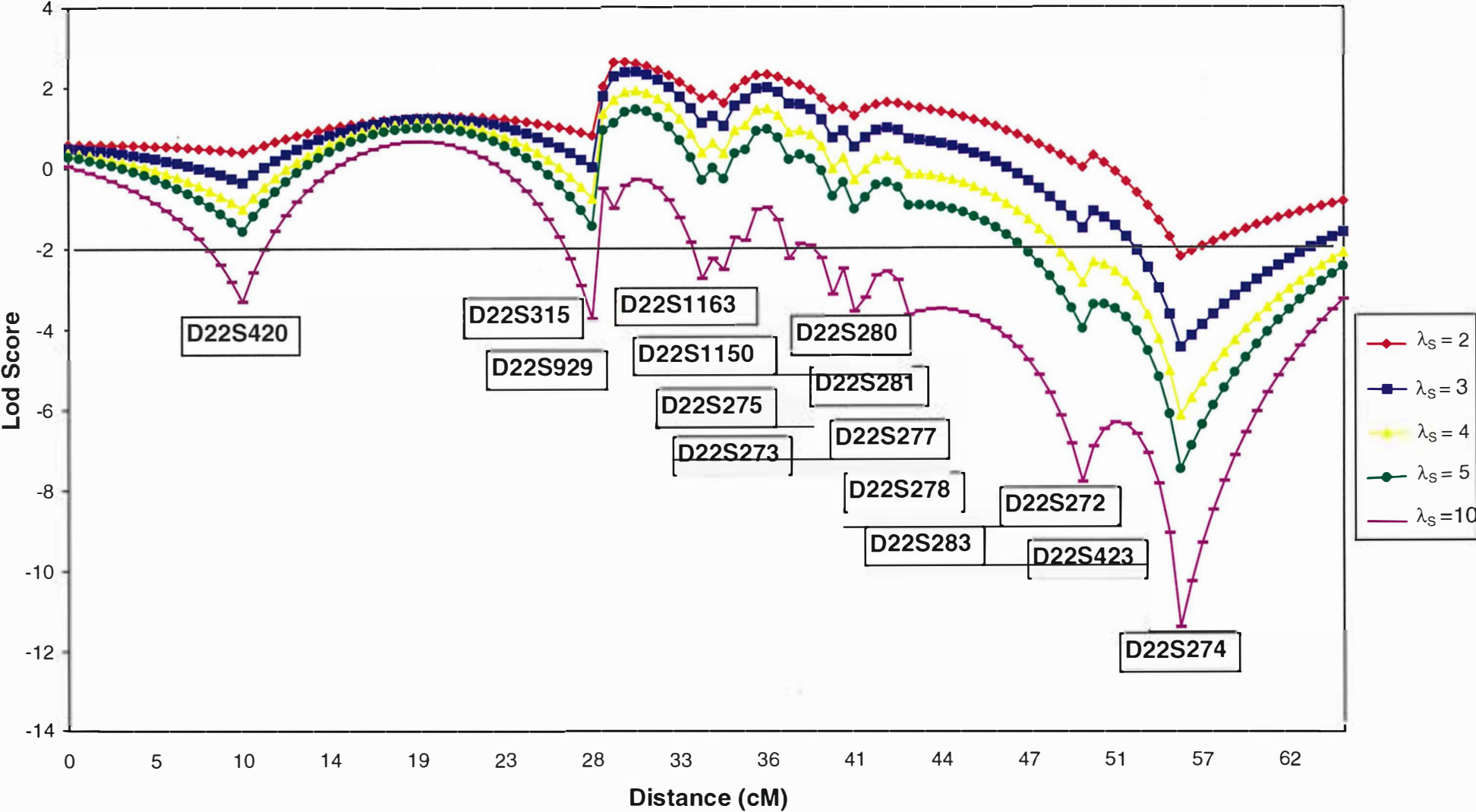


Table 11: Most Likely Lod Scores for Chromosome 22 (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D22S420	28	20	58.3%	1.33	26	22	54.2%	0.33	
D22S315	33	17	66.0%	5.12	26	32	44.8%	0.62	
D22S929	35	14	71.4%	9.00	27	22	55.1%	0.51	
D22S1163	39	17	69.6%	8.64	28	25	52.8%	0.17	
D22S1150	37	13	74.0%	11.52	22	21	51.2%	0.02	
D22S275	27	10	73.0%	7.81	21	21	50.0%	0.00	
D22S273	28	9	75.7%	9.76	25	17	59.5%	1.52	
D22S280	39	17	69.6%	8.64	33	25	56.9%	1.10	
D22S281	39	16	70.9%	9.62	27	31	46.6%	0.28	
D22S277	40	18	69.0%	8.34	33	30	52.4%	0.14	
D22S278	23	5	82.1%	11.57	18	11	62.1%	1.69	
D22S283	34	19	64.2%	4.25	35	27	56.5%	1.03	
D22S272	23	11	67.6%	4.24	25	26	49.0%	0.02	
D22S423	13	15	46.4%	0.14	21	20	51.2%	0.02	
D22S274	27	22	55.1%	0.51	21	27	43.8%	0.75	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D22S420	55	43	56.1%	1.47	58.2	0.17	0.49	0.34	0.65
D22S315	60	50	54.5%	0.91	58.1	0.18	0.49	0.34	0.86
D22S929	63	37	63.0%	6.76	65.4	0.12	0.45	0.43	3.08
D22S1163	69	44	61.1%	5.53	60.8	0.15	0.48	0.37	1.68
D22S1150	60	35	63.2%	6.58	60.7	0.15	0.48	0.37	1.65
D22S275	48	31	60.8%	3.66	60.1	0.16	0.48	0.36	1.47
D22S273	56	29	65.9%	8.58	62.5	0.14	0.47	0.39	2.19
D22S280	73	43	62.9%	7.76	62.5	0.14	0.47	0.39	2.21
D22S281	67	48	58.3%	3.14	60.2	0.16	0.48	0.36	1.45
D22S277	73	48	60.3%	5.17	59.9	0.16	0.48	0.36	1.37
D22S278	42	17	71.2%	10.59	60.8	0.15	0.48	0.37	1.57
D22S283	71	48	59.7%	4.45	61.6	0.15	0.47	0.38	1.81
D22S272	49	38	56.3%	1.39	57.6	0.18	0.49	0.33	0.74
D22S423	34	35	49.3%	0.01	58.8	0.17	0.48	0.35	0.97
D22S274	50	51	49.5%	0.01	50.5	0.24	0.5	0.26	0.00

Table 12: Two-Point Lod Scores for Chromosome 22 (U.S.A. Families)

Locus	Paternal				Maternal						
	+	-	%	χ^2	+	-	%	χ^2			
D22S420	28	20	58%	1.33	26	22	54.20%	0.33			
D22S315	33	17	66%	5.12	26	32	44.80%	0.62			
D22S929	35	14	71%	9.00	27	22	55.10%	0.51			
D22S1163	39	17	70%	8.64	28	25	52.80%	0.17			
D22S1150	37	13	74%	11.52	22	21	51.20%	0.02			
D22S275	27	10	73%	7.81	21	21	50.00%	0.00			
D22S273	28	9	76%	9.76	25	17	59.50%	1.52			
D22S280	39	17	70%	8.64	33	25	56.90%	1.10			
D22S281	39	16	71%	9.62	27	31	46.60%	0.28			
D22S277	40	18	70%	8.34	33	30	52.40%	0.14			
D22S278	23	5	82%	11.57	18	11	62.10%	1.69			
D22S283	34	19	64%	4.25	35	27	56.50%	1.03			
D22S272	23	11	68%	4.24	25	26	49.00%	0.02			
D22S423	13	15	46%	0.14	21	20	51.20%	0.02			
D22S274	27	22	55%	0.51	21	27	43.80%	0.75			
Locus	Combined				Most –Likely Model						
	+	-	%	χ^2	%shr	z0	z1	z2	Pat	Mat.	MLS
D22S420	55	43	56%	1.47	56.2	0.19	0.49	0.32	0.3	0.0	0.33
D22S315	60	50	54%	0.91	54.6	0.21	0.50	0.30	0.6	-0.4	0.20
D22S929	63	37	63%	6.76	63.3	0.13	0.46	0.40	1.7	-0.2	1.52
D22S1163	69	44	61%	5.53	61.5	0.15	0.47	0.38	1.6	-0.3	1.26
D22S1150	60	35	63%	6.58	63.4	0.13	0.46	0.40	2.1	-0.6	1.48
D22S275	48	31	61%	3.66	60.8	0.15	0.48	0.37	1.2	-0.4	0.80
D22S273	56	29	66%	8.58	67.1	0.11	0.44	0.45	1.9	0.1	2.04
D22S280	73	43	63%	7.76	63.2	0.14	0.47	0.40	1.7	0.0	1.73
D22S281	67	48	58%	3.14	58.4	0.17	0.49	0.34	1.4	-0.7	0.70
D22S277	73	48	60%	5.17	60.3	0.16	0.48	0.36	1.5	-0.3	1.13
D22S278	42	17	71%	10.59	71.9	0.08	0.40	0.52	2.4	0.1	2.46
D22S283	71	48	60%	4.45	60.0	0.16	0.48	0.36	0.9	0.2	1.01
D22S272	49	38	56%	1.39	56.5	0.19	0.49	0.32	0.6	-0.2	0.31
D22S423	34	35	49%	0.01	50	0.25	0.5	0.25	0.0	0.0	0.00
D22S274	50	51	50%	0.01	50	0.25	0.5	0.25	0.0	0.0	0.00

3.1.3 Candidate Regions

The 90 U.S.A. multiplex families were also used to conduct a refined analysis of genomic regions that have produced positive results in previous studies carried out by other groups. These regions included the HLA region on chromosome 6 and a large section of chromosome 15. In this analysis the entire length of chromosome 15 was studied.

3.1.3.1 HLA Region on Chromosome 6

In total, 8 markers with an average inter-marker distance of 3cM were genotyped in the HLA region on chromosome 6. Multipoint sib-pair analysis results under different gene effects (λ_S) are presented in Figure 17. The entire HLA region on chromosome 6 could be excluded from linkage to a gene of moderate effect for autism ($\lambda_S \geq 2$). All markers showed a distribution of sharing that would be expected under a model of no linkage (ie. $z_0=0.25$, $z_1=0.50$, $z_2=0.25$) (Table 13). Therefore the amount of sharing of alleles at each locus did not exceed 50%, corresponding to a most-likely lod score of zero at each marker. When viewed separately by transmitting parent, the shared alleles at each locus also did not exceed what was expected by chance, with the majority of markers showing a percentage sharing lower than 50%.

3.1.3.2 Chromosome 15

Nineteen microsatellites that were approximately 7.5cM apart from each other and spanned the entire length of chromosome 15 were analysed in the 97 U.S.A. affected sib-pairs. The results for this chromosome were very similar to those

displayed by the HLA region on chromosome 6. Linkage to a susceptibility gene that has a large effect on autism could be excluded throughout the entire chromosome ($\lambda_s \geq 5$) (Figure 18). ASPEX results also showed that linkage to a gene of moderate effect could be excluded for 81% of the chromosome between the region spanning D15S165 and D15S120. However, a gene of small effect was only excluded from 10% of this chromosome near markers D15S117, D15S152 and D15S205. When viewing the maximum likelihood calculations for this chromosome (Table 14), it can be seen that allele sharing greater than 55% was not observed for any of the chromosome 15 markers. The highest lod score within this region was 0.22 at D15S158. Allele sharing by parental origin also showed no significant deviation from the null hypothesis of no linkage.

Table 13: Most Likely Lod Scores for Chromosome 6 (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D6S285	27	26	50.9%	0.02	28	28	50.0%	0.00	
D6S1029	12	17	41.4%	0.86	16	23	41.0%	1.26	
D6S276	27	36	42.9%	1.29	32	35	47.8%	0.13	
D6S1014	29	30	49.2%	0.02	21	27	43.8%	0.75	
D6S497	18	19	48.6%	0.03	20	32	38.5%	2.77	
D6S439	20	23	46.5%	0.21	17	24	41.5%	1.20	
D6S1019	20	23	46.5%	0.21	21	27	43.8%	0.75	
D6S426	30	39	43.5%	1.17	33	37	47.1%	0.23	
Locus	Combined				Most-Likely Model				
	+	-	%	χ^2	%shr	z0	Z1	z2	MLS
D6S285	56	55	50.5%	0.01	50	0.25	0.5	0.25	0.00
D6S1029	33	45	42.3%	1.85	50	0.25	0.5	0.25	0.00
D6S276	60	72	45.5%	1.09	50	0.25	0.5	0.25	0.00
D6S1014	52	59	46.8%	0.44	50	0.25	0.5	0.25	0.00
D6S497	39	52	42.9%	1.86	50	0.25	0.5	0.25	0.00
D6S439	37	47	44.0%	1.19	50	0.25	0.5	0.25	0.00
D6S1019	44	53	45.4%	0.84	50	0.25	0.5	0.25	0.00
D6S426	64	77	45.4%	1.20	50	0.25	0.5	0.25	0.00

Figure 17: Chromosome 6 multipoint results using USA families

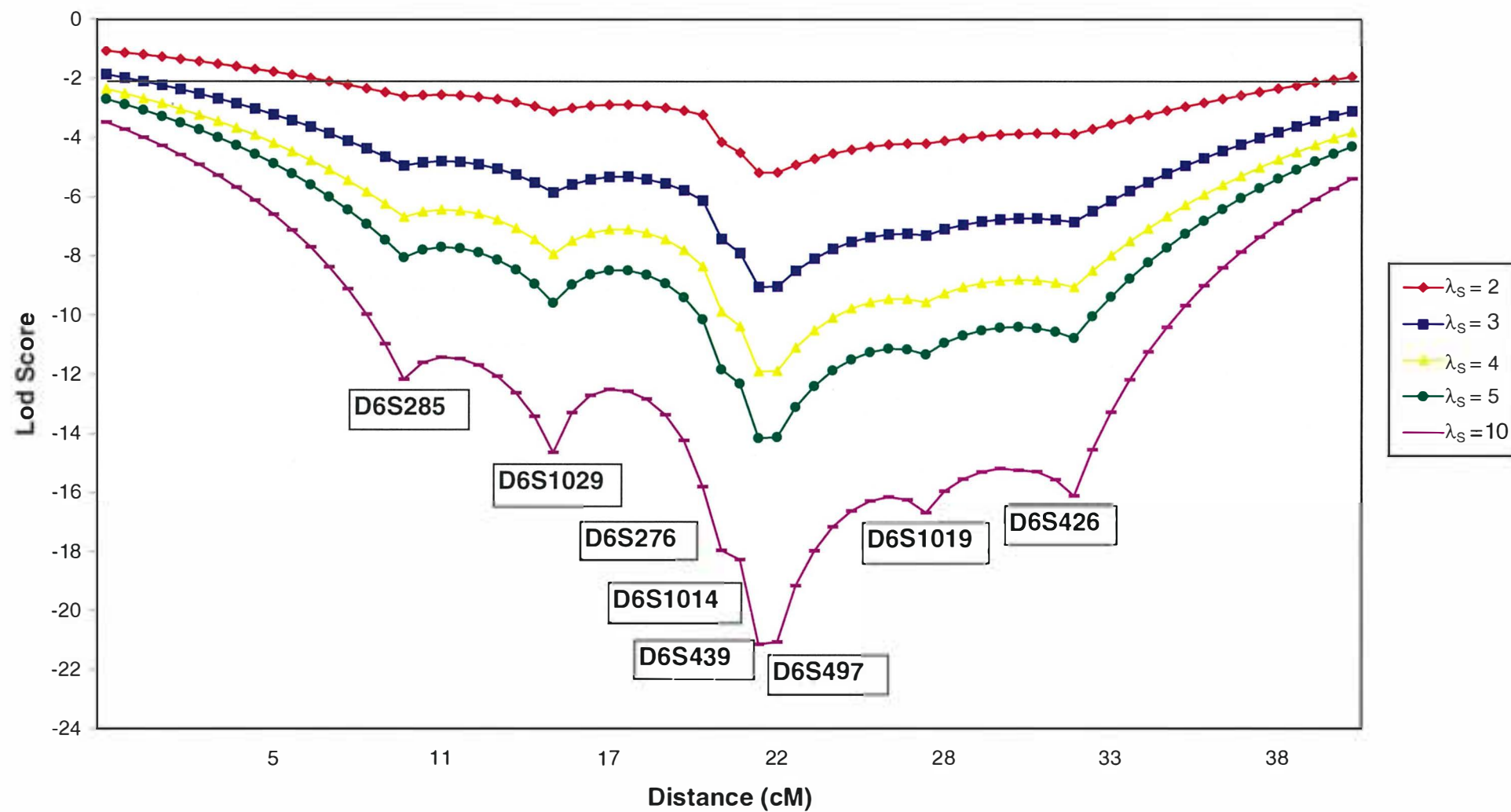


Figure 18: Chromosome 15 multipoint results using USA families

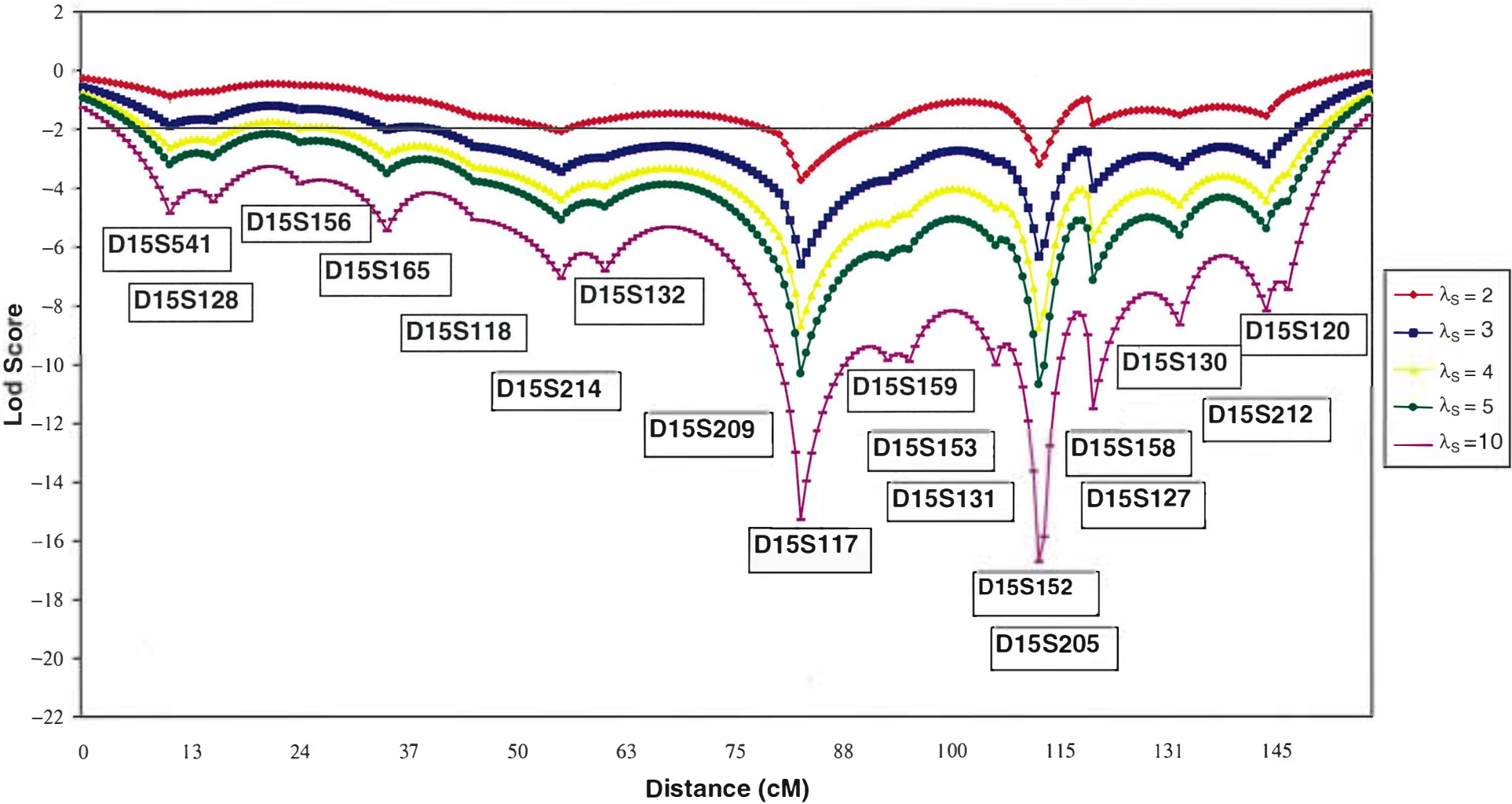


Table 14: Most Likely Lod Scores for Chromosome 15 (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D15S541	11	9	55.0%	0.20	7	13	35.0%	1.80	
D15S128	22	15	59.5%	1.32	23	19	54.8%	0.38	
D15S156	8	4	66.7%	1.33	8	7	53.3%	0.07	
D15S165	14	13	51.9%	0.04	22	17	56.4%	0.64	
D15S118	5	10	33.3%	1.67	8	5	61.5%	0.69	
D15S214	2	4	33.3%	0.67	4	13	23.5%	4.76	
D15S132	11	8	57.9%	0.47	7	15	31.8%	2.91	
D15S209	10	11	47.6%	0.05	15	11	57.7%	0.62	
D15S117	26	32	44.8%	0.62	24	38	38.7%	3.16	
D15S159	6	6	50.0%	0.00	8	6	57.1%	0.29	
D15S153	28	33	45.9%	0.41	32	28	53.3%	0.27	
D15S131	20	21	48.8%	0.02	22	20	52.4%	0.10	
D15S152	7	8	46.7%	0.07	4	12	25.0%	4.00	
D15S205	36	31	53.7%	0.37	38	39	49.4%	0.01	
D15S158	9	7	56.3%	0.25	7	6	53.8%	0.08	
D15S127	35	34	50.7%	0.01	34	31	52.3%	0.14	
D15S130	18	19	48.6%	0.03	25	21	54.3%	0.35	
D15S212	8	7	53.3%	0.07	12	11	52.2%	0.04	
D15S120	22	20	52.4%	0.10	26	23	53.1%	0.18	
Locus	Combined				Most-Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D15S541	20	24	45.5%	0.36	51.9	0.23	0.5	0.27	0.03
D15S128	46	35	56.8%	1.49	53.0	0.22	0.5	0.28	0.08
D15S156	16	11	59.3%	0.93	54.0	0.21	0.5	0.29	0.10
D15S165	36	30	54.5%	0.55	52.1	0.23	0.5	0.27	0.03
D15S118	13	15	46.4%	0.14	50.0	0.25	0.5	0.25	0.00
D15S214	6	17	26.1%	5.26	50.0	0.25	0.5	0.25	0.00
D15S132	18	23	43.9%	0.61	50.0	0.25	0.5	0.25	0.00
D15S209	25	22	53.2%	0.19	50.0	0.25	0.5	0.25	0.00
D15S117	50	70	41.7%	3.33	50.0	0.25	0.5	0.25	0.00
D15S159	16	14	53.3%	0.13	50.6	0.24	0.5	0.26	0.00
D15S153	61	62	49.6%	0.01	52.4	0.23	0.5	0.27	0.07
D15S131	42	41	50.6%	0.01	53.9	0.21	0.5	0.29	0.18
D15S152	11	20	35.5%	2.61	50.0	0.25	0.5	0.25	0.00
D15S205	74	70	51.4%	0.11	50.3	0.25	0.5	0.25	0.00
D15S158	16	13	55.2%	0.31	54.2	0.21	0.5	0.29	0.22
D15S127	70	66	51.5%	0.12	51.5	0.23	0.5	0.27	0.03
D15S130	43	40	51.8%	0.11	51.8	0.23	0.5	0.27	0.03
D15S212	20	18	52.6%	0.11	50.8	0.24	0.5	0.26	0.01
D15S120	49	44	52.7%	0.27	54.6	0.21	0.5	0.3	0.21

3.1.4 Duplication of Positive Regions Detected by the Stanford and Perth Research Groups

At the completion of stage one of the project, 346 autosomal and 14 X-linked markers had been analysed in all 97 affected sib-pairs from the U.S.A. The results from the multipoint analysis on this data identified eleven regions that gave a maximum lod score greater than or equal to one ($MLS \geq 1$). These included chromosome 1p ($MLS = 1.87$), 1q ($MLS = 1.19$), 7p ($MLS = 1.00$), 10q ($MLS = 1.00$), 11p ($MLS = 1.25$), 13q ($MLS = 1.49$), 15q ($MLS = 1.75$), 17p ($MLS = 1.30$), 18q ($MLS = 1.00$), 20p ($MLS = 1.09$) and 22q ($MLS = 3.08$). MLS curves for each of these chromosomes were generated from maximum likelihood calculations and are shown in Risch *et al.* (1999). Results for chromosome 22 are also presented in section 3.1.2.4.

The analysis of six of these regions (1p, 1q, 10q, 15q, 17p & 18q) was duplicated in this study using the same families as an additional control. The genetic maps utilised, however, contained different microsatellites to those analysed in stage one of the project (refer to Appendix 6.1 for the microsatellites analysed in each chromosomal region and the distances between each marker). The results of these analyses are stated below.

3.1.4.1 Chromosome 1p

A region on the short arm of chromosome 1 showed a $MLS > 1$ when analysed by the Stanford group ($MLS = 1.87$). When three markers within this region were re-analysed by the Perth group, positive linkage to this region was still found and no area could be excluded for any gene effect (λ_s) (Figure 19). Based on

the calculated sharing probabilities in the interval between D1S1631 to D1S534 ($z_0 = 0.13$, $z_1 = 0.46$, $z_2 = 0.41$), this region had a MLS of 1.55 (Table 15). The MLS at D1S1631 was 1.26 (60.9% sharing). Although slightly lower than that calculated by the Stanford group, this linkage result ($MLS > 1$) still suggests that a susceptibility locus to autism may reside in this region. We performed two-point analysis as an additional test for linkage at this locus (Table 16). Results again were suggestive of an autism gene near D1S1631 ($MLS = 1.13$, 60.1% sharing). The percentage of allele sharing at this locus in the combined data set was also significant at the designated 5% level of significance ($\chi^2 = 5.14$, $p = 0.0234$).

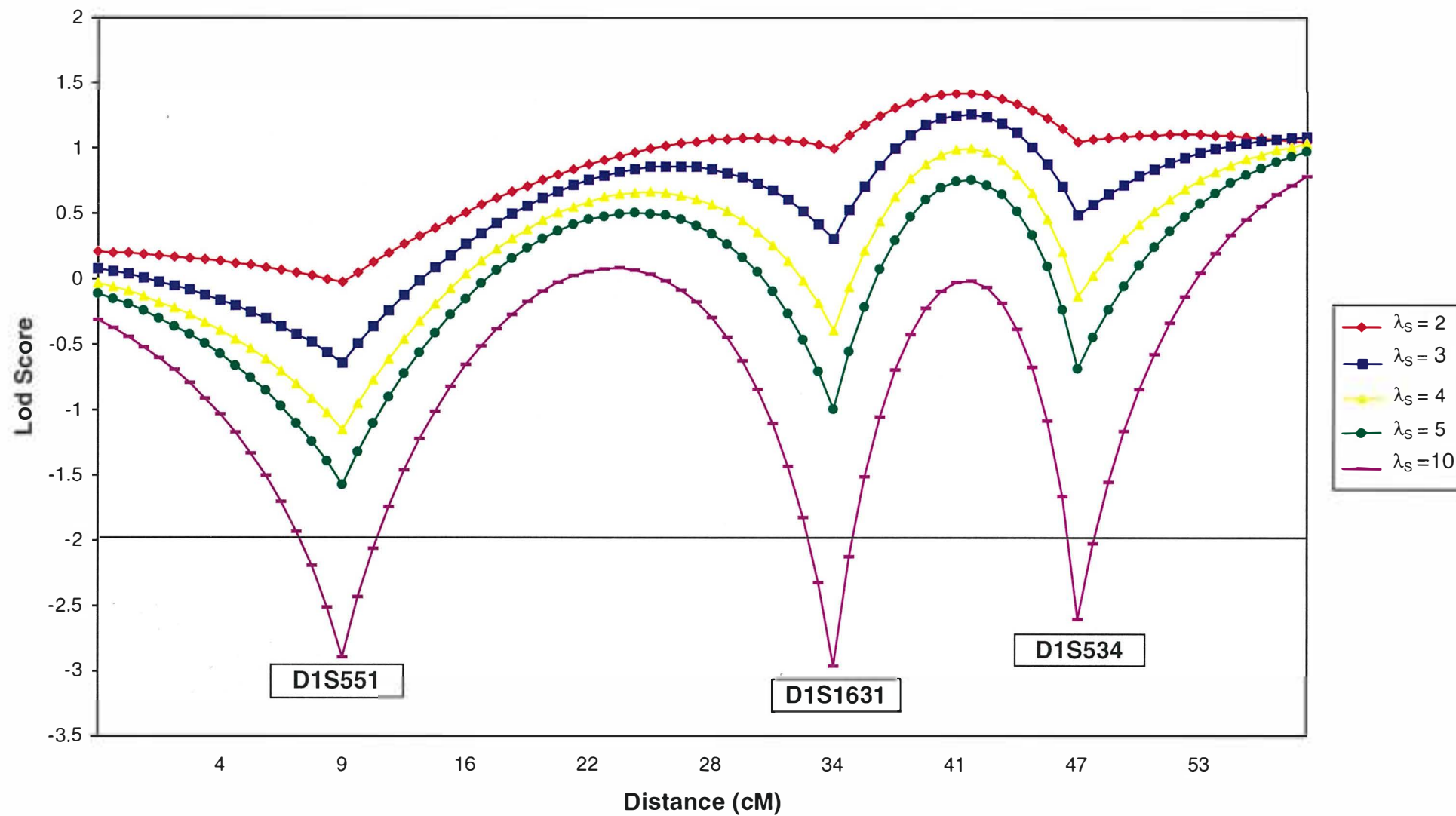
Table 15: Most Likely Lod Scores for Chromosome 1p (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%		χ^2
D1S551	22	13	62.9%	2.31	17	20	45.9%		0.24
D1S1631	32	17	65.3%	4.59	36	27	57.1%		1.29
D1S534	26	19	57.8%	1.09	34	22	60.7%		2.57
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D1S551	40	34	54.1%	0.49	56.1	0.19	0.49	0.31	0.25
D1S1631	68	44	60.7%	5.14	60.9	0.15	0.48	0.37	1.26
D1S534	61	42	59.2%	3.50	60.7	0.15	0.48	0.37	1.14

Table 16: Two-Point Lod Scores for Chromosome 1p (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D1S551	22	13	62.9%	2.31	17	20	45.9%	0.24	
D1S1631	32	17	65.3%	4.59	36	27	57.1%	1.29	
D1S534	26	19	57.8%	1.09	34	22	60.7%	2.57	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D1S551	40	34	54.1%	0.49	54.2	0.21	0.50	0.29	0.11
D1S1631	68	44	60.7%	5.14	60.7	0.15	0.48	0.37	1.13
D1S534	61	42	59.2%	3.50	59.4	0.16	0.48	0.35	0.78

Figure 19: Chromosome 1p multipoint results using USA families



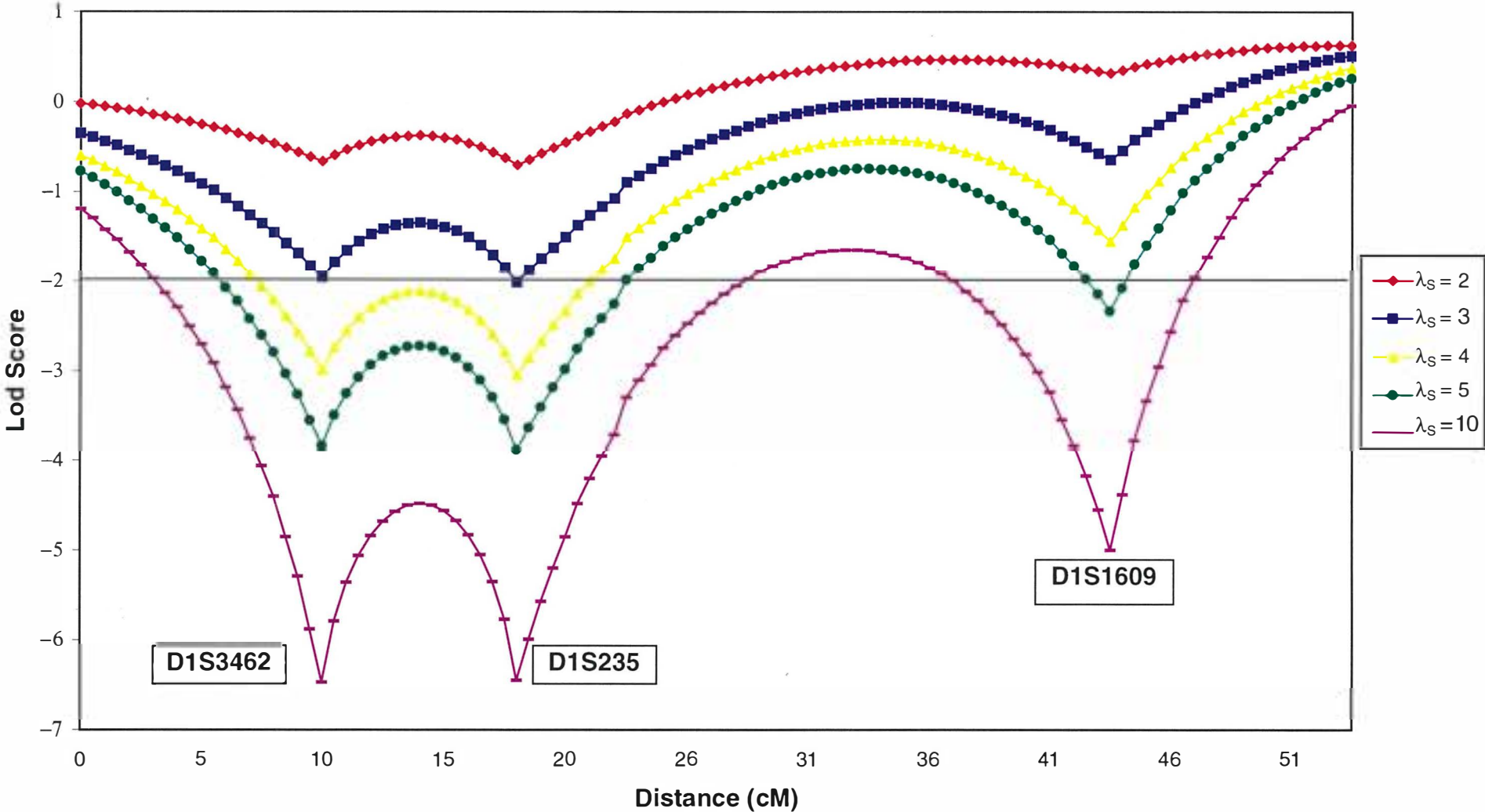
3.1.4.2 Chromosome 1q

A region on chromosome 1q suggesting linkage to an autism gene in stage one of the genome scan (MLS = 1.19), was re-analysed in the same affected sib-pairs using microsatellite markers mapping in the same area. Whilst no region in this section of chromosome 1 could be excluded from linkage to autism, the multipoint lod scores calculated at all λ_s values were less than one (Figure 20). Maximum likelihood calculations were also lower than those calculated in stage one of the study (Table 17). A MLS of 0.71 (58.8% sharing) was calculated at D1S1609. The percentage of alleles shared by sib-pairs at each locus was also not significantly different to that expected under the null hypothesis of no linkage (50%). Our findings make it unlikely that a gene for autism resides in this region of the genome.

Table 17: Most Likely Lod Scores for Chromosome 1q (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D1S3462	18	22	45.0%	0.40	26	19	57.8%	1.09	
D1S235	26	18	59.1%	1.45	24	23	51.1%	0.02	
D1S1609	36	22	62.1%	3.38	24	20	54.5%	0.36	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D1S3462	46	43	51.7%	0.10	54.8	0.2	0.50	0.30	0.21
D1S235	53	44	54.6%	0.84	54.6	0.21	0.50	0.30	0.20
D1S1609	60	42	58.8%	3.18	58.8	0.17	0.48	0.35	0.71

Figure 20: Chromosome 1q multipoint results using USA families



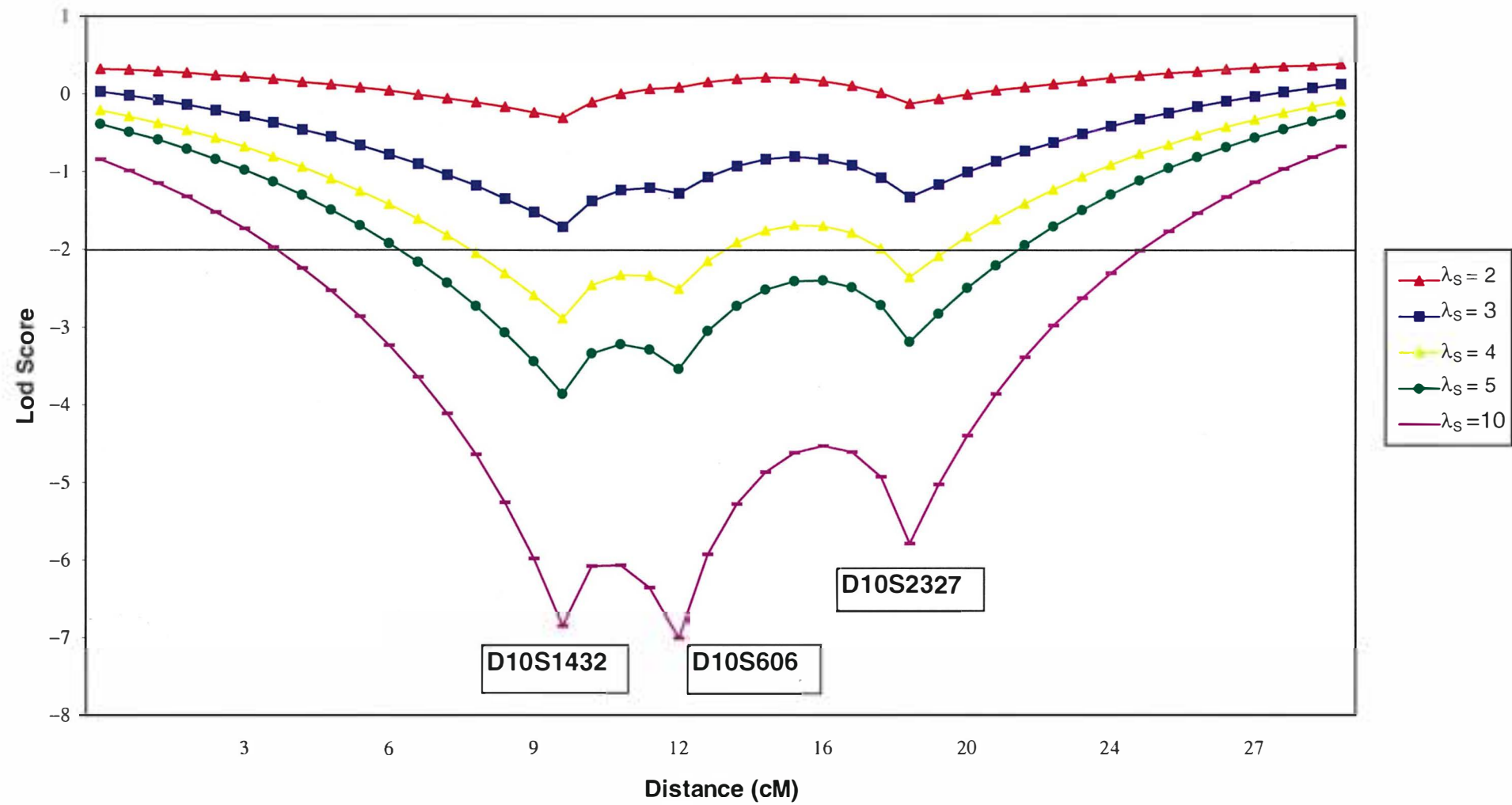
3.1.4.3 Chromosome 10q

Re-analysis of a 10cM region of chromosome 10q excluded this area from containing a gene conferring susceptibility to autism. This region showed a MLS = 1.00 when analysed by the Stanford research group. Through the analysis of three markers approximately 5cM apart, chromosome 10q was excluded from linkage to a gene of large effect on the disorder (Figure 21). A gene of small to moderate effect, however, could not be excluded over this entire region of chromosome 10q. Multipoint analysis using the “most-likely” model showed no significant excess of allele sharing in the combined data set with the calculation of a MLS = 0.76 (58.3% sharing) at D10S606 (Table 18). Significant excess sharing, however, was noted for the maternally inherited alleles at locus D10S2327 ($\chi^2 = 4.57, p = 0.033$).

Table 18: Most Likely Lod Scores for Chromosome 10q (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D10S1432	25	14	64.1%	3.10	24	24	50.0%	0.00	
D10S606	30	18	62.5%	3.00	20	21	48.8%	0.02	
D10S2327	17	20	45.9%	0.24	25	12	67.6%	4.57	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D10S1432	49	38	-56.3%	1.39	56.8	0.19	0.49	0.32	0.49
D10S606	50	39	-56.2%	1.36	58.3	0.17	0.49	0.34	0.76
D10S2327	42	32	-56.8%	1.35	57.6	0.18	0.49	0.33	0.54

Figure 21: Chromosome 10q multipoint results using USA families



3.1.4.4 Chromosome 15q

Results similar to those obtained for chromosome 10q (3.1.4.1) were also found in the region of chromosome 15q, identified by the Stanford group as possibly containing a gene for autism (MLS = 1.75). The results from this analysis are shown in Figure 18 and Table 14 and are explained previously in section 3.1.3.2. The area previously identified in stage one of the study as possibly containing a susceptibility gene to autism was between markers D15S159 and D15S130. Our results (section 3.1.3.2) clearly show the exclusion of a gene of moderate to large effect in this region and the results of the maximum likelihood calculations did not differ significantly from those expected by chance.

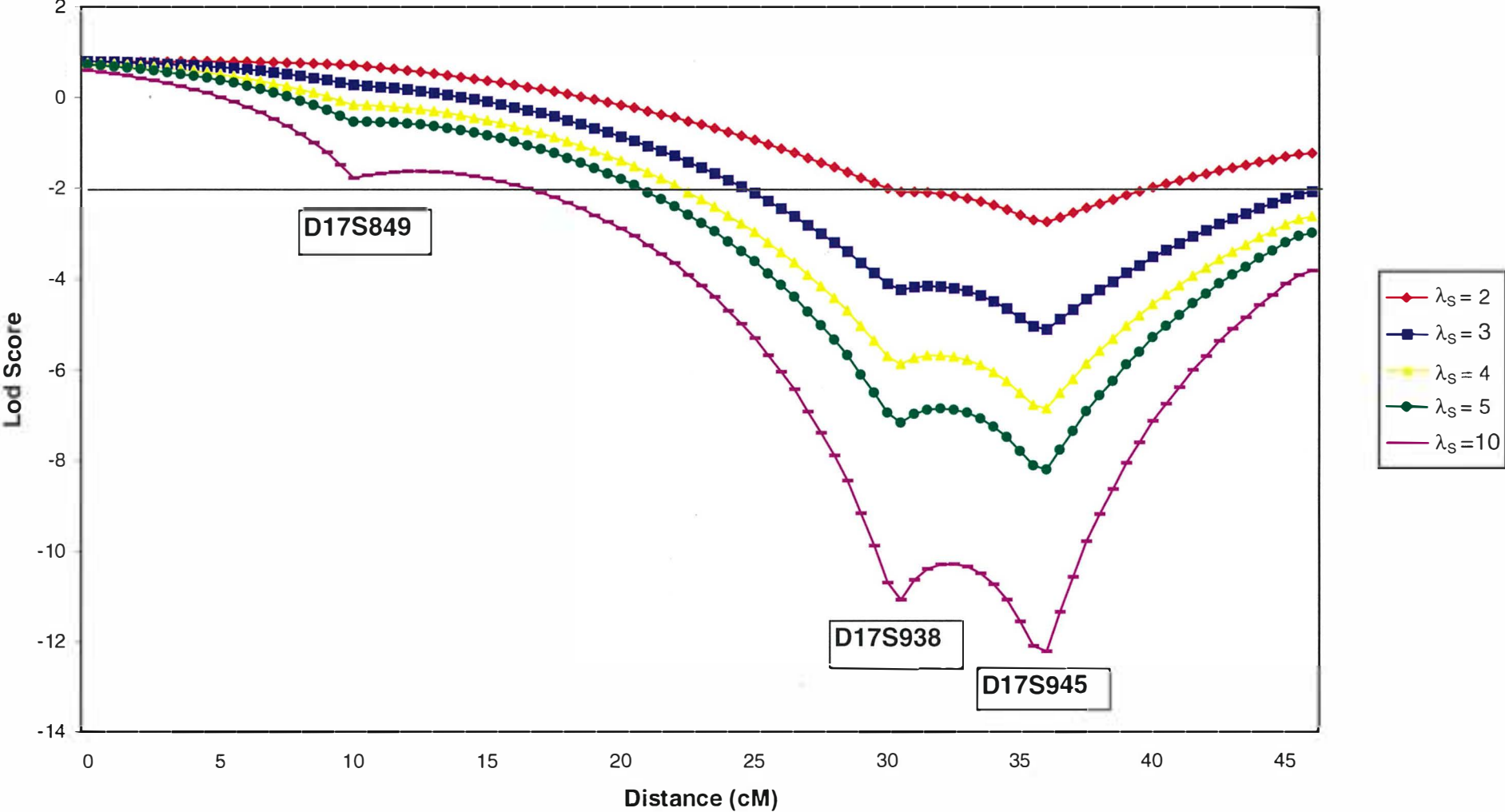
3.1.4.5 Chromosome 17p

When chromosome 17p was re-examined in the 90 U.S.A. multiplex families using three different microsatellites, suggestive linkage (MLS > 1) to an autism gene was excluded at all loci. This region previously showed a MLS = 1.30 when analysed by the Stanford research group. Two of these markers (D17S938 and D17S945) could be excluded from linkage to a gene of small effect on the disorder (Figure 22). The marker, which could not be excluded as linked to a gene of small effect, was D17S849. This marker also showed significant excess of sharing in the paternal and combined data sets ($\chi^2 = 5.16$ & 4.05 respectively; Table 19). The MLS that was calculated at this marker on chromosome 17p was 0.81 (60.4% sharing) (Table 19), which was not significant at the genome-wide level of significance.

Table 19: Most Likely Lod Scores for Chromosome 17p (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D17S849	26	12	68.4%	5.16	22	18	55.0%	0.40	
D17S938	21	23	47.7%	0.09	29	27	51.8%	0.07	
D17S945	23	32	41.8%	1.47	28	27	50.9%	0.02	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D17S849	49	31	61.3%	4.05	60.4	0.16	0.48	0.36	0.81
D17S938	51	51	50.0%	0.00	50.6	0.24	0.50	0.26	0.00
D17S945	52	60	46.4%	0.57	50.0	0.25	0.50	0.25	0.00

Figure 22: Chromosome 17p multipoint results using USA families



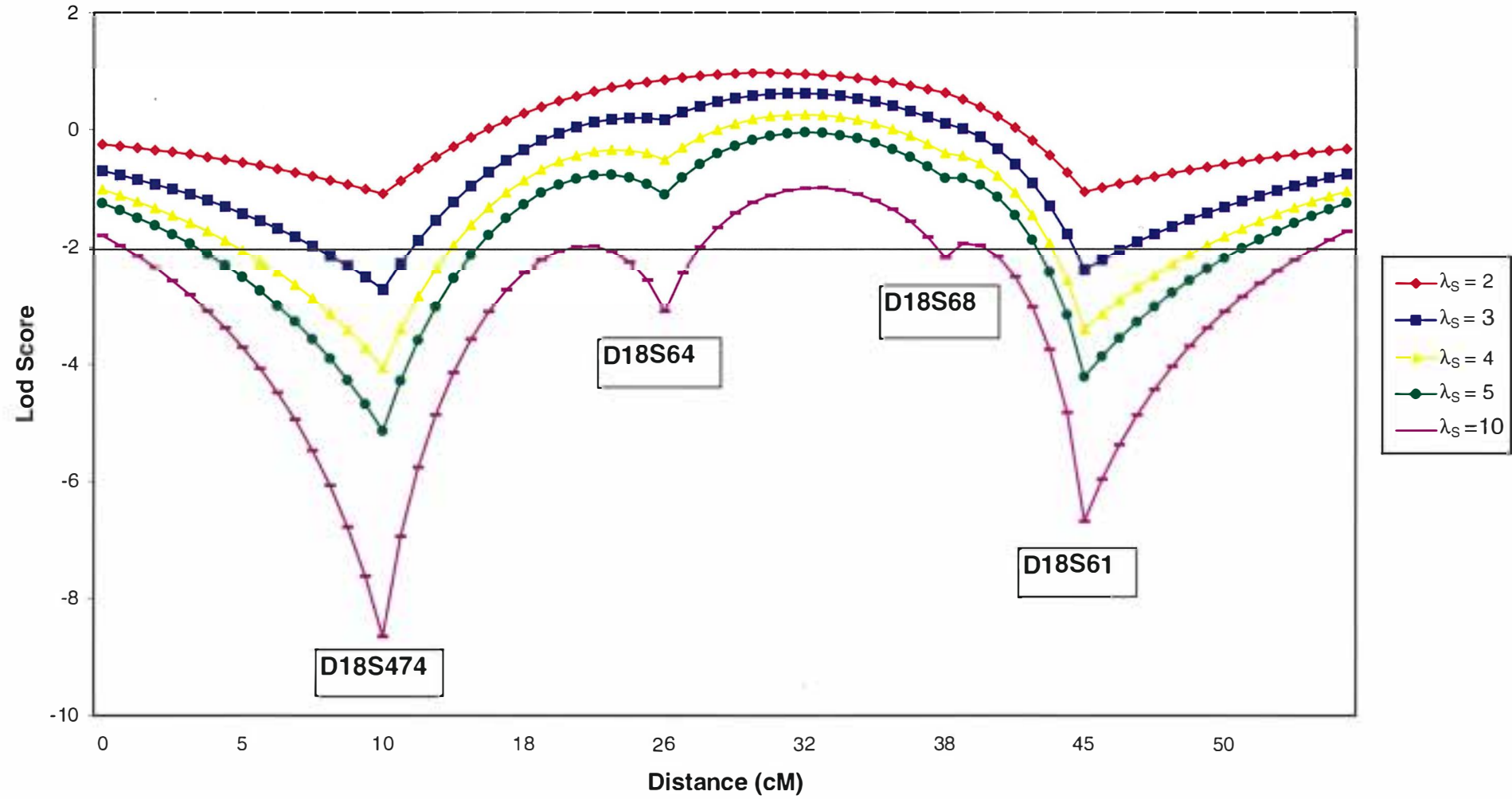
3.1.4.6 Chromosome 18q

Four markers spanning the region of chromosome 18 that showed a ML S of 1.00 for the Stanford group were tested using the same families. Both D18S474 and D18S61 could be excluded from linkage to an autism gene of small to large effect on the disorder (Figure 23). However, the region between these markers could not be excluded. The MLS calculated in this region was 1.05 at D18S64 (60.5% sharing) based on the estimated sharing probabilities at this marker ($z_0 = 0.16, z_1 = 0.48, z_2 = 0.37$) (Table 20). When viewed separately, the percentage of shared compared to non-shared alleles inherited from the father were not significant at any locus when related to a χ^2 distribution. Similar results were obtained for the sharing of maternally inherited alleles except at marker D18S64, which showed significant excess of sharing ($\chi^2 = 5.00, p = 0.025$).

Table 20: Most Likely Lod Scores for Chromosome 18q (U.S.A. Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D18S474	23	26	46.9%	0.18	27	22	55.1%	0.51	
D18S64	21	18	53.8%	0.23	30	15	66.7%	5.00	
D18S68	22	21	51.2%	0.02	29	18	61.7%	2.57	
D18S61	26	27	49.1%	0.02	26	23	53.1%	0.18	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D18S474	52	50	51.0%	0.04	53.5	0.22	0.50	0.29	0.11
D18S64	53	35	60.2%	3.68	60.5	0.16	0.48	0.37	1.05
D18S68	52	40	56.5%	1.57	57.3	0.18	0.49	0.33	0.57
D18S61	53	51	51.0%	0.04	51.9	0.23	0.50	0.27	0.04

Figure 23: Chromosome 18q multipoint results using USA families



3.2 Linkage Results from Australian Families

3.2.1 Extension Study of Positive Regions Detected Using U.S.A. Families

An “extension study” refers to the process of testing additional families in the hope of reaching the genome-wide significance level (Lander and Kruglyak, 1995). After re-analysis of regions of the genome showing $MLS \geq 1$ in stage one of the study which was performed on the 97 U.S.A. sib-pairs, two regions still showed suggestive evidence of linkage to an autism susceptibility gene. These included a region within chromosome 1p and a large part of chromosome 22. These regions were genotyped in the sub-set of 41 families using the same microsatellite markers and identical analysis procedures. Due to the small number of available Australian families at present, the U.S.A. and the Australian families were analysed together rather than treating the Australian families as a "replication" data set. The results from this analysis are described below.

3.2.1.1 Chromosome 1p

With the addition of the Australian families the overall lod scores dropped slightly with 20% of this region on chromosome 1p excluded from linkage to a gene of large effect on autism (Figure 24). This compares to no exclusion of any gene effect when the first 90 U.S.A. multiplex families were analysed. When the Australian data were added to the maximum likelihood calculations, the maximum lod score at this region dropped from 1.55 (6cM centromeric of D1S1631) to 1.21 (8cM centromeric of D1S1631) (Table 21). Two-point linkage analysis on this

region showed a reduction in the MLS at D1S1631 from 1.13 to 0.58 (Table 22).

The MLS at this region on chromosome 1p was 0.88 at D1S534. This marker, when analysed using multipoint and two-point analysis, also had a percentage of allele sharing significantly different to that expected under the null hypothesis of no linkage ($\chi^2 = 3.98, p = 0.046$).

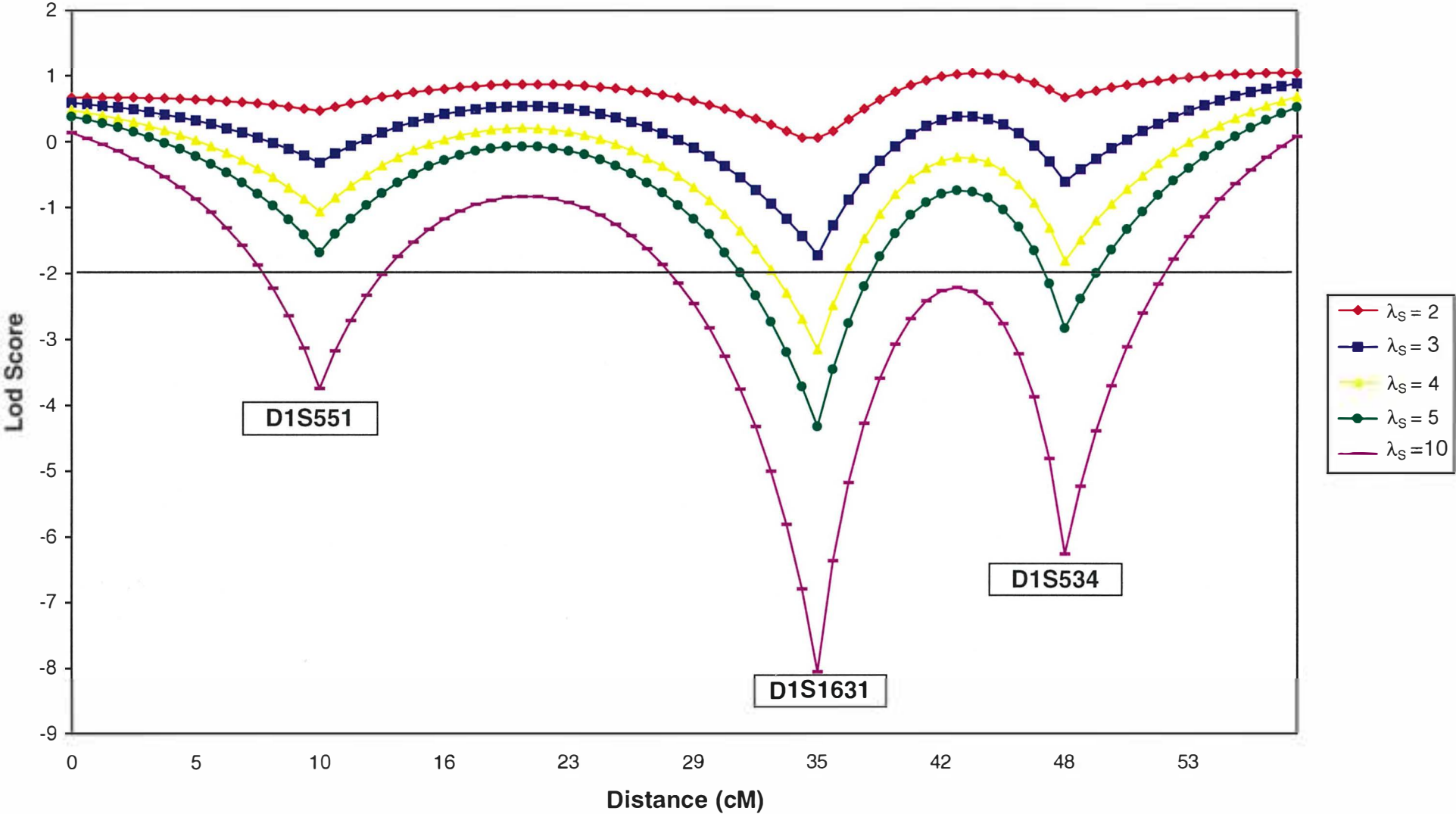
Table 21: Most Likely Lod Scores for Chromosome 1p (All Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D1S551	38	23	62.3%	3.69	32	31	50.8%	0.02	
D1S1631	44	35	55.7%	1.03	49	37	57.0%	1.67	
D1S534	42	32	56.8%	1.35	55	39	58.5%	2.72	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D1S551	71	55	56.3%	2.03	57.4	0.18	0.49	0.33	0.62
D1S1631	93	72	56.4%	2.67	56.9	0.19	0.49	0.32	0.75
D1S534	98	72	57.6%	3.98	58.1	0.18	0.49	0.34	1.04

Table 22: Two-Point Lod Scores for Chromosome 1p (All Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D1S551	38	23	62.3%	3.69	32	31	50.8%	0.02	
D1S1631	44	35	55.7%	1.03	49	37	57.0%	1.67	
D1S534	42	32	56.8%	1.35	55	39	58.5%	2.72	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D1S551	71	55	56.3%	2.03	56.5	0.19	0.49	0.32	0.45
D1S1631	93	72	56.4%	2.67	56.4	0.19	0.49	0.32	0.58
D1S534	98	72	57.6%	3.98	57.7	0.18	0.49	0.33	0.88

Figure 24: Chromosome 1p multipoint results using all families



3.2.1.2 Chromosome 22

Chromosome 22 results showed similar trends to those seen in chromosome 1p when the Australian autistic families' data were added. With the new data set, 20% of the chromosome could be excluded from linkage to a gene of large effect on the disorder (37%) (Figure 25). Genes of small to moderate effect on autism, however, could not be excluded from a location on chromosome 22.

Maximum likelihood calculations using the most-likely model showed a drop in MLS below the genome-wide significance level for linkage ($\text{lod} > 3$) at the marker locus D22S929. With the addition of the Australian families the lod score was reduced from 3.08 to 2.70 (Table 23). Percentage sharing compared to non-sharing between alleles in the combined data set was significantly different for a number of markers on this chromosome, including D22S929 ($\chi^2 = 6.92, p = 0.009$), D22S1163 ($\chi^2 = 7.11, p = 0.008$), D22S1150 ($\chi^2 = 8.26, p = 0.004$), D22S273 ($\chi^2 = 10.45, p = 0.001$), D22S280 ($\chi^2 = 7.41, p = 0.006$), D22S281 ($\chi^2 = 4.03, p = 0.045$) and D22S278 ($\chi^2 = 10.59, p = 0.001$).

When these alleles were divided into those inherited from the father and the mother, significant differences were noted only for the paternally inherited alleles. The number of alleles shared by siblings compared to those not shared was most significant at D22S1150 ($\chi^2 = 13.6, p = 0.0002$), with 73% of alleles inherited from the father by one sibling being identical to the other member of the sibling pair. When maximum likelihood calculations were performed on just the paternally inherited alleles, the MLS for this chromosome again exceeded the genome-wide level for significant linkage ($\text{lod} > 3$). This was noted at D22S280 with a MLS of 3.2. D22S280 is located 8.6cM telomeric of D22S929, the marker where the most significant linkage result was detected using only the 90 U.S.A. sib-pairs (MLS =

3.08). However, maximum likelihood calculations using the data on sharing of maternal alleles produced negative lod scores throughout most of the chromosome (except D22S420, MLS = 0.5). When the multipoint lod scores for the paternal and maternal inherited alleles were graphed under different lambda values, a similar trend was observed (Figures 26 and 27, respectively). As seen in these figures, the region most positive in the paternal multipoint graph (D22S315 to D22S281) was also the region most negative in the maternal multipoint graph. Sixty three percent of chromosome 22 could be excluded from containing a maternally inherited gene of large effect on autism. No region of this chromosome could be excluded from containing a paternally inherited gene for autism. These results indicate that a paternally derived gene could reside in the chromosome 22 region.

When two-point linkage analysis was run on this chromosome, the lod scores calculated showed a shift in MLS to marker D22S273. The calculated MLS was 2.47 with 67.4% sharing (Table 24). This marker is located 6.1 cM distal from D22S929. Percentage sharing at markers between D22S929 and D22S278 was still significantly different to that expected by chance when the paternally inherited alleles were analysed. Alleles inherited from the mother at all loci were not different to what was expected under the null hypothesis of no linkage.

Although the addition of the results of the Australian families to the chromosome 22 data reduced the overall lod scores, they were still above 1 for some markers and therefore suggestive of linkage to an autism gene. To increase the power to detect linkage, we combined our data with those obtained by the Stanford group. Chromosome 22 was re-analysed by the Stanford group to determine if the high positive linkage results obtained on this chromosome in stage one of the project could be duplicated. The results of the Stanford group were obtained using a number

of different microsatellite markers and these were analysed in the entire group of 139 U.S.A. multiplex families. The additional 49 US families were collected for stage two of the project and used by the Stanford group to saturate regions giving positive linkage evidence in stage one of the project (Risch *et al.*, 1999). The results of this analysis are shown in Table 25. With the addition of these new U.S.A. families, the maximum likelihood lod scores dropped significantly. The MLS for this chromosome was still at the marker locus D22S929 (the marker showing significant linkage to autism in the multipoint sib-pair analysis using the first 90 U.S.A. families), however it dropped well below the threshold for suggestive linkage with a value of 0.37. The presence of a gene on this chromosome could not be excluded since the percentage sharing, throughout the entire chromosome, was still above that expected by chance. Also, the percentages of shared alleles at the some loci were still significantly different in the combined data set when compared to a χ^2 distribution. These loci include D22S1163 ($\chi^2 = 4.44$, $p = 0.035$), D22S273 ($\chi^2 = 4.26$, $p = 0.039$), D22S278 ($\chi^2 = 5.33$, $p = 0.021$) and D22S283 ($\chi^2 = 4.41$, $p = 0.036$). Significant sharing was also noted in the percentage of paternal alleles shared between sib-pairs at all loci between D22S429 and D22S281, as well as at D22S278 and D22S277. The most significant result was at D22S278 ($\chi^2 = 8.91$, $p = 0.003$), with 82% of alleles inherited from the father by one sibling being identical to the other member of the sibling pair. No significant deviation from the null hypothesis of 50% sharing, however, was noted in the percentage of maternal alleles shared between sib-pairs at all markers on chromosome 22 in this combined data set. This strengthens the indication that if a gene for autism does reside on chromosome 22, it is most probably paternally derived.

Figure 25: Chromosome 22 multipoint results using all families

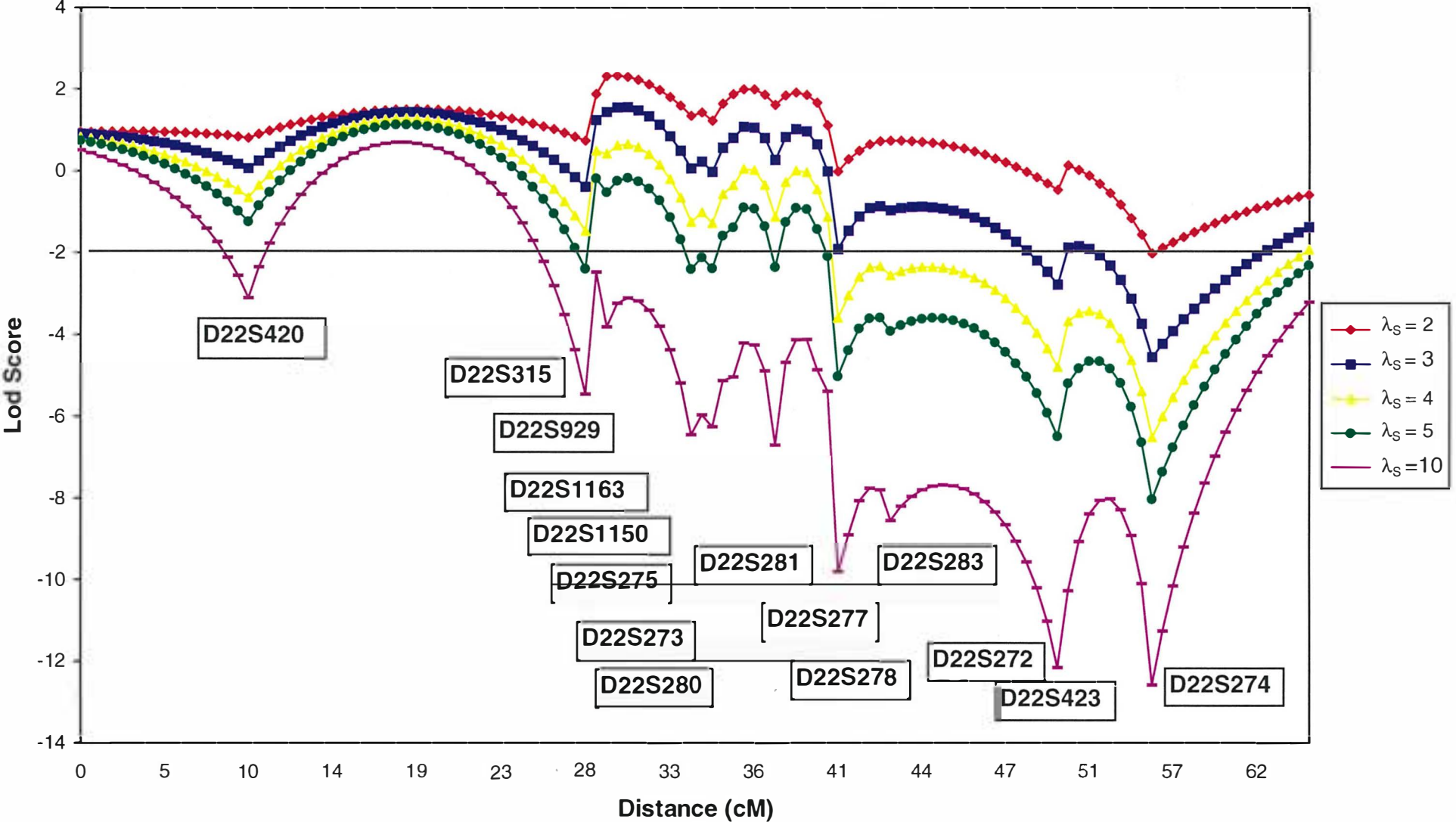


Table 23: Most Likely Lod Scores for Chromosome 22 (All Families)

Locus	Paternal				Maternal						
	+	-	%	χ^2	+	-	%	χ^2			
D22S420	33	25	57%	1.1	36	25	59%	2.0			
D22S315	40	22	65%	5.2	34	39	47%	0.3			
D22S929	43	20	68%	8.4	36	29	55%	0.8			
D22S1163	51	22	70%	11.5	35	32	52%	0.1			
D22S1150	48	18	73%	13.6	29	27	52%	0.1			
D22S275	38	17	69%	8.0	29	31	48%	0.1			
D22S273	31	12	72%	8.4	31	18	63%	3.5			
D22S280	52	25	68%	9.5	42	35	55%	0.6			
D22S281	50	22	69%	10.9	39	42	48%	0.1			
D22S277	51	26	66%	8.1	39	43	48%	0.2			
D22S278	23	5	82%	11.6	18	11	62%	1.7			
D22S283	44	27	62%	4.1	42	40	51%	0.1			
D22S272	30	15	67%	5.0	31	36	46%	0.4			
D22S423	21	19	53%	0.1	26	31	46%	0.4			
D22S274	33	24	58%	1.4	27	31	47%	0.3			
Locus	Combined				Most –Likely Model						
	+	-	%	χ^2	%shr	z0	z1	z2	P	M	MLS
D22S420	70	51	58%	2.98	59	0.17	0.48	0.35	0.5	0.5	1.0
D22S315	75	62	55%	1.23	57	0.18	0.49	0.33	1.2	-0.3	0.9
D22S929	80	50	62%	6.92	63	0.14	0.47	0.39	2.4	0.3	2.7
D22S1163	88	56	61%	7.11	60	0.16	0.48	0.35	2.5	-0.9	1.6
D22S1150	78	46	63%	8.26	59	0.16	0.48	0.35	2.7	-1.0	1.6
D22S275	67	48	58%	3.14	59	0.17	0.48	0.35	2.6	-1.1	1.5
D22S273	65	33	66%	10.45	61	0.15	0.48	0.37	2.9	-0.7	2.1
D22S280	95	61	61%	7.41	61	0.15	0.48	0.37	3.2	-1.1	2.1
D22S281	90	65	58%	4.03	60	0.16	0.48	0.36	2.6	-0.9	1.8
D22S277	90	69	57%	2.77	57	0.19	0.49	0.32	1.5	-0.6	0.8
D22S278	42	17	71%	10.59	58	0.18	0.49	0.33	1.6	-0.6	1.0
D22S283	88	69	56%	2.30	59	0.17	0.48	0.35	1.6	-0.1	1.5
D22S272	62	52	54%	0.88	57	0.18	0.49	0.33	1.0	-0.2	0.8
D22S423	47	50	49%	0.09	59	0.17	0.49	0.34	1.4	-0.3	1.2
D22S274	62	57	52%	0.21	53	0.22	0.5	0.28	0.4	-0.2	0.1

Figure 26: Chromosome 22 multipoint results (paternal inheritance) using all families

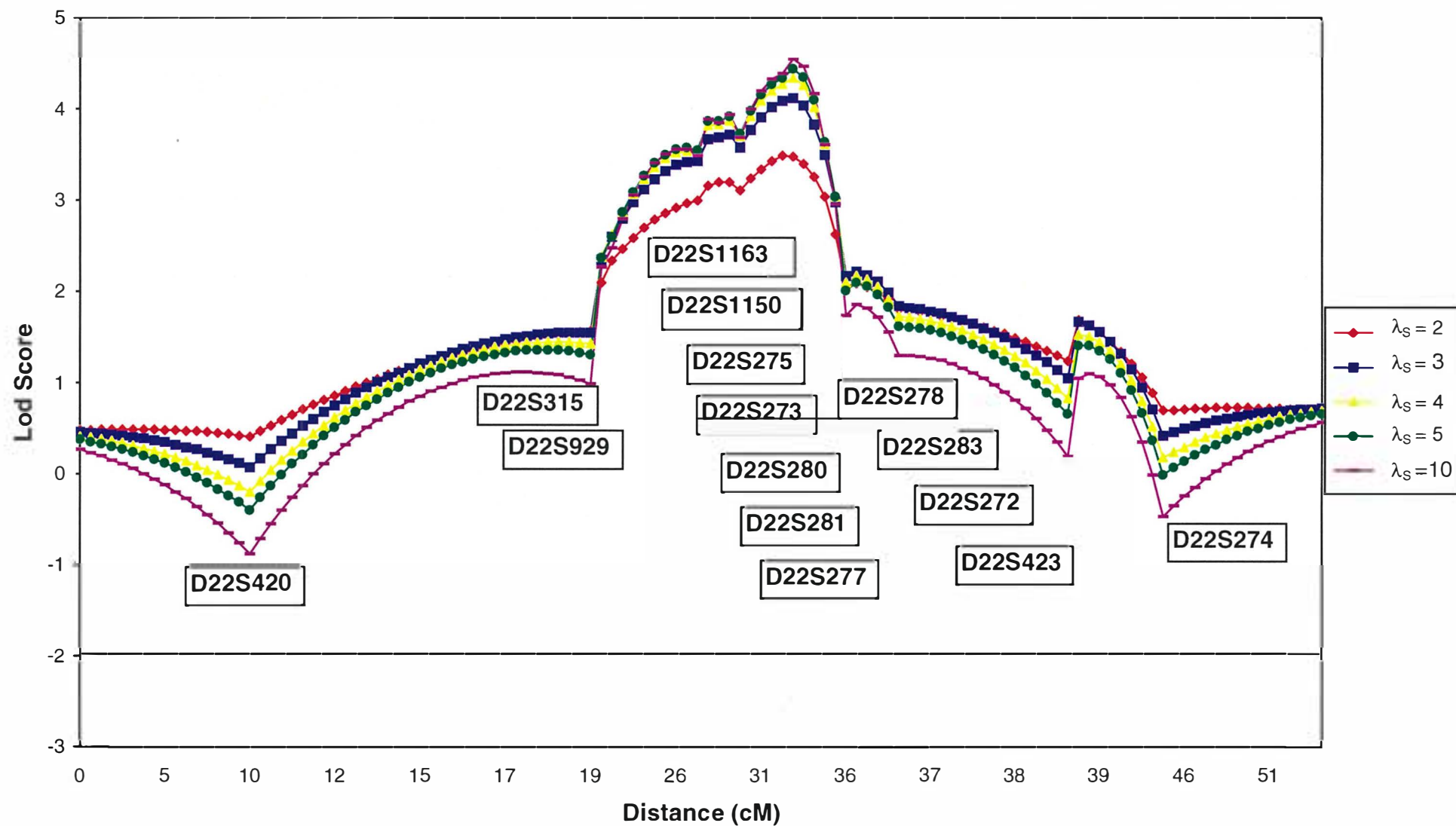


Figure 27: Chromosome 22 multipoint results (maternal inheritance) using all families

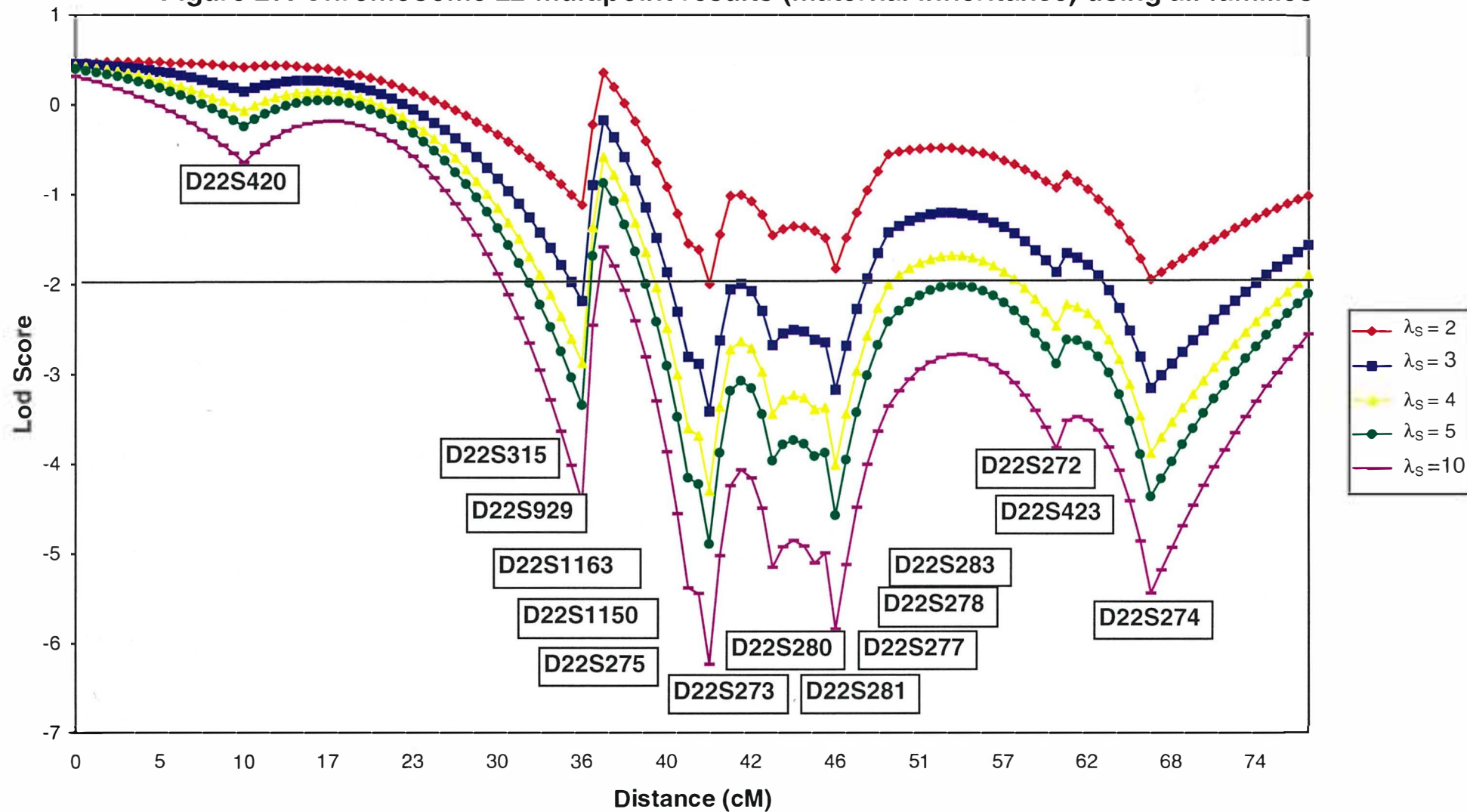


Table 24: Two-Point Lod Scores for Chromosome 22 (All Families)

Locus	Paternal				Maternal						
	+	-	%	χ^2	+	-	%	χ^2			
D22S420	33	25	57%	1.10	36	25	59.00%	1.98			
D22S315	40	22	65%	5.23	34	39	46.60%	0.34			
D22S929	43	20	68%	8.40	36	29	55.40%	0.75			
D22S1163	51	22	70%	11.52	35	32	52.20%	0.13			
D22S1150	48	18	73%	13.64	29	27	51.80%	0.07			
D22S275	38	17	69%	8.02	29	31	48.30%	0.07			
D22S273	31	12	72%	8.40	31	18	63.30%	3.45			
D22S280	52	25	68%	9.47	42	35	54.50%	0.64			
D22S281	50	22	69%	10.89	39	42	48.10%	0.11			
D22S277	51	26	66%	8.12	39	43	47.60%	0.20			
D22S278	23	5	82%	11.57	18	11	62.10%	1.69			
D22S283	44	27	62%	4.07	42	40	51.20%	0.05			
D22S272	30	15	67%	5.00	31	36	46.30%	0.37			
D22S423	21	19	52%	0.10	26	31	45.60%	0.44			
D22S274	34	24	59%	1.72	28	31	47.50%	0.15			
Locus	Combined				Most –Likely Model						
	+	-	%	χ^2	%shr	z0	z1	z2	Pat	Mat.	MLS
D22S420	70	51	57.9%	2.98	58.0	0.18	0.49	0.34	0.2	0.4	0.66
D22S315	75	62	54.7%	1.23	54.8	0.20	0.50	0.30	0.6	-0.4	0.27
D22S929	80	50	61.5%	6.92	61.7	0.15	0.47	0.38	1.6	-0.1	1.54
D22S1163	88	56	61.1%	7.11	61.4	0.15	0.47	0.38	2.1	-0.5	1.60
D22S1150	78	46	62.9%	8.26	63.1	0.14	0.47	0.40	2.5	-0.6	1.84
D22S275	67	48	58.3%	3.14	58.3	0.17	0.49	0.34	1.2	-0.5	0.68
D22S273	65	33	66.3%	10.45	67.4	0.11	0.44	0.45	1.8	0.7	2.47
D22S280	95	61	60.9%	7.41	61.0	0.15	0.48	0.37	1.8	-0.2	1.64
D22S281	90	65	58.1%	4.03	58.2	0.17	0.49	0.34	1.6	-0.7	0.89
D22S277	90	69	56.6%	2.77	56.6	0.19	0.49	0.32	1.2	-0.5	0.60
D22S278	42	17	71.2%	10.59	71.9	0.08	0.4	0.52	2.4	0.1	2.46
D22S283	88	69	56.1%	2.30	56.2	0.19	0.49	0.32	0.7	-0.2	0.51
D22S272	62	52	54.4%	0.88	54.5	0.21	0.5	0.3	0.5	-0.3	0.19
D22S423	47	50	48.5%	0.09	50.0	0.25	0.5	0.25	0.0	0.0	0.00
D22S274	64	57	52.9%	0.40	53.0	0.22	0.5	0.28	0.2	-0.1	0.09

Table 25: Most Likely Lod Scores for Chromosome 22 (Combined Data)

Locus	Paternal				Maternal			
	+	-	%	χ^2	+	-	%	χ^2
D22S420	8	10	44.4%	0.22	11	11	50.0%	0.00
D22S264	37	24	60.7%	2.77	37	37	50.0%	0.00
D22S311	29	25	53.7%	0.30	28	31	47.5%	0.15
D22S686	28	20	58.3%	1.33	32	25	56.1%	0.86
GCT10C10	21	12	63.6%	2.45	16	18	47.1%	0.12
D22S315	49	39	55.7%	1.14	42	43	49.4%	0.01
D22S429	30	15	66.7%	5.00	22	26	45.8%	0.33
D22S929	25	12	67.6%	4.57	16	18	47.1%	0.12
D22S1163	32	13	71.1%	8.02	22	21	51.2%	0.02
D22S689	36	19	65.5%	5.25	24	38	38.7%	3.16
D22S1150	31	13	70.5%	7.36	17	19	47.2%	0.11
D22S275	24	10	70.6%	5.76	18	19	48.6%	0.03
D22S273	24	8	75.0%	8.00	20	18	52.6%	0.11
D22S280	36	20	64.3%	4.57	28	31	47.5%	0.15
D22S281	33	15	68.8%	6.75	22	29	43.1%	0.96
D22S685	35	23	60.3%	2.48	25	30	45.5%	0.45
D22S683	42	27	60.9%	3.26	35	41	46.1%	0.47
D22S278	18	4	81.8%	8.91	13	11	54.2%	0.17
D22S277	18	7	72.0%	4.84	16	15	51.6%	0.03
D22S283	16	9	64.0%	1.96	17	9	65.4%	2.46
D22S445	27	18	60.0%	1.80	26	22	54.2%	0.33
D22S272	13	5	72.2%	3.56	10	15	40.0%	1.00
D22S423	35	29	54.7%	0.56	40	34	54.1%	0.49
D22S274	12	6	66.7%	2.00	10	8	55.6%	0.22

Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D22S420	19	21	47.5%	0.10	54.6	0.21	0.5	0.30	0.32
D22S264	74	61	54.8%	1.25	52.3	0.23	0.5	0.27	0.08
D22S311	60	59	50.4%	0.01	52.0	0.23	0.5	0.27	0.06
D22S686	60	45	57.1%	2.14	52.5	0.23	0.5	0.28	0.10
GCT10C10	39	32	54.9%	0.69	52.3	0.23	0.5	0.27	0.09
D22S315	93	84	52.5%	0.46	52.3	0.23	0.5	0.27	0.10
D22S429	52	41	55.9%	1.30	53.1	0.22	0.5	0.28	0.19
D22S929	41	30	57.7%	1.70	54.5	0.21	0.5	0.30	0.37
D22S1163	55	35	61.1%	4.44	52.8	0.22	0.5	0.28	0.14
D22S689	60	57	51.3%	0.08	52.6	0.22	0.5	0.28	0.12
D22S1150	50	34	59.5%	3.05	52.6	0.22	0.5	0.28	0.12
D22S275	42	29	59.2%	2.38	52.6	0.22	0.5	0.28	0.12
D22S273	47	29	61.8%	4.26	54.6	0.21	0.5	0.30	0.36
D22S280	64	51	55.7%	1.47	53.4	0.22	0.5	0.28	0.19
D22S281j	56	45	55.4%	1.20	52.2	0.23	0.5	0.27	0.09
D22S685	61	54	53.0%	0.43	52.3	0.23	0.5	0.27	0.09
D22S683	77	68	53.1%	0.56	53.1	0.22	0.5	0.28	0.16
D22S278	32	16	66.7%	5.33	53.1	0.22	0.5	0.28	0.16
D22S277	34	22	60.7%	2.57	53.1	0.22	0.5	0.28	0.16
D22S283	33	18	64.7%	4.41	54.0	0.21	0.5	0.29	0.20
D22S445	55	42	56.7%	1.74	54.0	0.21	0.5	0.29	0.23
D22S272	24	21	53.3%	0.20	54.0	0.21	0.5	0.29	0.23
D22S423	77	65	54.2%	1.01	54.1	0.21	0.5	0.29	0.25
D22S274	24	16	60.0%	1.60	54.5	0.21	0.5	0.30	0.21

3.2.2 Replication of Potential Susceptibility Regions

A genome screen by IMGSAC (1998) on 39 multiplex autism families, with follow up in an additional set of 60 families, found suggestive evidence of linkage to six chromosomal regions. The long arm of chromosome 7 from D7S530 to D7S684 (16.6cM apart) was the most significant region, with a multipoint MLS of 2.53 ($p = 0.0022$) calculated when all families were analysed. When the subset of 66 UK families were used, a MLS of 3.55 ($p = 0.00057$) was detected. Based on the sharing probabilities in this region ($z_0 = 0.05$, $z_1 = 0.50$, $z_2 = 0.45$), the significant linkage result in the UK families corresponds to a gene effect (λ_s) of 5.0.

Results from our genome scan in the same region showed a maximum lod score of 0.62 at the marker D7S684 (Risch *et al.*, 1999). This result was based on a total of 139 families and 147 affected sib-pairs. A slightly more positive result was found at marker D7S1804 (MLS = 0.93) which was located 10.3cM proximal of D7S684.

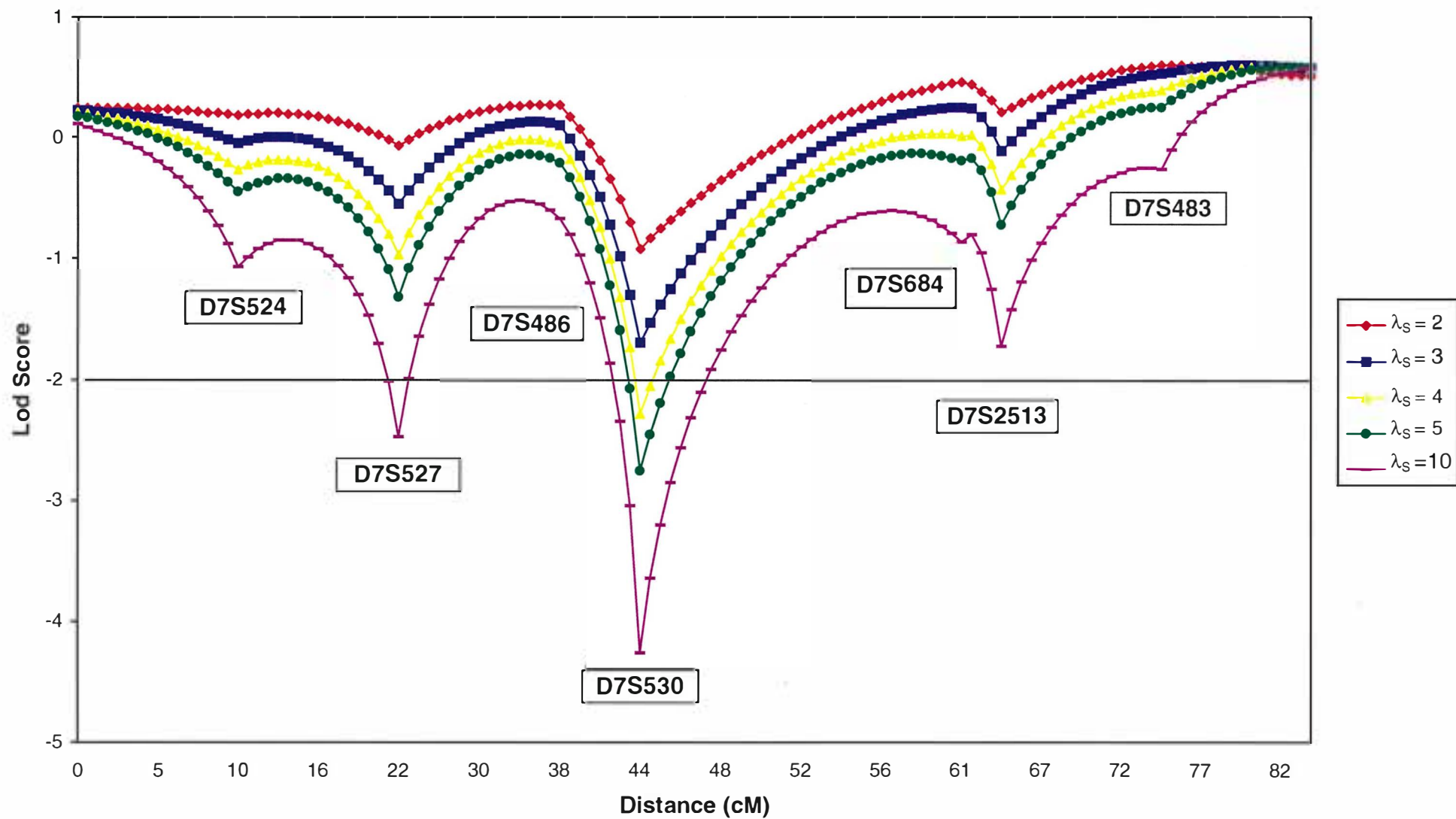
Using our Australian sub-set of families we examined 7 markers in the region found to be positive by IMGSAC (1998). These microsatellites were the same as utilised by the IMGSAC group to produce their suggestive evidence of linkage to the chromosome 7q region. The results of the analysis are shown in Figure 28. No region within this section of chromosome 7 could be excluded from linkage to a gene of large effect, with a MLS of 0.62 at D7S684 (Table 26). The MLS calculated using the Australian families was equal to that found in the U.S.A. data set. Significant excess of sharing was also found at D7S684 for the paternally inherited alleles, when compared to a χ^2 distribution ($\chi^2 = 5.40$, $p = 0.02$; Table 26). However, if a

susceptibility locus does reside in this region, its effect is likely to be small since no suggestive linkage ($MLS > 1$) was detected at any marker on chromosome 7q.

Table 26: Most Likely Lod Scores for Chromosome 7 (Australian Families)

Locus	Paternal				Maternal				
	+	-	%	χ^2	+	-	%	χ^2	
D7S524	10	7	58.8%	0.53	8	5	61.5%	0.69	
D7S527	10	9	52.6%	0.05	9	7	56.3%	0.25	
D7S486	7	5	58.3%	0.33	11	7	61.1%	0.89	
D7S530	8	9	47.1%	0.06	8	8	50.0%	0.00	
D7S684	12	3	80.0%	5.40	5	5	50.0%	0.00	
D7S2513	6	5	54.5%	0.09	9	8	52.9%	0.06	
D7S483	9	8	52.9%	0.06	13	6	68.4%	2.58	
Locus	Combined				Most –Likely Model				
	+	-	%	χ^2	%shr	z0	z1	z2	MLS
D7S524	18	12	60.0%	1.20	59.3	0.17	0.48	0.35	0.25
D7S527	20	17	54.1%	0.24	57.9	0.18	0.49	0.34	0.21
D7S486	18	12	60.0%	1.20	58.7	0.17	0.48	0.35	0.25
D7S530	16	17	48.5%	0.03	50.0	0.25	0.50	0.25	0.00
D7S684	17	8	68.0%	3.24	63.7	0.13	0.46	0.41	0.62
D7S2513	16	14	53.3%	0.13	59.7	0.16	0.48	0.36	0.31
D7S483	22	14	61.1%	1.78	62.8	0.14	0.47	0.39	0.57

Figure 28: Chromosome 7q multipoint results using Australian families



3.3 Linkage Disequilibrium Results

Linkage disequilibrium can be used in an exploratory fashion to investigate regions of the genome with suggestive evidence of linkage. Using the genotyping data from the combined U.S.A. and Australian families, transmission/disequilibrium tests (TDT) were run on chromosomes 1p and 22q to calculate linkage disequilibrium. The results from these analyses are shown below (Table 27 and 28).

For each marker, the table lists the number of alleles (Al) and the percentage of parents who are heterozygous at that position (%Het). The table also lists the most significant results on allele transmission frequency at each locus in the form of a χ^2 score (χ^2 Max).

Table 27: TDT Tests on Chromosome 1p Using All Families

Locus	Al	%Het	χ^2 Max
DIS551	7	72.0	2.06 ($p = 0.619$)
DIS1631	10	81.5	3.85 ($p = 0.542$)
DIS534	12	85.4	2.51 ($p = 0.787$)

Table 28: TDT Tests on Chromosome 22 Using All Families

Locus	Al	%Het	χ^2 Max
D22S420	14	65.6	0.72 ($p = 0.990$)
D22S315	17	79.8	6.00 ($p = 0.419$)
D22S929	11	78.3	4.45 ($p = 0.465$)
D22S1163	9	76.1	12.34 ($p = 0.005$)
D22S1150	13	73.4	3.77 ($p = 0.676$)
D22S275	10	80.0	2.28 ($p = 0.770$)
D22S273	9	68.5	2.13 ($p = 0.856$)
D22S280	10	79.8	2.47 ($p = 0.752$)
D22S281	11	83.5	3.00 ($p = 1.000$)
D22S277	17	86.5	4.00 ($p = 0.756$)
D22S278	7	77.0	2.00 ($p = 1.000$)
D22S283	16	86.5	5.31 ($p = 0.281$)
D22S272	13	66.1	2.25 ($p = 0.867$)
D22S423	16	83.7	5.83 ($p = 0.171$)
D22S274	11	77.5	3.33 ($p = 0.601$)

There was no evidence of linkage disequilibrium between the markers on chromosome 1p and autism, however this could be due to the large inter-marker distances (~19cM). In contrast to this, significant linkage disequilibrium was detected on chromosome 22q at marker D22S1163 ($P = 0.005$). This marker (D22S1163) is only 4.9cM away from D22S929, the marker showing significant linkage ($MLS \geq 3$) to autism in the multipoint sib-pair analysis using the first 90 U.S.A. families (see Appendix 6.1). These TDT results provide further evidence that a susceptibility gene for autism may reside in this region on chromosome 22.

CHAPTER 4

DISCUSSION

In this study, a large genome-wide screen was conducted by collaborating groups from Stanford, USA and Perth, Australia with the aim of detecting susceptibility loci for autism. Linkage to markers in genes that may be involved in possible pathobiochemical pathways of the disorder was also examined. The project is the largest genome screen of autism families conducted to date, with a total of 519 markers analysed in the initial set of 90 U.S.A. multiplex families and 149 markers studied in the second set of 49 U.S.A. multiplex sibships. An additional sample of 41 Australian families was analysed with 24 microsatellites in regions showing suggestive evidence of linkage to autism. The results of this PhD project formed part of an overall study whose findings were published recently in the *American Journal of Human Genetics* (Risch *et al.* 1999; see Appendix 6.3), the *Journal of Autism and Developmental Disorders* (Rogers *et al.*, 1999), and the *American Journal of Medical Genetics* (Salmon *et al.*, 1999).

4.1 Findings from the Study

4.1.1 Genome Scan

4.1.1.1 Susceptibility Loci

Since the first linkage study in autism, by Spence *et al.* (1985), there has been a number of advances in molecular biology techniques that have allowed genome

screens for susceptibility loci to be performed on a much larger scale. However, to date, there have been a limited number of linkage and association studies in autism and only three published genome scans (in addition to our study) that systematically span the entire genome (see section 1.8.5). Table 29 shows a summary of the positive findings reported in these studies. Only regions displaying significant associations or suggestive linkage results ($MLS \geq 1$) are shown.

Table 29: Chromosomal Regions Showing Positive Linkages and Associations to Autism (NOTE: MLS scores shown in brackets represent those results calculated by IMGSAC (1998) using the UK families only).

Chromosome (Locus)	Cytogenetic Position (in cM)	MLS or TDT χ^2	Publication
1	-	≥ 1	Bass <i>et al.</i> (1998)
2	-	≥ 1	Bass <i>et al.</i> (1998)
3	-	≥ 1	Bass <i>et al.</i> (1998)
4 (D4S412)	p16.3 (4.8)	1.55 (1.10)	IMGSAC (1998)
6 (HLA-DRB1)	p21.3 (56.7)	TDT $\chi^2 = 19.8$	Warren <i>et al.</i> (1996a)
6 (D6S283)	q21-23 (132.8)	2.23	Philippe <i>et al.</i> (1999)
7	-	≥ 1	Bass <i>et al.</i> (1998)
7 (D7S487)	q31-35 (135.1)	1.21	Pericak-Vance <i>et al.</i> (1998)
7 (D7S530)	q31-35 (144.7)	2.53 (3.55)	IMGSAC (1998)
9 (ABO blood group)	q34.1-34.2 (138.1)	1.54	Spence <i>et al.</i> (1985)
10 (D10S197)	q11.2 (51.9)	1.36 (0.69)	IMGSAC (1998)
11 (c-Harvey-ras gene)	p15.5 (6.2)		Herauld <i>et al.</i> (1994); Herauld <i>et al.</i> (1995); Comings <i>et al.</i> (1996)
15 (GABRB3 locus)	q11.2-q12 (4.1)	TDT $\chi^2 = 28.63$	Cook <i>et al.</i> (1998)
15 (D15S156)	q11-q12 (9.6)	2.5	Pericak-Vance <i>et al.</i> (1997)
15 (D15S118)	q14-q21 (41.1)	1.10	Philippe <i>et al.</i> (1999)
16 (D16S407)	p13.3 (17.3)	1.51 (1.97)	IMGSAC (1998)
16 (haptoglobin)	q22.2 (86.7)	1.04	Spence <i>et al.</i> (1985)
17 (5-HTTLPR; serotonin)	q11.2-q12 (57.4)	TDT $\chi^2 = 4.69$	Cook <i>et al.</i> (1998)
18	-	≥ 1	Bass <i>et al.</i> (1998)
19 (D19S226)	p13.3 (24.1)	1.37	Philippe <i>et al.</i> (1999)
19 (D19S49)	p13.1 (48.2)	0.99 (1.11)	IMGSAC (1998)
20	-	≥ 1	Bass <i>et al.</i> (1998)
22 (D22S264)	q11 (5.0)	1.39 (1.36)	IMGSAC (1998)
X (DXS424)	q24-q26 (125.1)	1.24	Hallmayer <i>et al.</i> (1996)

Our genome scan of 97 affected U.S.A. sib-pairs revealed 11 chromosomal regions linked to autism with MLS values greater than 1 (Table 30). These included two regions on chromosome 1 (1p [MLS = 1.87], 1q [MLS = 1.19]) and one region on each of chromosomes, 7p (MLS = 1.00), 10q (MLS = 1.0), 11p (MLS = 1.25), 13q (MLS = 1.49), 15q (MLS = 1.75), 17p (MLS = 1.30), 18q (MLS = 1.00), 20p (MLS = 1.09) and 22q (MLS = 3.08). While maximum lod scores as low as 1.0 are not regarded as conclusive evidence for linkage, they may indicate intervals worthy of further investigation (Elston, 1992; Hauser *et al.*, 1996). The proportion of regions that represent false-positive findings is, however, difficult to assess. Hauser *et al.* (1996) demonstrated by simulation that, when studying 100 nuclear families using a genetic map with an average intermarker distance of 10cM, there is a 2.4% false positive rate for regions with a lod score of 1.0. This represents approximately 8 false positives, on average, in a genome scan of 300 markers. A lod score of 2.0 represents no more than one false positive in the same data set. In our genome scan, where 362 microsatellites were used in the initial analysis (~10cM map), this would be equivalent to around 9 false positive localisations using the same assumptions. The observation of 11 regions with a lod score of ≥ 1.0 is therefore not different to what could be expected by chance.

To help discriminate between the false positive and genuine results, more markers in these regions were analysed in the same group of families, to provide more dense or completely different chromosomal maps. The maximum lod scores for all the suggestive regions, except those on chromosomes 1p and 22q, dropped below 1.0 (Table 30). It has recently been demonstrated that, for true positive localisations, the positive findings span wider genetic areas than false positive

localisations of similar lod scores (Terwilliger *et al.*, 1997). The two chromosomal regions showing $MLS > 1$ with the addition of more markers (regions 1p and 22q) were also flanked by further markers with lod scores > 0.5 . This increases the probability that these regions represent true positive localisations.

Table 30: Maximum Lod Scores (MLS) at Regions Previously Suggestive of Containing Susceptibility Loci for Autism, After the Addition of More Markers Within the Region.

Chromosome	MLS in Stage 1	MLS in Stage 2
1p	1.87	1.55
1q	1.19	0.71
7p	1.00	1.00
10q	1.00	0.76
11p	1.25	0.90
13q	1.49	0.70
15q	1.75	0.22
17p	1.30	0.81
18q	1.00	1.00
20p	1.39	0.80
22q	1.40	3.08

Chromosome 22q

The original analysis of 6 markers on chromosome 22 indicated that a gene might reside in a region on chromosome 22q ($MLS = 1.40$), and subsequent analysis using a more dense marker map still resulted in positive lod scores. A literature

search was undertaken to see if any relevant genes resided in this region. The gene for Neurofibromatosis 2 (NF2), designated Schwannomin (SCH), was found to be located in this region on chromosome 22 (22q12.2). A marker known to link to NF2, D22S929, was analysed for linkage to autism and the calculation of a MLS of 3.08 indicates that this gene may be involved in the disorder.

NF2 is a neurocutaneous disorder characterised by tumours within the central and peripheral nervous system (see section 1.5.1.2). There are many similarities between neurofibromatosis and autism. In particular, they are both associated with brain dysfunction and mental retardation (Gillberg *et al.*, 1987; Wadsby *et al.*, 1989). They both show characteristic impairment of cognitive functioning, displaying a verbal/language disability (Rutter, 1983; Eldridge *et al.*, 1989). Two genetic studies have also noted an association of autism with neurofibromatosis. Gillberg and Forsell (1984) reported mild cases of NF1 in 3 out of 51 (6%) autistic children collected through a community survey, and Gaffney and Tsai (1987) found the occurrence of both disorders in 2 out of 14 (15%) autistic children seen in a child psychiatric clinic over one year. However, a study (Mouridsen *et al.*, 1992) comprising 341 children with autism and other types of childhood psychoses seen in two university clinics of child psychiatry in Denmark (Copenhagen and Arrhus) during a 25-year period found only one case (0.3%) with both autism and neurofibromatosis. This sample has a bias of hospital selection with only seriously affected individuals who required hospitalization included in the study. This bias, as well as the fact that there was no information about the diagnostic criteria used to diagnose neurofibromatosis in these subjects, could have resulted in milder cases being missed. Due to these published associations and suggestive linkage between

these two disorders in a sub-group of the analysed families in this study, NF2 should be considered as a candidate gene in future studies.

Further evidence suggesting that a gene for autism may reside on chromosome 22 comes from the full genome screen for autism by the IMGSAC (1998). They calculated a MLS of 1.39 (59.7% sharing) in a region close to D22S264 (22q11.2) using all 99 affected sib-pair families. This locus is approximately 21.5cM centromeric of D22S929. When only the UK families were analysed, the MLS at this region was 1.36 with 63.7% sharing. Figure 29 shows the relative positions of these positive findings on the chromosome 22 cytogenetic map. The results from this study are shown in bold.

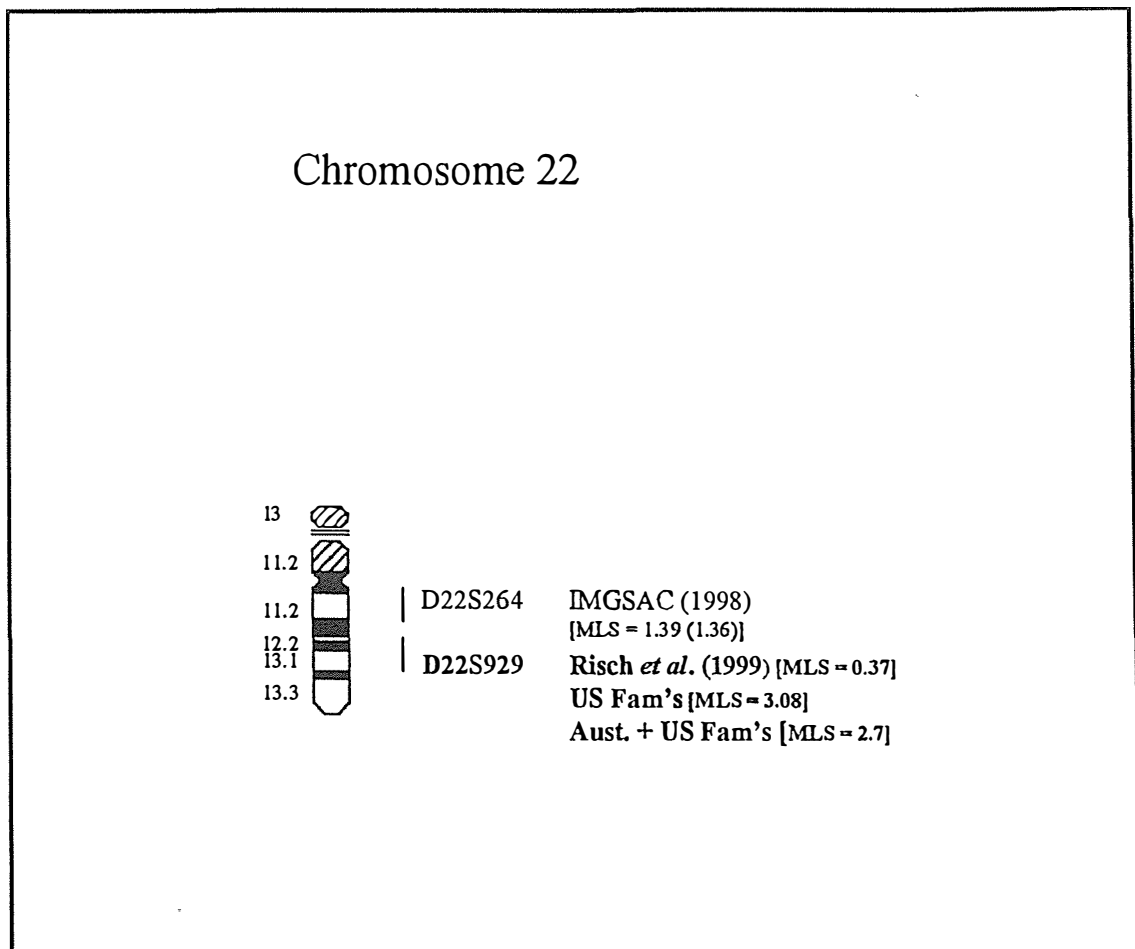


Figure 29: Cytogenetic map of chromosome 22 showing positive linkage results

To increase the power to detect true positive regions with statistical parameters that are significant at the genome-wide level, more families (45 Australian ASPs) were added to the analysis. With the addition of the 41 Australian families the lod score dropped below the genome-wide significance level to a MLS of 2.7 (D22S929; 22q12.2). The highest two-point lod score was calculated at D22S273 (MLS = 2.47), a marker 6.1cM telomeric of D22S929. It was also noted that with the addition of these families, the MLS scores for the paternally inherited alleles were still very positive, with the highest multipoint MLS calculated at D22S280 (MLS = 3.2) and the highest two-point MLS computed at D22S1150 (MLS = 2.5). The percentage of maternally inherited alleles shared by the siblings, however, was not greater than that expected by chance. If a gene resides in this region on chromosome 22, it is most probably paternally derived.

TDT analyses on this chromosome using the combined U.S.A. (97 ASPs) and Australian (45 ASPs) data showed significant association between D22S1163 and autism. This marker is 4.9cM telomeric of D22S929, providing further evidence that a susceptibility gene may reside in this region.

Multipoint linkage results from a further group of 50 affected U.S.A. sib-pairs were added to the combined U.S.A. (97 ASPs) and Australian (45 ASPs) data on chromosome 22q. The lod scores in chromosome 22q dropped significantly with the highest MLS of only 0.37 at D22S929. The percentage sharing over this region was still above that expected by chance, therefore a gene of small effect could not be entirely excluded from the region on chromosome 22q.

Chromosome 1p

Overall, the most significant finding in the present study was on chromosome 1p. In the initial analysis of this region with 90 multiplex families, a MLS of 1.55 was calculated between markers D1S1631 and D1S534 (1p13.3). The lod score at this locus dropped to 1.21 with the addition of the Australian families (45 ASPs). However, with the addition of 50 affected U.S.A. sib-pairs to the original U.S.A. data (97 ASPs) already analysed on chromosome 1p, linkage results showed an increase in MLS to 2.15 with 60% allele sharing between the siblings (results not shown, see Risch *et al.*, 1999 [Appendix 6.3]). Since the lod scores at a region on chromosome 1p stayed positive ($MLS \geq 1.0$) with the addition of more markers and the expansion on the number of families analysed, this suggests that a gene causing susceptibility to autism may reside in this region.

Preliminary analysis by Bass *et al.* (1998) on 52 multiplex families also identified chromosome 1 as a promising region with the calculation of a lod score ≥ 1.0 . The exact chromosomal position of this finding was not explained in the Bass *et al.* (1998) report, therefore it remains unclear whether the positive region on chromosome 1 identified in our study overlaps with that found by Bass *et al.* Our results on this chromosome are shown diagrammatically on Figure 30.

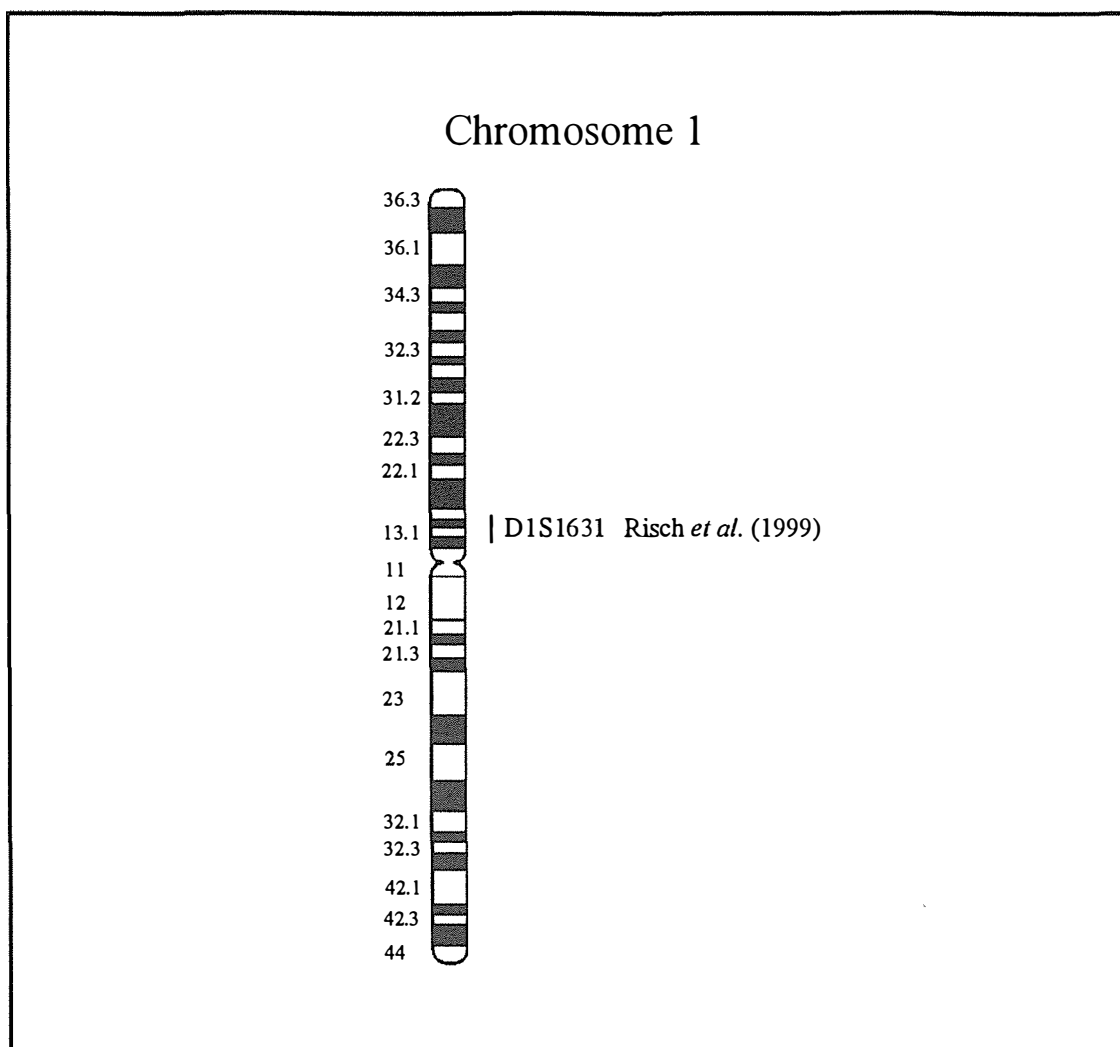


Figure 30: Cytogenetic map of chromosome 1 showing positive linkage results from the present study

Possible candidate genes in the 1p region, expressed in the brain, include glutathione-S-transferase 3 (GSTM3) and adenosine receptor A3 (ADORA3). The adenosine receptor is of particular interest, since these receptors have been utilised as targets for the development of clinically viable drugs against ischaemic brain disorders. A gene for Waardenburg syndrome type II (WS2B) has also been mapped to this region (1p13.3) (Lalwani *et al.*, 1994). This neurocutaneous disorder is characterised by deafness with pigmentary abnormalities, and, like autism, it is associated with mental retardation (Fujimoto *et al.*, 1998).

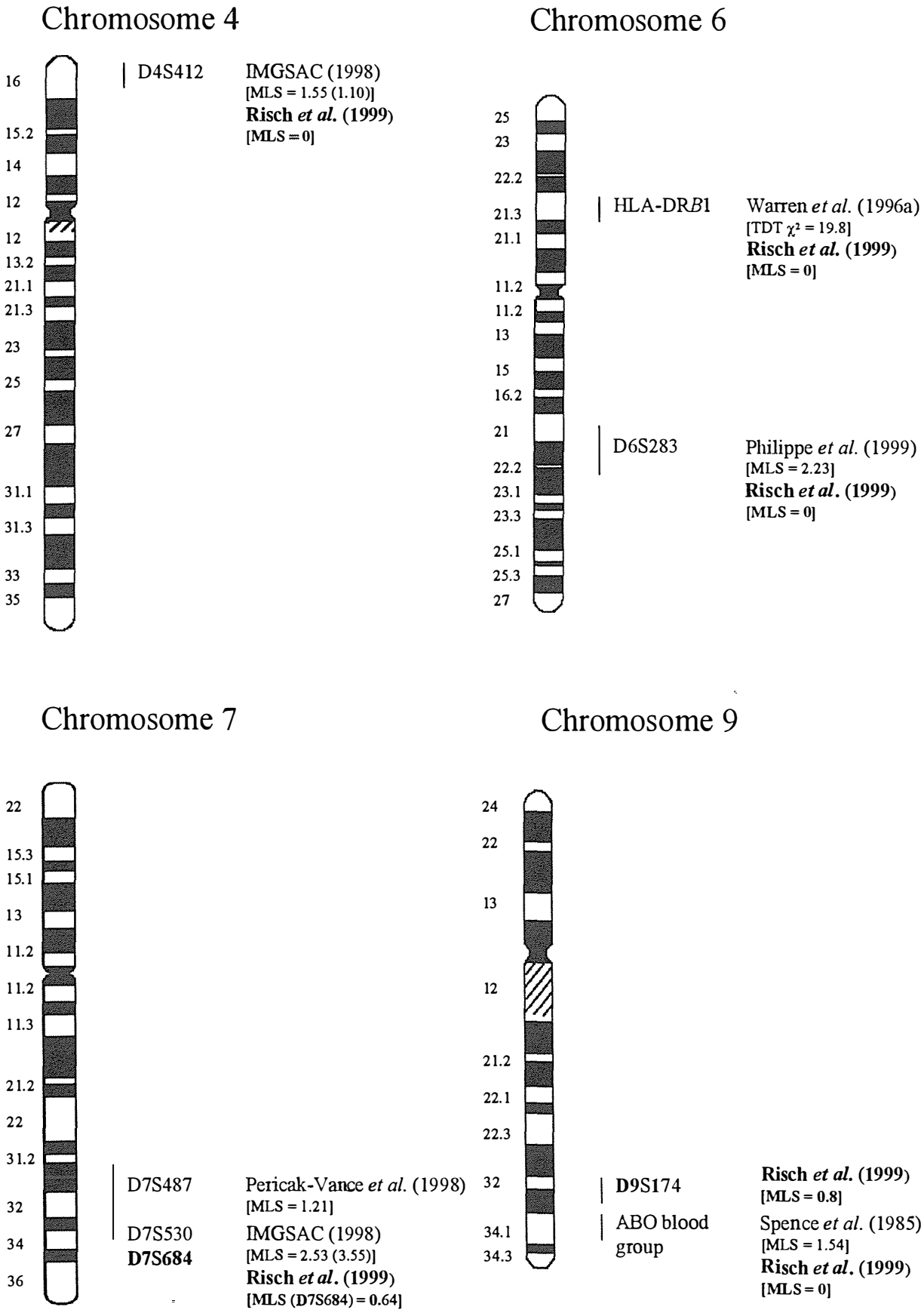
4.1.1.2 Excluded Regions

Results from this PhD study can exclude an average of 94% of chromosomes 15, 19, 20 and 21 from containing a gene of large effect on susceptibility to autistic disorder. Genes of moderate effect can also be excluded from residing in 65% of these chromosomes. A gene of moderate to large effect on autism was also excluded in the HLA region of chromosome 6 (see section 4.1.2) and large parts of 17q and 10q. However, a gene of moderate to large effect on autism could not be excluded from possibly residing on chromosomes 1p, 7q, 18q or 22q.

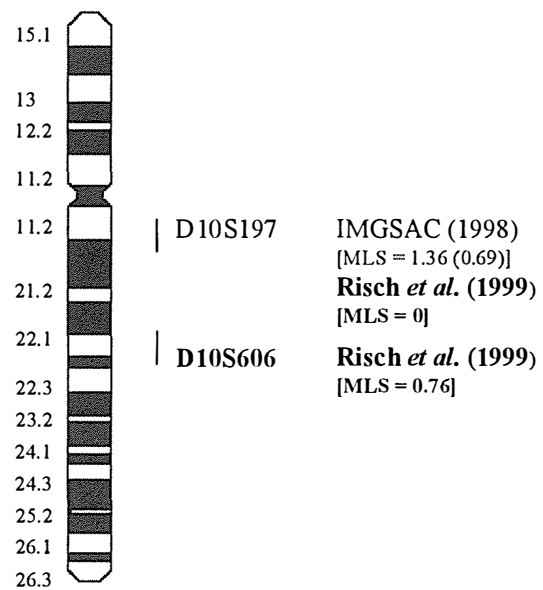
Combining the results of the entire genome scan (Risch *et al.*, 1999), we were able to exclude approximately 95% of the genome for a value of $\lambda_S = 3.0$, using the strict exclusion criterion of -2 (Hauser *et al.*, 1996). Using a less strict exclusion criterion of -1 , 99% of the genome could be excluded for genes of moderate to large effect. Figure 31 shows the diagrams of regions from other studies that may contain a gene involved in the aetiology of autism, with our results in the same region shown in bold. It can be seen from these diagrams that our results exclude chromosome 4p, 6, 15q, 16, 19p and Xq from containing an autism-causing gene. However, regions on chromosomes 7q, 9q, 10q, 11p, 17, 22q previously showing suggestive evidence could not be formally excluded in this study and should be investigated further as possible candidate regions for predisposing loci.

Chromosome 7q is of particular interest as a site of susceptibility loci in autism, due to the significant lod score (MLS = 3.55) obtained in this region in the IMGSAC (1999) study (see section 1.8.5.9). This chromosome was also examined in detail by Pericak-Vance *et al.* (1998). They examined 7q in 65 multiplex families using sib-pair analysis (SIBPAL), and identified 5 markers showing suggestive evidence of linkage ($p < 0.05$), with their most significant result at D7S2527 ($p =$

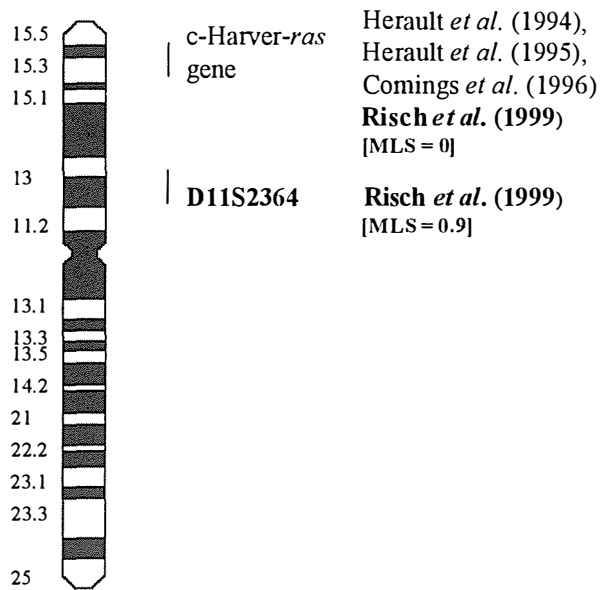
Figure 31: Cytogenetic maps of chromosomal regions showing $MLS \geq 1$ (The results of the present study are shown in bold)



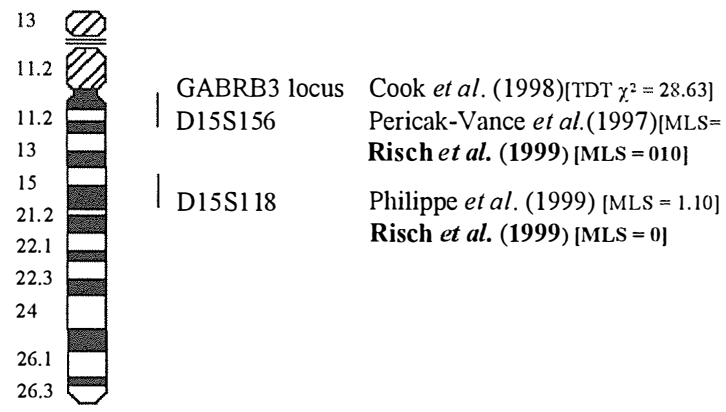
Chromosome 10



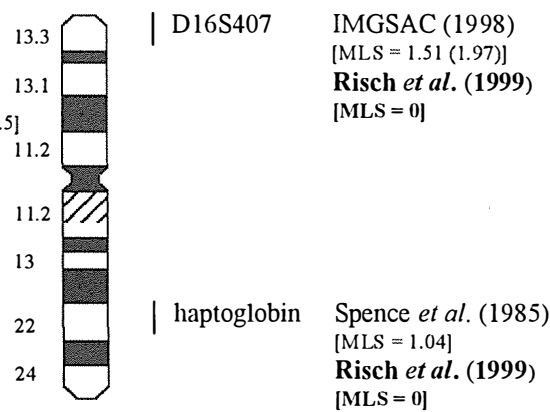
Chromosome 11



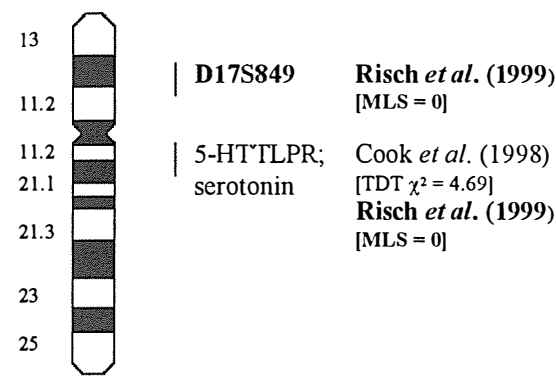
Chromosome 15



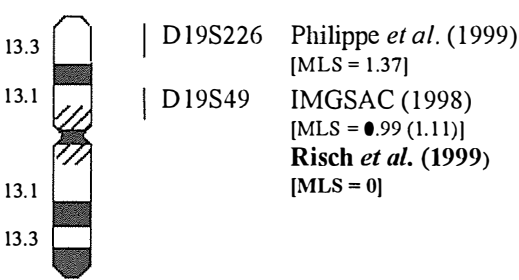
Chromosome 16



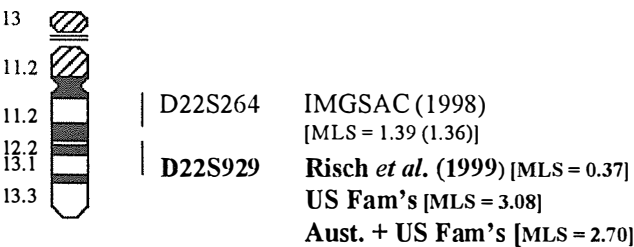
Chromosome 17



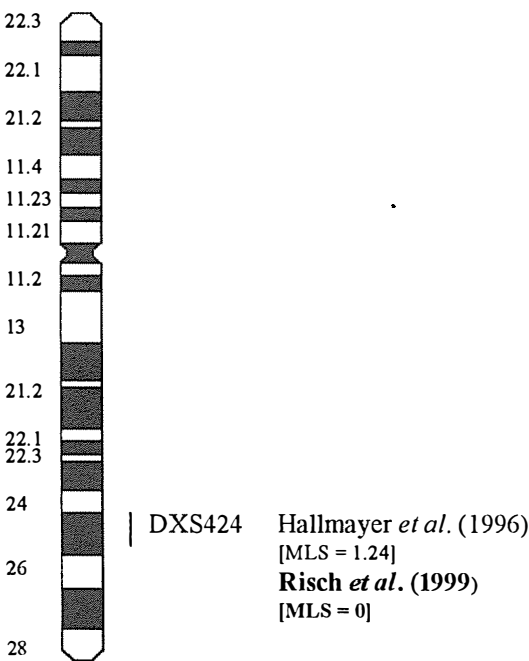
Chromosome 19



Chromosome 22



Chromosome X



0.002). A maximum lod score of 1.21 was found at D7S487. A chromosomal abnormality has also been identified on chromosome 7 that overlaps the potential linkage region identified by IMGSAC (1998). A family with two male siblings who have autism (as diagnosed using the ADI), and a female sibling with significant developmental delays and expressive language disorder, each have an identical paracentric inversion (Inv (7) (q22q31.2)).

4.1.2 Candidate Regions

The HLA and 15q11-q13 regions were considered as possible candidates for containing a gene(s) involved in the aetiology of autism, due to the positive association being found in other studies (see section 1.8.5). The results from our analysis of these regions have recently been published in relevant journals (see Appendix 6.3).

Linkage results from the analysis of the HLA region on chromosome 6 in 90 U.S.A. sib-pair families demonstrated that a large to moderate gene effect could be excluded for 100% of this region (see section 3.1.3; Rogers *et al.*, 1999). With the addition of 49 multiplex U.S.A. sibships, the data in this region remained negative (Risch *et al.*, 1999). A gene of small effect ($\lambda_s \leq 2$), however, could not be excluded from this region. Similarly, TDT analysis did not produce any positive findings (see Rogers *et al.*, 1999; Appendix 6.3). These results do not entirely exclude a possible association between autism and the HLA region. However, if an association does exist, it must be weak.

The reported immunological abnormalities found in autistic children (Warren *et al.*, 1986; Stubbs and Crawford, 1977; Weizman *et al.*, 1982; Singh *et al.*, 1993; Plioply *et al.*, 1994) are unlikely to be related to susceptibility genes in the HLA region. The immune system is connected to the brain in an integrated regulatory circuit, which is only slowly beginning to be understood. One of the integral organs in this network is the hypothalamus, which also plays an important role in cognitive and emotional processes. An altered immune response during early childhood might produce changes in brain development. Although it appears that genes in the HLA region have little or no effect on this process, the possibility that genes of immunological interest other than HLA might be involved in the pathophysiology of autism has not been excluded.

A paper describing the results of the analysis of 5 markers in the proximal region of chromosome 15q in 90 multiplex U.S.A. families has been published in the U.S.A. Journal of Medical Genetics (Salmon *et al.*, 1999; see Appendix 6.3). These results exclude this region for the presence of a susceptibility gene of strong or moderate effect ($\lambda_s > 2$) on autism. Allele-sharing results at marker D15S156 (maximum lod score of 2.5 detected by Pericak-Vance *et al.* (1997) in this region) was slightly increased in our data (57%), however other nearby markers which were more informative (eg. D15S822) showed a decrease in sharing (48%). The multipoint results across this entire region show no significant increase in sharing (51%). The International Autism Consortium (1998) also scanned chromosome 15q in 39 multiplex families and found no evidence of a major susceptibility gene for autism, consistent with the results presented here.

The GABRB3 locus revealed linkage disequilibrium in the sample of autistic individuals studied by Cook *et al.* (1998; TDT 28.63, $p = 0.0014$). A marker within

this locus (GABRB3CA) was tested in our families for association with autistic disorder. We found no evidence of linkage disequilibrium with GABRB3CA. This marker is 250kB distal from the microsatellite (GABRB3 155CA-2) at which Cook *et al.* (1998) found suggestive evidence of linkage disequilibrium, therefore the two markers may not be in linkage disequilibrium. An association study by Menold *et al.* (1998) also attempted to replicate the Cook *et al.* (1998) findings. Using 147 autistic families, they failed to detect significant linkage disequilibrium between autism and the GABRB3 155CA-2 marker. A study by Pericak-Vance *et al.* (1997), however, obtained a lod score of 1.4 at the GABRB3 locus when analysing 38 sib-pairs and 9 cousin pairs. Their peak lod score, assuming an autosomal recessive model, was 2.5 for the microsatellite marker D15S156, which is located approximately 5cM distal to the GABRB3 locus. When non-parametric linkage analysis was employed using the affected sib-pair analysis, the lod score dropped to a MLS of 1.1.

Our results clearly indicate that, for the majority of affected sib-pairs, there is no major autism gene in the 15q11-q13 region. Nonetheless, chromosome 15, in particular 15q11-q13, must still be viewed as an area of interest, given the number of reports showing association between autism and cytogenetic abnormalities in this region.

4.2 Replication Studies

For linkage results to be credible, they must be replicated. Failure to replicate some of the positive results in one study does not necessarily disprove a hypothesis, since genes may have small effects on the disorder, which may turn out to be smaller

in a second study if the sample collected selects against this gene. Thus positive results may be replicated in some studies and not in others. To assess whether the overall evidence for linkage is convincing, a meta-analysis of all studies may be necessary. Table 31 lists those regions where other studies have suggested a gene may reside, but which were not confirmed in this study.

Table 31: Susceptibility Loci Not Replicated in this Genome Scan

Chromosome (Locus)	Cytogenetic Position	Publication
4 (D4S412)	p16.3	IMGSAC (1998)
6 (HLA-DRB1)	p21.3	Warren <i>et al.</i> (1996a)
6 (D6S283)	q21-23	Philippe <i>et al.</i> (1999)
15 (GABRB3 locus)	q11.2-q12	Cook <i>et al.</i> (1998)
15 (D15S156)	q11-q12	Pericak-Vance <i>et al.</i> (1997)
15 (D15S118)	q14-q21	Philippe <i>et al.</i> (1999)
16 (haploglobin)	q22.2	Spence <i>et al.</i> (1985)
19 (D19S226)	p13.3	Philippe <i>et al.</i> (1999)
19 (D19S49)	p13.1	IMGSAC (1999)
X (DXS424)	q24-q26	Hallmayer <i>et al.</i> (1996)

There could be several reasons why we did not replicate the positive results reported by IMGSAC (1998), Philippe *et al.* (1999) and other studies. These include: false negative results due to a smaller sample size than required to replicate studies, false positive results in the initial study, genetic heterogeneity between data sets within or between studies, and differences in the diagnostic criteria used in these studies. These points are discussed below.

The accurate replication of previous results requires a very large sample of families. Suarez *et al.* (1994) demonstrated that the number of families required to reliably replicate linkage in a polygenic disease is approximately $n(k-1)$, where n is the number of sib-pairs that were needed to demonstrate the initial linkage and k is the putative number of susceptibility genes. This is the case in autism where it is generally accepted that multiple genes are involved in the aetiology. Multiple loci that interact to increase disease risk are difficult to detect, since each parent may carry several susceptibility alleles and multiple, distinct combinations of these alleles might contribute to disease in different children and in different pedigrees. Thus, for any one locus, the pattern of allele sharing among affected relatives is variable and complex. With smaller samples, statistical support for linkage will vary because of random (stochastic) variation in the proportion of families in whom disease is associated with inheritance of a particular parental susceptibility allele (Suarez *et al.*, 1994). Therefore some samples will produce significant lod scores in the chromosomal regions containing such genes and other samples will produce weakly positive or negative scores in the same region. Scores in the same range may also be observed by chance in regions that do and do not contain disease genes, making true and false positive results difficult to distinguish statistically.

The type 1 error rate could also have been higher in the initial study (ie. the IMGSAC [1998] study). The appearance of positive linkage results in the original study could be due to reasons such as the use of multiple testing in the analysis (Clerget-Darpoux *et al.*, 1990). The presence of false positives in the IMGSAC (1998) and Philippe *et al.* (1999) studies could be a reason why these results were not found in our genome scan.

The failure of replication between these studies could also be due to the different populations utilised in the analysis. The failure of replication between populations can be explained by inter-ethnic genetic heterogeneity, in which predisposing genotypes and their interaction with the surrounding environment may vary between populations (Risch and Botstein, 1996).

4.3 The Genetics of Autism – An Overview of Findings to Date

This study was designed under the assumption of an oligogenic model of inheritance, with 2 to 6 susceptibility loci involved in the aetiology of autism. The detailed and comprehensively defined nature of this project should have led to the detection of these loci if our hypothesis was true. The results from the study show that our hypothesis is most unlikely and they do not support the assumption of a small number of loci (see Risch *et al.*, 1999).

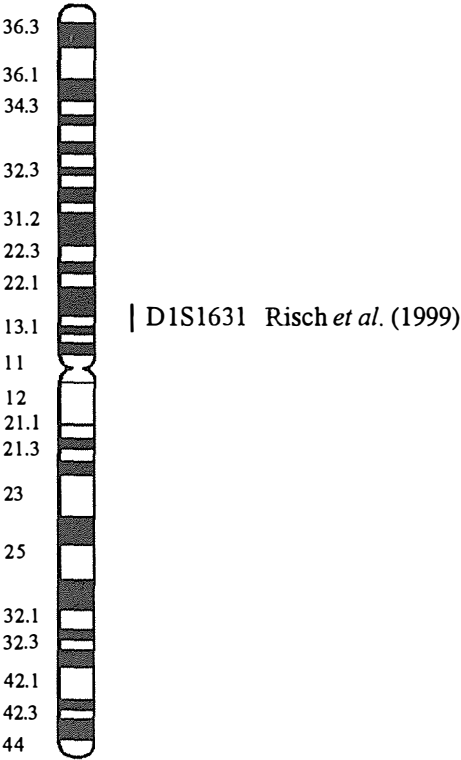
Using the results from our genome scan, a statistical analysis was performed assuming a multiplicative epistatic model of inheritance to determine the number of loci most probably involved in this disorder (Risch *et al.*, 1999). The data most closely approximate the expectations for a model with 20 susceptibility loci, with models involving ≤ 10 loci unlikely and those with ≥ 15 more plausible. These results are consistent with the observed recurrence risks between siblings and twins. The very high (25-fold) MZ:DZ concordance ratio is also indicative of the involvement of multiple interacting loci. Our results also show that whilst a multilocus inheritance model does not preclude the possibility of one or a few loci with larger effects, these effects can not be too large given the non significant linkage results that

were found. Our genome scan therefore reveals that the inheritance pattern for autism is multigenic, with ≥ 15 susceptibility loci, each with a small effect on disease risk ($\lambda_s < 2$), most probably involved in the aetiology. These results go against the Pickles *et al.* (1995) model for autism which suggested that 3 interactive loci were involved in the aetiology of autism, with a range of 2-10 loci probable (see section 1.7). The results of Pickles *et al.* (1995) were based on probands with a broad autistic phenotype, whereas our findings were based on individuals with a "core" diagnosis of autism.

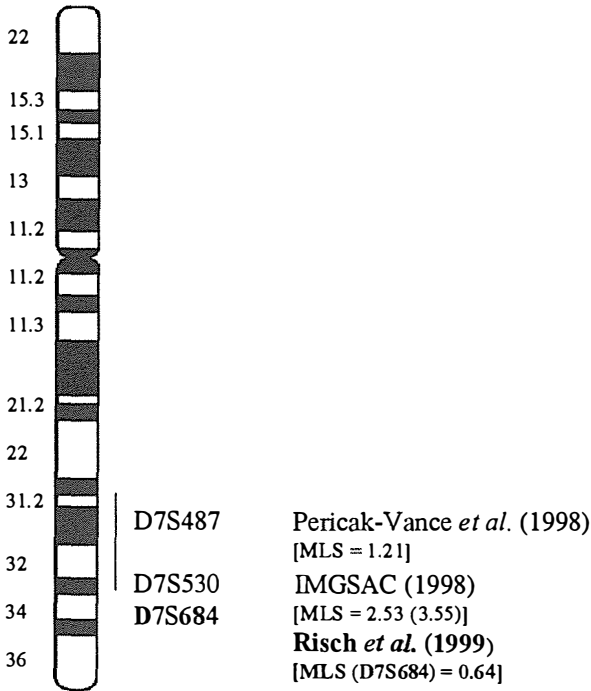
No chromosomal region with significant evidence of linkage ($MLS > 3$) was identified in the present study, however the possibility of one or more susceptibility loci of moderate effect on the disorder could not be excluded. On the basis of our results, the most likely location for such a locus or loci is on chromosomes 1p and 22q. Our results could also not exclude chromosomes 7q, 9q, 10q, 11p and 17, regions previously suggested by other studies to contain disease-predisposing loci. Whilst none of these regions show linkage results above the genome-wide level of significance, the calculated lod scores are above 1 and therefore the regions should be considered in future studies as candidates for further analysis. Figure 32 shows those regions most likely to contain autism-predisposing loci, with our results shown in bold.

Figure 32: Cytogenetic maps of chromosomal regions most likely to contain susceptibility loci for autism

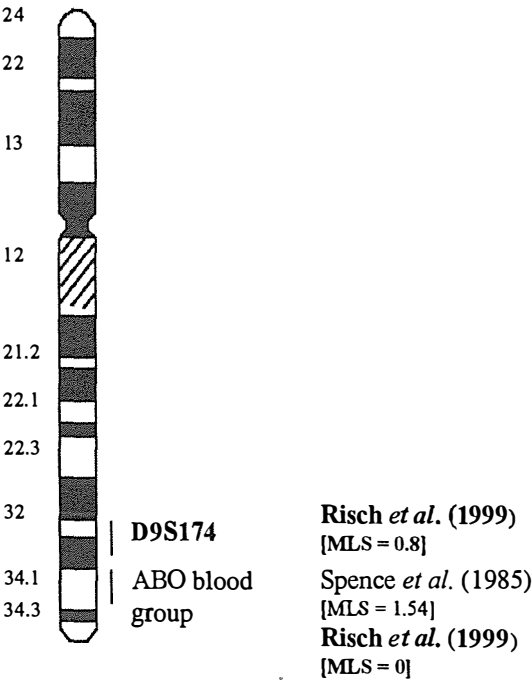
Chromosome 1



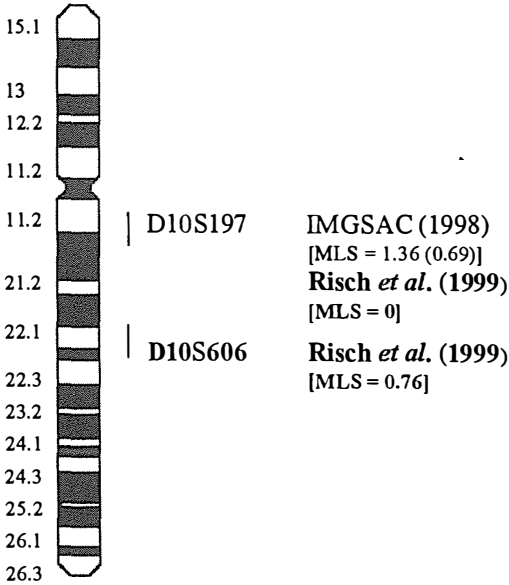
Chromosome 7



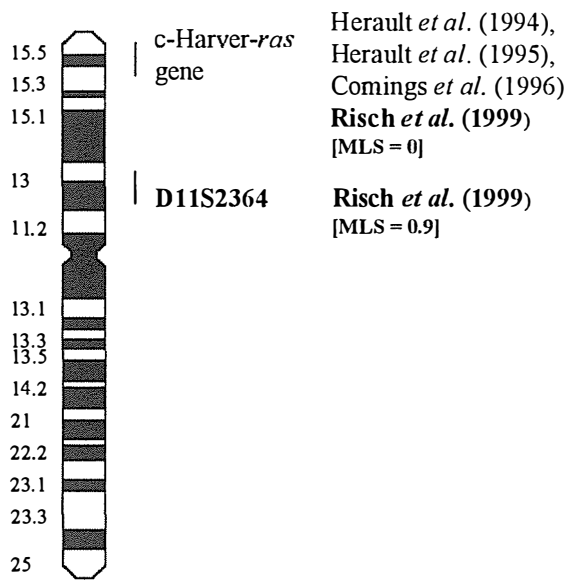
Chromosome 9



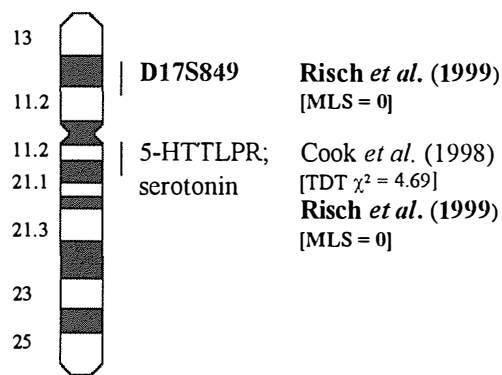
Chromosome 10



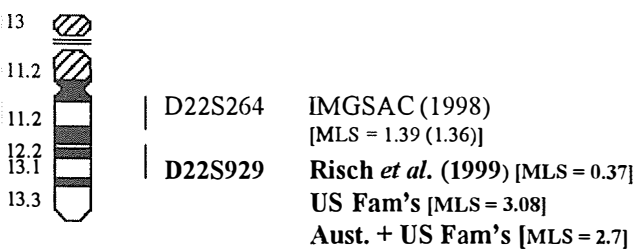
Chromosome 11



Chromosome 17



Chromosome 22



4.4 Reliability of Results

The genome scan conducted by the Stanford and Perth groups was designed to maximise the chance of detecting a gene of moderate to large effect on susceptibility to autism. The important features of the study design include: (a) a statistical method that reduced ascertainment bias; (b) collection of a homogeneous set of families; (c) a map of highly polymorphic markers; (d) use of positive controls throughout the genotyping; and (e) a robust, multipoint model-free method of analysis. Problems do exist, however, with the current linkage maps and the statistical programme used. Whilst the concept and design of this project provide confidence in the validity of our conclusions, these problems should still be considered. The design of this study and the problems that exist are discussed below.

4.4.1 Ascertainment

Historically, it has been assumed that the mode of ascertainment used in obtaining data for linkage studies does not have any effect on the results, provided that individuals were sampled on the basis of trait data and not marker genotypes (Cox Matise *et al.*, 1996). However, studies by Vieland and Hodge (1995, 1996) have shown the effect of ascertainment on maximum likelihood estimates of genetic parameters. These authors claim that the most commonly used ascertainment schemes for linkage studies (non-single ascertainment and proband-dependent sampling, in which the observed pedigree structure depends upon who the probands were) result in asymptotically biased estimates of genetic parameters, for example

the recombination fraction. Although this bias is small in many situations, the calculated maximum likelihood based on the observed data can not be corrected to accommodate the bias. This project was designed to avoid the bias effect shown by Vieland and Hodge (1995, 1996), by utilising a statistical method that did not require the estimation of genetic parameters.

The ethnic origin of the families used could have affected the detection of significant linkage in this study. By restricting the study to individuals likely to be descended from a small founding population, a group of related individuals with a reduced number of genetic and allelic contributions to the disease can be identified. A homogeneous group of strictly defined affected subjects should maximise the chances of detecting linkage for any plausible genetic model (Risch *et al.*, 1999). Our sample contained 88% Caucasian, 5% African or African-U.S.A., 3% Hispanic, and 4% Asian in the U.S.A. sample of families. In the Australian group, the families were primarily Caucasian. Although the majority of subjects included in the genome scan were Caucasian in origin, they can not be classified as a homogeneous population since they are of varied European and Middle Eastern background.

An example of a relatively homogeneous population isolate would be families in the Botnia region of western Finland. This isolated population contains Swedish-speaking inhabitants surrounded by a larger Finnish-speaking population, and was believed to have been settled over 1,000 years ago, with little immigration until at least the middle of the 14th century (Virtaranta-Knowles *et al.*, 1991). This population has been used in search for inherited diseases including non-insulin dependent diabetes mellitus (NIDDM; Mahtani *et al.*, 1996) and diastrophic dysplasia (DTD; Hastbacka *et al.*, 1992).

How the families are grouped together in the analysis also has an effect on the power of the overall study. Whilst it is generally accepted that an individual must inherit more than one gene in order to express the autism phenotype (see section 4.3), there are multiple ways in which these genes could interact to produce the phenotype. For example, the same genes could act together in every instance or they could work together in different combinations to cause autism. In the later scenario, there may be considerable variation from family to family with respect to which genes the affected individual has inherited. Another possibility is that a certain number of ‘core’ genes are required for autism, with several other genes influencing the severity or the expression of the phenotype. This could explain why the severity of autism varies widely from one affected individual to another. It could also explain the variance within families, where two siblings with autism may have significantly different language and cognitive abilities, while their parents or other siblings may have only social constraints or a preference for routines. How the genes interact is important in the analysis of the families. If individuals with different causative genes are grouped together, the overall power to detect one or more of these susceptibility loci is reduced, especially if the loci have a small effect on the disorder ($\lambda_s \leq 2$).

4.4.2 Diagnostic Criteria

Subject misclassification has been proven to affect genetic linkage analysis (Ott, 1991). Problems of phenotypic definition and diagnostic accuracy are particularly important in genetic studies of psychiatric disorders. For this reason, this

study adopted a conservative approach to phenotyping. Good inter-rater reliability was observed for the first 44 multiplex families. The kappa coefficient between our evaluators for autistic vs. non-autistic groups was 0.90. Both ADI and ADOS were used to evaluate the children and any questionable cases were excluded from the study. A total of 45 families were excluded from further analysis. One reason for exclusion of some families was that the second affected sibling did not have a strict diagnosis of autism by agreement of all diagnosticians.

Another reason for the exclusion of a family was low mental age or extremely low IQ scores observed in all affected children. This approach was adopted to avoid the difficulty in diagnosing autism in the very young or in profoundly retarded children. It also helped to reduce potential heterogeneity in the sample. Families with more than one child with both autism and low IQ scores may have distinct aetiologies related to their severe to profound mental retardation. However, one could also argue that, by excluding these families, which could represent the more severe forms of autism caused by genes with a larger effect, the ability to detect susceptibility genes is reduced.

The stringency of the design means that the families in this study included no broadly-defined, questionable, or mild cases (eg. clinical diagnoses of PDD-NOS or Asperger's Syndrome). This was done in order to enhance the chance of detecting linkage by creating a more homogeneous set of families, as well as to gain confidence that any false negative or false positive results obtained would not be due to using too broad a diagnostic spectrum whose genetic basis is still uncertain. As stated previously, a homogeneous group of strictly defined affected subjects should maximise the chances of detecting linkage for any plausible genetic model (Risch *et al.*, 1999). However, some investigators argue strongly for the inclusion of "mild" or

autism “spectrum” cases in genetic studies, which would increase the number of families in the analysis, and in turn the number of informative meioses. However, including individuals with a mild or less strict diagnosis of autism in the analysis, while increasing the power of the study, may also increase the probability of false positive results. The present study was designed to establish a balance between the power to accurately discover true linkage results and the rate of false positive findings.

4.4.3 Linkage Maps

In the design of this type of project, problems with current linkage maps and the distances between microsatellites could limit the detection of genes of moderate to large effect on a disorder. The average spacing between the autosomal markers analysed in stage one of this study was 10cM, assuming a sex-averaged total autosomal map length of 3500cM. Spacing of 20cM has been shown to be more efficient than either a 10 or 40cM in the context of screening for a simple dominant trait using affected relative pairs (Brown, Gorin, and Weeks, 1994). However, for a complex disease, a map density with a mean separation of 10-20cM probably gives the optimal balance between “effort” and power (Hauser *et al.*, 1996). Morton (1996) also found that it is preferable to increase sample size rather than marker density, since applying ever more dense sets of markers to the same cohort of families does not help to narrow potential domains. The effects of increasing marker map density have been explored by other studies, both theoretical (Hauser *et al.*, 1996; Holmans and Craddock, 1997; Chataway *et al.*, 1998) and experimental

(Davies *et al.*, 1996). Using computer simulations, Hauser *et al.* (1996) showed that the most efficient way to type sibling pairs was with a 10-20cM map, with subsequent follow-up of regions of interest. This two stage method was employed in our study.

The design minimises the chance of a missed linkage within the interval between flanking markers. Similarly, a linkage study of diabetes showed that saturation mapping of the 6q region increased the MLS from 2.2 to only 2.4 for a change in marker density from 8 to 2.5cM (Davies *et al.*, 1994). Simulations by our group on 100 families have also shown that, by increasing the map density from 10cM to 5cM, you can only expect to exclude an additional 5% (87% to 92%) of the genome for susceptibility loci of $\lambda_s = 2$ (see Table 7). When 200 families are analysed, increasing the map density from 10cM to 5cM has a minor effect (99% to 99%) on the proportion of the genome that you could expect to exclude. Hauser *et al.* (1996) stated that, unless the actual sibling population was dramatically expanded, a dividend from the effort involved in typing more markers over a region of interest was unlikely to be obtained. They recommended using singletons with parents and additional affected siblings, when possible, for transmission disequilibrium testing.

Errors in the order and distances between markers on a linkage map affect the probability of detecting linkage. Accurate construction of meiotic maps requires the use of correct data in the analysis (Buetow, 1991). In the database used to construct the CEPH consortium map of chromosome 1, errors occurred at a rate of 0.6% for genotypes at the same locus, even after extensive data checking by individual laboratories. In version 4 of the CEPH data for chromosome 4, errors were detected

at a rate of 1.4% for genotypes at the same locus produced by different laboratories (23/817 discordant pairs).

Since more genotypes are necessary for the construction of high-resolution maps, the opportunities for the above mentioned primary errors increase. A typing error that negates a true recombination event or introduces false recombination events could have a large effect on a meiotic map by substantially reducing the accuracy of the resolved map order (Buetow, 1991). Since it is unlikely that even the most rigorous laboratory quality-control procedures will be successful in eliminating all erroneous observations from a data set, high-resolution linkage maps should be used with caution. However it should also be noted that studies by Hauser *et al.* (1996) have shown that multipoint analysis is robust to errors in the map distance between the flanking markers, except when the errors are extreme.

Another problem of current linkage maps is the lack of microsatellite markers near the telomeric and the centromeric regions of chromosomes (Wintero *et al.*, 1992). Whilst minisatellites or variable number of tandem repeat (VNTR) polymorphisms are preferentially clustered at the telomeric regions of chromosomes (Royle *et al.*, 1988) or localised near the centromeres (Willard *et al.*, 1986), microsatellites are mainly distributed throughout the chromosomal arms. To ensure that resolution is sufficiently high to be of use in the identification of polygenic disorders, VNTRs are very important in the finalisation of a genetic map.

4.4.4 Reliability of Genotyping

Errors in genetic linkage data can occur because of errors in the pedigrees or during genotyping. A pedigree error usually involves misidentification of individuals and relationships (Ott, 1991), for example nonpaternity, unidentified adoption, and sample mix-ups. Typing errors include the misinterpretation of genotypes and data-entry errors (Buetow, 1991). Pedigree errors were minimised in this study since the parents were included in the analysis, allowing Mendelian segregation to be checked in all families.

A number of controls were employed in our analysis to identify potential typing errors (see section 3.1.1). The error rate was found to be 1.0% for the entire genome scan and 1.2% for those genotypes analysed as part of this PhD project. This result is comparable to other studies, such as the Levinson *et al.* (1998) schizophrenia genome scan which found results concordant in over 99% of subjects, and the Lathrop *et al.* (1983) study that estimated a typing error frequency of 1% in data obtained from a population study in the south Pacific island, Tokelau. The estimated reliability of these results should provide confidence in their validity.

4.4.5 ASPEX Programme

Another factor that may have an effect on the reliability of our data is the statistical programme utilised. The computer package used throughout this genome scan was the identity-by-descent (IBD) programme of the ASPEX package. This programme uses only unambiguous IBD sharing to compute a multipoint lod score

(maximised over λ_S), as well as an exclusion map along each chromosome. Whilst our analysis of marker DXYS154 located in the pseudoautosomal regions showed the programme to be sufficiently sensitive to detect a significant excess in sharing, when present (see section 3.1.1), there are a few shortcomings with the package that should be noted. Firstly, the number of parents typed per family affects the detection of linkage using the ASPEX-IBD programme. A study by Davis and Weeks (1997) comparing 23 different statistics from a total of 10 popular software packages for model-free linkage, found that when only one parent was typed in a family structure including only two affected siblings, the power to detect linkage dropped by over fifty percent (Figure 33a). When other siblings were included in the analysis, the parental genotypes could be reconstructed and the loss of power was lower (Figure 33b). In this genome scan, 23 out of the 90 U.S.A. autism families had only one parent available for typing. Fifteen of these families had no unaffected siblings that could be used to reconstruct the parental genotypes. In the other 8 families, the majority (six) had only 1 unaffected sibling, which was not always adequate to reconstruct the parental genotypes.

Secondly, the false positive rate for the ASPEX programme generally exceeds the nominal significance level of 0.05 (Figure 34 a and b). In the above mentioned study, Davis and Weeks (1997) found that the highest false positive rates in the ASPEX-IBD programme occurred when a family structure that included four siblings (at least two of which were affected) and one typed parent (at 2.1 unit) was used. Therefore, although the power to detect linkage was lower in families with only 2 affected siblings and one parent typed, the rate of false positive results was less using this family structure.

Davis and Weeks (1997) concluded in their study that the most powerful statistic for affected sibling linkage methods was SIB-PAL sage.asp. This statistic was reasonably powerful for all situations tested in the study, and they agreed with others who suggested that the t_2 test implemented in sage-asp is powerful and theoretically sound (Blackwelder and Elston, 1985; Knapp *et al.*, 1994).

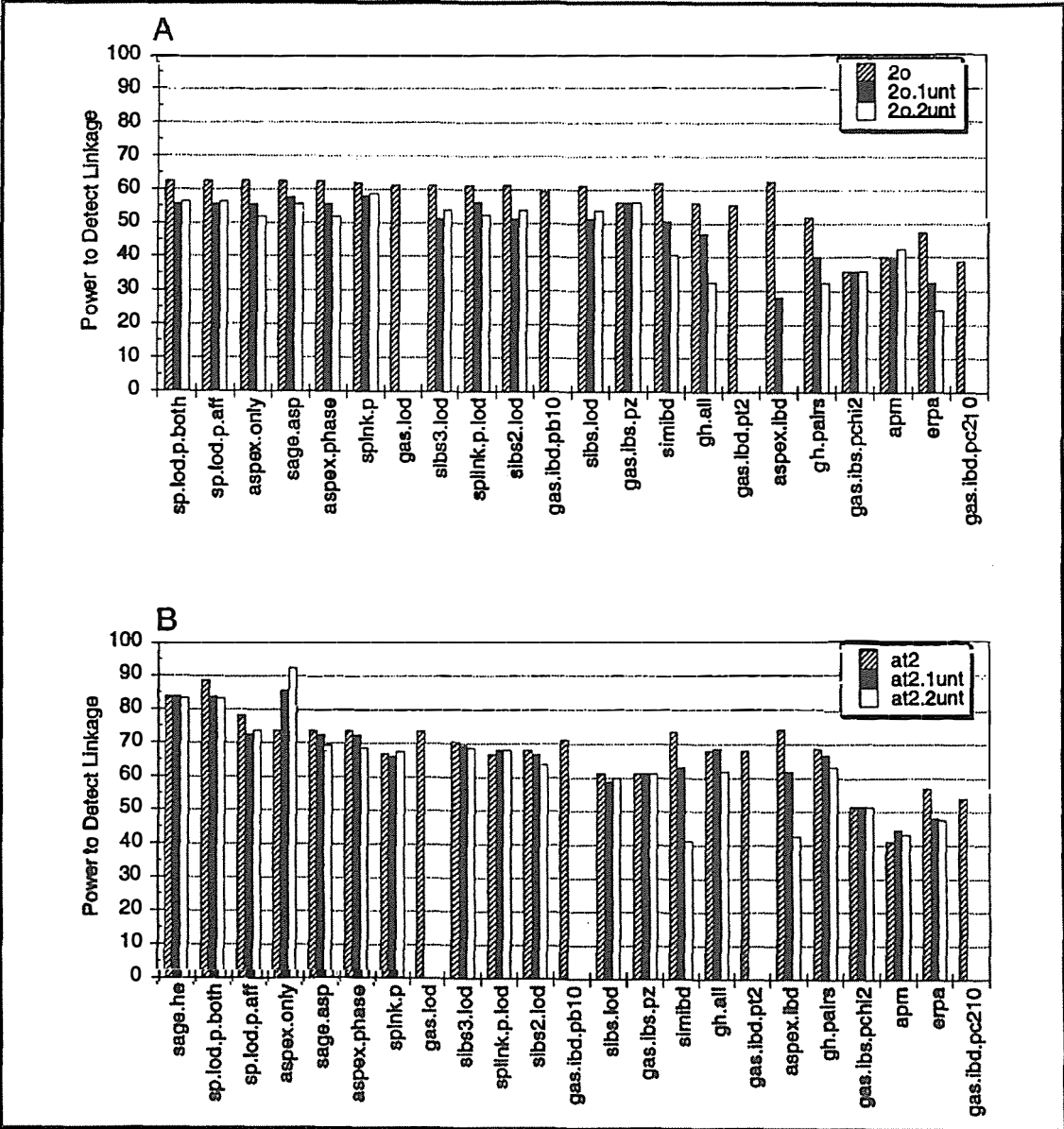


Figure 33: Power to detect linkage. Statistics are ordered in decreasing rank, by power. A, Family structure included two affected siblings and their parents, with zero ("2o"), one ("2o.1unt"), or both ("2o.2unt") parents untyped. B, Family structure included four siblings, at least two of whom were affected, and their parents, with zero ("at2"), one ("at2.1unt"), or both ("at2.2unt") parents untyped.

(Figure from Davis and Weeks, 1997)

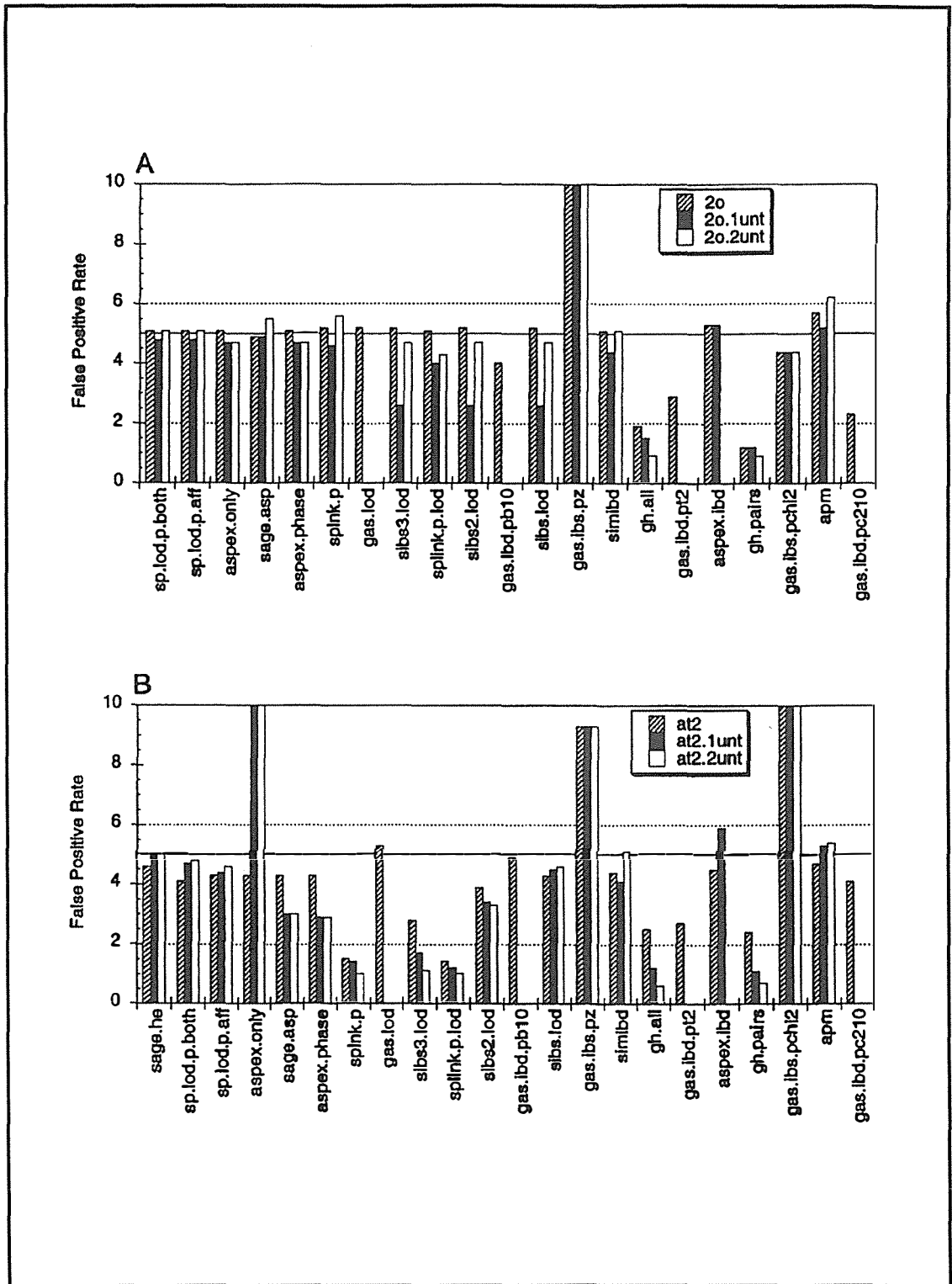


Figure 34: False positive rates. Statistics are ordered in decreasing rank, by power. A, Family structure included two affected siblings and their parents, with zero ("2o"), one ("2o.1unt"), or both ("2o.2unt") parents untyped. B, Family structure included four siblings, at least two of whom were affected, and their parents, with zero ("at2"), one ("at2.1unt"), or both ("at2.2unt") parents untyped.

(Figure from Davis and Weeks, 1997)

4.5 Implications for Future Studies of the Genetics of Autism

In light of our conclusions that autism is caused by multiple genes, each with a small effect on the disorder, other study designs and methods of detection may need to be utilised to detect these genes. In a recent perspective, Risch and Merikangas (1996) and Camp (1997) demonstrated that linkage analysis was likely to be successful only for loci with λ_s values in the range of 4 or larger, but not for loci with λ_s values of two or less. Detection of these small genetic effects using sib-pair analysis could require larger samples than are currently available to any independent research group, perhaps 500-1000 pedigrees (Hauser *et al.*, 1996; Kruglyak and Lander, 1995). This may be achieved if researchers pool their samples and data through large multicentre and multidisciplinary projects. A simultaneous evaluation of all available data with unified methods of analysis (same diagnostic techniques, markers and statistical analyses) may help to eliminate some genetic models for this disease, as well as define the most promising chromosomal regions for follow-up.

Other methods exist that may also assist in the detection of susceptibility genes in autism. Examples include: using an alternative phenotype than the disease itself, examining family structures other than affected sib-pairs and their parents, analysing candidate genes within regions showing susceptibility to autism and/or scanning the entire genome for association of single nucleotide polymorphisms (SNPs) to disease loci. The positive linkage results found by genome scans using sib-pair analysis could be used as candidate regions in these future analyses.

4.5.1 Refining the Phenotype Definition

By choosing a single trait within the illness as the target of genetic research, especially one that is principally accounted for by the effect of a major gene, the deficit should approximate a classical Mendelian pattern of segregation in families and increase the power to detect linkage. This strategy has not been widely exploited in psychiatric research (Freedman *et al.*, 1999). Since autism may result from different influences, some of them genetic and others environmental, each of the genetic traits should have its own biological phenotype that contributes to the pathogenesis of the illness. If this phenotype can be recognised independently from the full clinical syndrome, then the genes underlying the phenotype could be easier to detect and give a clearer understanding of the genetic, neurobiological, and neuropsychological basis of autism and related disorders.

This principle is illustrated by the study of families affected with haemochromatosis. Certain individuals were noted to have increased serum ferritin levels, and it was established that these serum levels were related to the presence of the major gene responsible for the illness (Borecki *et al.*, 1990). A few gene carriers had haemochromatosis, whereas many other gene carriers had only elevated serum ferritin levels. If a similar measurement were available for autism, then its presence could be used as an alternative phenotype for the illness itself.

There are several potential advantages to such alternative phenotypes. Firstly, they greatly increase the power of genetic analysis, because they allow consideration of all members of the study. For autism there are no diagnostic criteria for parents and no clear cut way of establishing whether parents and siblings affected by milder subclinical forms of the disorder are unaffected or have a different degree

of the same condition. In the present statistical models, these family members are considered unknown, compromising the chance of determining the location of abnormal alleles by linkage. By using an alternative phenotype all abnormal members of the pedigree could be identified, increasing the amount of data derived from the pedigree. It has been estimated that hundreds of sib-pair families are needed for sufficient power to find genes related to autism, however if a sub-phenotype definition with, for example, an autosomal dominant inheritance is found, linkage could be detected in as few as 20 families (Freedman *et al.*, 1999).

Secondly, possible biological differences between autistic probands could be detected. If different groups of patients had entirely different genetic predispositions but showed the same clinical phenotype, an alternative phenotype could discriminate between the two groups and select for a biologically more homogeneous set of patients, thus increasing the overall probability of detecting linkage.

One such alternative phenotype in autism could be cognitive impairment. Family studies have shown that the relatives of people with autism are affected by milder subclinical forms of the disorder, with a particularly high rate of cognitive deficits (see section 1.5.2.3). There are four cognitive processes known to be impaired in autism. These include theory of mind, executive function, central coherence, and structural language skills. By examining the patterns of segregation of these deficits in families, DNA markers may be found that link to these impairments. Identifying the genetic basis of these cognitive deficits would greatly assist in the early detection of children in high-risk families, so that early remediation programmes could be implemented. Alternative phenotypes could also provide a clearer understanding of the genetic and neurological basis of autism.

4.5.2 Alternative Family Structures

4.5.2.1 Cousin Pairs

Affected relative pairs that are more distantly related, for example cousins, contain more information on linkage than affected sib pairs if the susceptibility allele is rare (Badner *et al.*, 1998). This is because more distant relatives share fewer genes/markers ($\leq 25\%$) by chance than sibling pairs (50%), and those they do share that are identical by descent (IBD) are likely to be linked to the susceptibility locus. Also, since the power to detect genetic linkage is dependent on both the relationship of the affected pair and the number of pairs in the sample, the number of families needed to achieve the same power as the sib-pair approach is significantly smaller (Teng and Siegmund, 1997).

This approach would be useful in eliminating some of the irrelevant loci located using the sib-pair approach. Unfortunately, since cousin-pairs are quite uncommon, it is unlikely that enough cousin-pair families could be collected to provide sufficient power to detect susceptibility loci in a whole genome scan. What must be kept in mind is the fact that any gene located through this method is generally going to be a single gene with a large effect, since cousins have only a slight chance of sharing all of the genes that may be needed to cause this disorder.

4.5.2.2 Pairs Discordant for Disease

If we assume that the genes contributing to autism have some detectable behavioural manifestation, such as cognitive function (see section 4.5.1), siblings extremely discordant for a specific trait may be used as another approach to eliminate false positive linkage results detected using the sib-pair method. Discordant sib-pairs

are used to detect loci by assuming that allele sharing between these siblings will be significantly less than expected by chance (as opposed to linkage, which assumes the opposite). A discordant sib-pair is chosen by selecting an unaffected sibling who has no identifiable autistic-like social or cognitive traits, and an autistic individual from the same sibship. For such extremely discordant sib-pairs, the chromosomal regions of reduced allele sharing would likely contain the genes that cause autism.

The number of sib-pairs required for this method is far fewer than those needed in methods that rely on pairs concordant for the phenotype (ASP method). Risch and Zhang (1995) have shown that a sample size 10- to 40-fold smaller would be sufficient. This approach has also been shown to be the only method that is uniformly powerful regardless of how the disorder is inherited (with the exception of rare recessives), whether genes are common or rare in the population, and whether there are multiple genes or only one (Folstein *et al.*, 1998).

However, there are some limitations to this approach. Firstly, it was suggested by Folstein *et al.* (1998) that only approximately 35% of siblings are expected to have none of the traits that signal the presence of an autism gene, so many families will need to be screened to obtain a large enough sample size of discordant pairs. Secondly, there may not presently be adequate methods to detect all of the traits and behaviours resulting from genes that contribute to the autism phenotype. For example, some contributing genes may have no behavioural manifestations at all. Therefore, distinguishing between individuals who are concordant or discordant for a trait is not clear-cut in the majority of cases. Unless quantitative trait loci (QTL) can be identified for autism and the extreme phenotypes can be accurately identified, this procedure has limited applications for use in identifying the genes that cause autism.

4.5.2.3 Extended Families

Another alternative family structure includes families with multiple affected individuals. The proportion of extended families with several affected individuals is small for autism. For those that do exist (see Figure 35), autism will generally be caused by one gene. If traditional genetic linkage methods were to be used to identify the causative genes, autism research groups from around the world would have to collaborate to obtain a sufficient number of such families. A new approach has been developed that requires only a few affected individuals from a single extended family to identify susceptibility loci. This approach is called “shared segment analysis” and involves the pooling of DNA from all the affected individuals followed by a search for segments with reduced diversity, ie. where the same marker alleles are shared by the pooled samples (Sheffield *et al.*, 1997). This approach, however, has only been used successfully in locating genes for rare recessive disorders, and as yet has not been used with multigenic diseases.

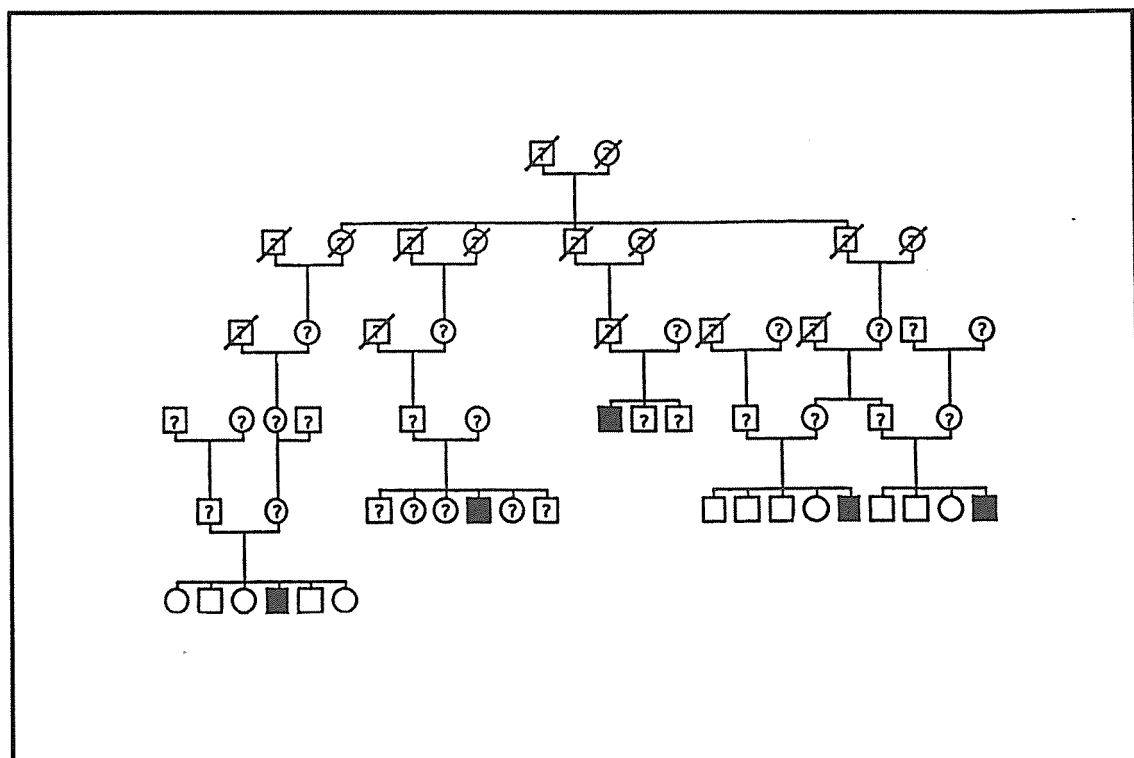


Figure 35: Autistic family with multiple affected individuals

4.5.3 Association Studies

Linkage analysis carries an advantage over association studies in that it extends over greater genetic distances and may therefore be detected using fewer markers. However, association has the advantage that it is considerably more powerful than linkage for genes with a small genetic influence ($\lambda_s < 2$) and therefore requires smaller study populations (Table 32). This is especially the case if there is a marker extremely close to or within the disease locus, and if polymorphisms of functional significance are used (Spielman *et al.*, 1993). For example, Risch and Merikangas (1996) showed that in a genome-wide association study using the transmission-disequilibrium test, 695 parent-child trios provide 80% power to detect an allele with a population frequency of 0.1, assuming that it confers a two-fold increase in disease risk in the heterozygous state (compared to homozygosity for the more common allele). This compares with a requirement of 5382 affected sib pairs for a model-independent linkage study (Table 32). Since trios are easier to ascertain, this difference is enormous.

Table 32: Comparison of Number of Sib-Pairs Required for Linkage vs. Association Studies.

Genotype relative risk (GRR)	Population Frequency of mutation (p)	Linkage (sib-pairs)	TDT (trios)
4	0.01	4260	1098
	0.1	185	150
2	0.01	296,710	5823
	0.1	5382	695
1.5	0.01	4,620,807	19,320
	0.1	67,816	2218

(Risch and Merikangas, 1996)

In autism, genetic studies have been carried out using linkage analysis with no single major locus being found. The suggestions that the genetic basis of the disease is controlled by a number of loci of modest effect indicate that genome-wide scans for linkage will probably be ineffectual. Other approaches, such as the candidate gene approach and/or genome-wide association studies, may be required to detect these susceptibility loci.

4.5.3.1 Candidate Gene Approach

Once a candidate region has been defined and then replicated by another research group, the candidate gene approach could be used to search for sequence variants. The choice of gene(s) that are most likely involved in the predisposition to the disorder among the many genes mapped in that interval (there could be up to several hundred), is based on their hypothetical function, expression pattern and so forth. In the case of autism these genes could be selected on the basis of known or expected role in brain development. More than 30,000 human genes, or Expressed Sequence Tags (ESTs), have been placed relative to a framework map as part of the Human Genome Project (HGP). These maps are very important for the successful identification of susceptibility genes for complex disorders. However, finding the DNA sequence variation that confers risk for a complex disorder, such as autism, will be more complicated than identifying the mutation responsible for a single gene disorder.

Chromosomal rearrangements or null mutations that abolish the function of a protein are unlikely to be found in complex disorders. More subtle sequence variations, such as amino acid substitutions producing a different protein isoform, can be fairly common in the general population and are more likely to represent the

susceptibility alleles in complex disorders. One such allelic sequence variation has been identified to cause susceptibility to the psychiatric disorder Alzheimer disease. This allelic sequence variation is in the apolipoprotein E (APOE) gene, and is referred to as the APOE-E4 allele. Different dosages of this allele have been shown to increase risk and decrease the age of onset in both late-onset familial Alzheimer disease and sporadic Alzheimer disease (Pericak-Vance and Haines, 1995).

In other cases, the susceptibility allele could lie outside the gene-coding region and influence the levels of expression and the regulation of the gene. This has been shown in insulin-dependent diabetes (IDDM) where sequence variation at a minisatellite 5' to the insulin gene influences susceptibility to the disease (Bennett *et al.*, 1995).

Singletons could be used to find the sequence variants in the genes identified as candidates for autism. These families, sometimes referred to as trios when both parents are used in the analysis, contain only one autistic child and are by far the most frequent type of autism family (~97%). Cook *et al.* (1997b) used trios to determine whether two distinct polymorphisms of the serotonin transporter gene (HTT) were associated with autism (see section 1.8.5.8). Genes that regulate serotonin biochemistry are likely candidates for causing autism, due to the inefficient serotonin metabolism and defective brain receptors for serotonin found in autistic individuals (Cook *et al.*, 1997b). A number of genes that have been identified and characterised are involved in this system (Table 33). Those that correspond to regions replicated by multiple groups could be used as candidate genes for further study.

Table 33: Chromosomal Positions of Some Serotonin Receptors

Serotonin Receptor	Chromosome
5-HT _{1A}	5
5-HT _{1B}	6q13
5-HT _{1D}	1p35
5-HT _{1E}	6q14-15
5-HT _{1F}	3p
5-HT _{2A}	13q14.1-14.2
5-HT _{2B}	2q36.3-2q37.1
5-HT _{2C}	Xq21
5-HT ₃	11q23.1-q23.2
5-HT ₄	5q
5-HT _{5A}	7q36.1
5-HT _{5B}	2q11-q13
5-HT ₆	1p35-36
5-HT ₇	10q
5-HTT	17q11.1-q12

(Cook and Leventhal, 1996)

4.5.3.2 Single Nucleotide Polymorphisms

As previously mentioned, association studies are limited in that the gene(s) must be tentatively identified before the test can be performed, and the actual polymorphism within the gene (or one in strong disequilibrium) must be available. Therefore, at present, genome-wide association studies are not feasible. However, with advances in the Human Genome Project (HGP) aiming to identify the majority

of human genes (50,000-100,000) and the simple, biallelic polymorphisms (SNPs) in these genes, Risch and Merikangas (1996) propose that association to genes with λ_s as low as 1.5 could be obtained with a significance level of 5×10^{-8} . SNPs are single nucleotide polymorphisms present throughout the genome in great abundance. These markers can be typed using high-throughput, low-cost automated genotyping technologies, such as DNA chips. SNPs throughout the entire genome can then be scanned for association to disease loci in many families in a fast and effective manner. With new advances in technology, the genes to autism may eventually be identified.

4.6 Conclusions and Outlook

Given the power of this study to reliably detect genes of moderate effect ($\lambda_s > 3.0$, see Table 5), the findings shown here and in the overall genome scan (Risch *et al.*, 1999) provide substantial evidence for the absence of such genes. The data were most consistent with a model of ≥ 15 susceptibility loci, each with a small effect on the disorder. Regions that may contain disease-predisposing loci, however, were identified. These include: regions on chromosomes 1p and 22q. Results from other smaller genome scans also indicate these regions as possible candidates for loci involved in the aetiology of autism. Our results could also not exclude chromosomes 7q, 9q, 10q, 11p and 17, which have previously been suggested to contain susceptibility loci.

These results are important for future studies. In particular, they have shown the requirement for a large sample size to detect genes of small effect. By combining

our data with those from the other linkage studies on autism (IMGSAC, 1998; Philippe *et al.*, 1999) there may be sufficient power to produce a significant linkage result. A simultaneous evaluation of all available data with unified methods of analysis may help to eliminate some genetic models for this disease, as well as define the most promising chromosomal regions for follow-up. Replication of promising loci could be achieved using a number of methods, including samples with different family structures, such as cousin pairs, extremely discordant sibling pairs and extended families. If these approaches fail to detect susceptibility loci, other methods such as the use of alternative phenotypes, linkage disequilibrium studies of candidate genes, and/or genome-wide association studies using SNPs may be required.

With advances in the Human Genome Project, the development of rigorous statistical approaches to the analysis of multifactorial diseases, and the availability of new technologies for the high-throughput assay of genetic markers, the molecular characterisation of the genes that control liability to complex disorders, such as autism, could soon be feasible. With these advances should come insight into the biology and pathology of brain development and behaviour, and improved diagnosis and management of the childhood disorder autism.

CHAPTER 5

BIBLIOGRAPHY

- Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, *et al.* (1993). Clues to the pathogenesis of familial colorectal cancer. *Science*, 260:812-816.
- Ahlsen G, Gillberg C, Lindblom R, Gillberg C (1994). Tuberous sclerosis in western Sweden: a population study of cases with early childhood onset. *Archives of Neurology*, 51:76-81.
- Ahlsen G, Rosengren L, Belgrave M, Palm A, Haglid K, Hamberger A, Gillberg C (1993). Glial fibrillary acidic (GFA) protein in the cerebrospinal fluid of children with autism and other neuropsychiatry disorders. *Biological Psychiatry*, 33:734-743.
- Akefeldt A, Gillberg C (1991). Hypomelanosis of Ito in three cases with autism and autistic-like conditions. *Developmental Medicine and Child Neurology*, 33:737-743.
- American Psychiatric Association (1952). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, Washington, Dc.
- American Psychiatric Association (1968). *Diagnostic and Statistical Manual of Mental Disorders, Second Edition*. American Psychiatric Association, Washington, DC.
- American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. American Psychiatric Association, Washington, DC.
- American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*. American Psychiatric Association, Washington, DC.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. American Psychiatric Association, Washington, DC.
- Anderson MA, Gusella JF (1984). Use of cyclosporine A in establishing Epstein-Barr virus transformed human lymphoblastoid cell lines. *In Vitro*, 20:856-858.
- Arrieta I, Lobato MN, Martinez B, Criado B (1994). Parental origin of Robertsonian translocation (15q22q) and Pader Willi syndrome associated with autism. *Psychiatric Genetics*, 4:63-65.
- Arvidsson T, Danielsson B, Forsberg P, Gillberg C, Johansson M, Kallgren G (1997). Autism in 3-6 year-old children in a suburb of Göteborg, Sweden. *Autism*, 1:163-173.
- Asperger H (1944). Die autistischen Psychopathen im Kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten*, 117:76-136.
- Ausubel AM, Brent R, Kingston RE, Smith JA, Seidmann JG, Struhl D (1991). *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.
- Bachevalier J (1994). Medial temporal lobes structures and autism: A review of clinical and experimental findings. *Neuropsychologia*, 32:627-648.

- Badner JA, Gershon ES, Goldin LR (1998). Optimal ascertainment strategies to detect linkage to common disease alleles. *American Journal of Human Genetics*, 63:880-888.
- Bailey A, Bolton P, Butler L, Le Couteur A, Murphy M, Scott S, Webb T, Rutter M (1993a). Prevalence of the fragile X anomaly amongst autistic twins and singletons. *Journal of Child Psychology and Psychiatry*, 34:673-688.
- 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. *Psychological Medicine*, 25:63-77.
- Bailey A, Luthert P, Bolton P, LeCouteur A, Rutter M (1993b). Autism and megalencephaly. *Lancet*, 34:1225-1226.
- Bailey A, Phillips W, Rutter M (1996). Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *Journal of Child Psychology and Psychiatry*, 37:89-126.
- Bailey AJ (1993). The biology of autism. *Psychological Medicine*, 23:7-11.
- Baker P, Piven J, Schwartz S, Patil S (1994). Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. *Journal of Autism and Developmental Disorders*, 24:529-535.
- Barker D, Wright E, Nguyen K, Cannon L, Fain P, Goldgar D, Bishop DT, Carey J, B B, Kivlin J, Willard H, Wayne JS, Greig G, Leinwand L, Nakamura Y, O'Connell P, Leppert M, Lalouel JM, White R, Skolnick M (1987). Gene for vonRecklinghausen neurofibromatosis is in the pericentric region of chromosome 17. *Science*, 236:1100-1102.
- Baron-Cohen S, Leslie A, Frith U (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21:37-46.
- Bartak L, Rutter M, Cox A (1975). A comparative study of infantile autism and specific developmental receptive language disorder. I. The children. *British Journal of Psychiatry*, 126:127-145.
- Bass MP, Wolpert CM, Menold MM, Donnelly SL, Ravan SA, Hauser ER, Church TD, Slotterbeck B, Curtis AM, Zimmerman A, Gilbert JR, Wright HH, Abramson RK, DeLong GR, Vance JM, Cuccaro ML, Pericak-Vance MA (1998). *American Journal of Human Genetics*, 63(Suppl):A281.
- Bauman ML (1996). Brief report: neuroanatomic observations of the brain in pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 26:199-203.
- Bene E (1958). A Rorschach investigation into the mothers of autistic children. *British Journal of Medical Psychology*, 38:226-227.
- Bennett ST, Lucassen AM, Gough SCL, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Dawaguchi Y, Dronsfield MJ, Pociot F, Nerup J, Bouzekri N, Cambon-Thomsen A, Ronningen KS, Barnett AH, Bain SC, Todd JA (1995). Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature Genetics*, 9:284-292.

- Bentler PM (1989). *EQS Structural Equations Program Manual*. BMDP Statistical Software, Los Angeles.
- Bernard GA (1949). Statistical inference. *Journal of the Royal Statistical Society*, B11:115-39.
- Bettelheim B (1967). *The Empty Fortress - Infantile Autism and the Birth of the Self*. The Free Press, Collier - MacMillan, New York.
- Blackwelder WC, Elston RC (1985). A comparison of sib-pair linkage tests for disease susceptibility loci. *Genetic Epidemiology*, 2:85-98.
- Bleeker-Wagemakers LM, Gal A, Kumar-Singh R, Ingeborgh van den Born L, Li Y, Schwinger E, Sandkuijl LA, Bergen AA, Kenna P, Humphries P, *et al.* (1992). Evidence for nonallelic genetic heterogeneity in autosomal recessive retinitis pigmentosa. *Genomics*, 14:811-812.
- Bleuler E, Brill A (1924). *Textbook of Psychiatry*. Macmillan Publishing Co. Inc., New York.
- Bohman M, Bohman IL, Bjork P, Sjöholm E (1983). *Childhood Psychosis in a Northern Swedish Country: Some Preliminary Findings from an Epidemiological Survey*. Stuttgart, Georg Thieme.
- 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. *Journal of Child Psychology and Psychiatry*, 35:877-900.
- Bolton P, Rutter M (1990). Genetic influences in autism. *International Review of Psychiatry*, 2:67-80.
- Borecki IB, Rao DC, Yaouanq J, Lalouel JM (1990). Serum ferritin as a marker of affection for genetic hemochromatosis. *Human Heredity*, 40:159-66.
- Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32:314-331.
- Brask BH (1972). *A Prevalence Investigation of Childhood Psychosis*. Barnepsykiatrisk forening: Nordic Symposium on the Comprehensive Care of the Psychotic Children. Oslo, Norway.
- Bregman JD, Volkmar FR (1988). Autistic social dysfunction and Down syndrome. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27:440-441.
- Bristol MM, Cohen DJ, Costello EJ, Denckla M, Eckberg TJ, Kallen R, Kraemer HC, Lord C, Maurer R, McIlvane WJ, Minshew N, Sigman M, Spence MA (1996). State of the science of autism: report to the National Institutes of Health. *Journal of Autism and Developmental Disorders*, 26:121-154.
- Brown DL, Gorin MB, Weeks DE (1994). Efficient strategies for genomic searching using the affected-pedigree-member method of linkage analysis. *American Journal of Human Genetics*, 54:544-552.

- Brown MA, Pile KD, Kennedy LG, Campbell D, Andrew L, March R, Shatford JL, Weeks DE, Calin A, Wordsworth BP (1998). A genome-wide screen for susceptibility loci in ankylosing spondylitis. *Arthritis and Rheumatism*, 41:588-595.
- Bryson SE, Clark BS, Smith IM (1988). First report of a Canadian epidemiological study of autistic syndromes. *Journal of Child Psychology and Psychiatry*, 29:433-445.
- Buetow KH (1991). Influence of aberrant observations on high-resolution linkage analysis outcomes. *American Journal of Human Genetics*, 49:985-994.
- Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema, Duyk GM, Sheffield VC, Wang Z, Murray JC (1994). Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genetics*, 6:391-393.
- Bundey S, Hardy C, Vickers S, Kilpatrick MW, Corbett JA (1994). Duplication of the 15q11-13 in two individuals with autism, epilepsy and ataxia. *Developmental Medicine and Child Neurology*, 36:736-742.
- Burd L, Fisher W, Kerbeshian J (1987). A prevalence study of pervasive developmental disorders in North Dakota. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26:704-710.
- Burd L, Martsolf JT, Kerbeshian J, Jalal SM (1988). Partial 6p trisomy associated with infantile autism. *Clinical Genetics*, 33:356-359.
- Camp NJ (1997). Genomewide transmission/disequilibrium testing-consideration of the genotypic relative risks at disease loci. *American Journal of Human Genetics*, 61:1424-1430.
- Cantwell DP, Baker L, Rutter M (1978). *Family Factors*. In: Rutter M, Schopler E (eds), *Autism: A Reappraisal of Concepts and Treatment*. Plenum Press, New York, :pp 269-296.
- Carrano AV, Lamerdin J, Ashworth LK, Watkins B, Branscomb E, Slezak T, Raff M, Dejong PJ, Keith D, McBride L, Meister S, Kronick M (1989). A high-resolution fluorescence-based, semiautomated method for DNA fingerprinting. *Genomics*, 4:129-136.
- Chan CTJ, Clayton-Smith J, Chong XI, Buxton J, Webb T, Pembrey ME, Malcolm S (1993). Molecular mechanisms in Angelman syndrome: A survey of 93 patients. *Journal of Medical Genetics*, 30:895-902.
- Chapman CJ (1993-1996). *Cyrillic for Pedigree Drawing*. Cherwell Scientific Publishing, Ltd., Great Britain.
- Chataway J, Feakes R, Coraddu F, Gray J, Deans J, Fraser M, Robertson N, Broadley S, Jones H, Clayton D, Goodfellow P, Sawcer S, Compston A (1998). The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome scan. *Brain*, 121:1869-1887.
- Chen CH, Hsiao KJ (1989). A Chinese classic phenylketonuria manifested as autism. *British Journal of Psychiatry*, 155:251-253.
- Chess S (1971). Autism in children with congenital rubella. *Journal of Autism and Child Schizophrenia*, 1:33-47.

- Chess S (1977). Follow-up report on autism in congenital rubella. *Journal of Autism and Developmental Disorders*, 7:79-81.
- Cialdella P, Mamelie N (1989). An epidemiological study of infantile autism in a French department (Rhône): a research note. *Journal of Child Psychology and Psychiatry*, 30:165-175.
- Clerget-Darpoux F, Babron MC, Bonaiti-Pellie C (1990). Assessing the effect of multiple linkage tests in complex diseases. *Genetic Epidemiology*, 7:245-253.
- Coleman M, Blass JP (1985). Autism and lactic acidosis. *Journal of Autism and Developmental Disorders*, 15:1-8.
- Collins A, Frezal J, Teague J, Morton NE (1996). A metric map of humans: 23,000 loci in 850 bands. *Proceedings of the National Academy of Science, USA*, 93:14771-5.
- Collins FS (1995). Positional cloning moves from perditional to traditional. *Nature Genetics*, 9:347-350.
- Comings DE, Chiu C, Muhleman D, Sverd J (1996). Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Research*, 63:25-32.
- Connell C, Fung S, Heiner C, Bridgham J, Chakerian V, Heron E, Jones B, Menchen S, Mordan W, Raff M, Recknor M, Smith L, Springer J, Woo S, Hunkapiller M (1987). Automated DNA sequence analysis. *Biotechniques*, 5:342-348.
- Cook EH (1990). Autism: review of neurochemical investigation. *Synapse*, 6:292-308.
- Cook EH, 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. *American Journal of Human Genetics*, 62:1077-1083.
- Cook EH, Leventhal BL (1996). The serotonin system in autism. *Current Opinion in Pediatrics*, 8:348-354.
- Cook EHJ, Courchesne R, Lord C, Cox NJ, Yan S, Lincoln A, Haas R, Courchesne E, Leventhal BL (1997b). Evidence of linkage between the serotonin transporter and autistic disorder. *Molecular Psychiatry*, 2:247-250.
- Cook EHJ, Lindgren V, Leventhal BL, Courchene R, Lincoln A, C, Lord C, Courchesne E (1997a). Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *American Journal of Human Genetics*, 60:928-934.
- Cooper DN, Krawczak M (1993). *Human Gene Mutation*. BIOS Scientific Publishers Limited, United Kingdom, :pp 25-30.
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, 261:921-923.
- Courchesne E, Hesselink JR, Jernigan TL, Yeung-Courchesne R (1987). Abnormal neuroanatomy in a nonretarded person with autism. Unusual findings with magnetic resonance imaging. *Archives of Neurology*, 44:335-341.

- Courchesne E, Saitoh O, Yeung-Courchesne R, Press GA, Lincoln AJ, Haas RH, Schreibman L (1994). Abnormality of cerebellar vermal lobules VI and VII in patients with infantile autism: identification of hypoplastic and hyperplastic subgroups with MR imaging. *American Journal of Roentgenology*, 164:123-130.
- Courchesne E, Yeung-Courchesne R, Press G, Hesselink JR, Jernigan TL (1988). Hypoplasia of cerebellar vermal lobules VI and VII in autism. *New England Journal of Medicine*, 318:1349-1354.
- Couturier J, Delattre O, Kujas M, Philippon J, Peter M, Rouleau G, Aurias A, Thomas G (1990). Assessment of chromosome 22 anomalies in neurofibromas by combined karyotype and RFLP analysis. *Cancer Genetics and Cytogenetics*, 45:55-62.
- Cox Matisa T, Donis-Keller H, Ott J (1996). Statistical methods in genetic mapping. *Genomics*, 36:223-225.
- Cox Matisa T, Perlin M, Chakravarti A (1994). Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genetics*, 6:384-390.
- Critchley M, Earl CJC (1932). Tubero sclerosis and allied conditions. *Brain*, 55:311-346.
- Crolla J, Harvey J, Sitch F, Dennis N (1995). Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. *Human Genetics*, 95:161-170.
- Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, Wilson AJ, Cordell HJ, Bain SC, Todd JA (1998). A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nature Genetics*, 19:301-302.
- Cumings JN, Shorman RC, Skrbic T (1965). Lipid studies in the blood and brain in multiple sclerosis and motor neurone disease. *Journal of Clinical Pathology*, 18:641-644.
- Daniels WW, Warren RP, Odell JD, Maciulis A, Burger RA, Warren WL, Torres AR (1995). Increased frequency of the extended or ancestral haplotype B44-SC30-DR4 in autism. *Biological Psychiatry*, 32:120-123.
- Davidovitch M, Patterson B, Gartside P (1996). Head circumference measurements in children with autism. *Journal of Child Neurology*, 11:389-393.
- Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, *et al.* (1994). A genome-wide search for human type 1 diabetes susceptibility genes. *Nature*, 371:130-136.
- Davies K (1991). Breaking the fragile X. *Nature*, 351:439-440.
- Davies S, Schroeder M, Goldin LR, Weeks DE (1996). Nonparametric simulation-based statistics for detecting linkage in general pedigrees. *American Journal of Human Genetics*, 58:867-880.
- Davis S, Weeks DE (1997). Comparison of nonparametric statistics for detection of linkage in nuclear families: Single-marker evaluation. *American Journal of Human Genetics*, 61:1431-1444.

- Dawson G (1996). Brief report: neuropsychology of autism: a report on the state of the science. *Journal of Autism and Developmental Disorders*, 26:179-184.
- DeLong GR (1992). Autism, amnesia, hippocampus, and learning. *Neuroscience and Biobehavioral Reviews*, 16:63-70.
- DeLong RG, Dwyer JT (1988). Correlation of family history with specific autistic subgroups: Asperger's syndrome and bipolar affective disorders. *Journal of Autism and Developmental Disorders*, 18:593-600.
- Despert J (1956). Some considerations relating to the genesis of autistic behavior in children. *American Journal of Orthopsychiatry*, 12:366-371.
- Deykin EY, MacMahon B (1979). Viral exposure and autism. *American Journal of Epidemiology*, 109:628-38.
- Deykin EY, MacMahon B (1980). Pregnancy, delivery and neonatal complications among autistic children. *American Journal of Disabled Children*, 134:860-864.
- Donis-Keller H, Barker DF, Knowlton RG, Schumm JW, Braman JC, Green P (1986). Highly polymorphic RFLP probes as diagnostic tools. *Cold Spring Harbor Symposia on Quantitative Biology*, 51:317-324.
- Ebers GC, Bulman DE, Sadovnick AD, Paty DW, Warren S, Hader W, Murray TJ, Seland TP, Duquette P, Grey T. *et al.* (1986). A population-based study of multiple sclerosis in twins. *New England Journal of Medicine*, 315:1638-1642.
- Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, Armstrong H, Cousin K, Bell RB, Hader W, Paty DW, Hashimoto S, Oger J, Duquette P, Warren S, Gray T, O'Connor P, Nath A, Auty A, Metz L, Francis G, Paulseth JE, Murray TJ, Pryse-Phillips W, Nelson R, Freedman M, Brunet D, Bouchard J-P, Hinds D, Risch N (1996). A full genome search in multiple sclerosis. *Nature Genetics*, 13:472-476.
- Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12:241-253.
- Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, Deriel JK, Forget BG, Weissman SM, Slighton JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980). The structure and evolution of the human beta-globin gene family. *Cell*, 21:653-668.
- Ehlers S, Gilberg C (1993). The epidemiology of Asperger syndrome. A total population study. *Journal of Child Psychology and Psychiatry*, 34:1327-1350.
- Eldridge R, Denckla MB, Bien E, Myers S, Kaiser-Kupfer MJ, Pikus A, Schlesinger SL, Dilys MP, Dambroisa JM, Zasloff MA, Mulvihill JJ (1989). Neurofibromatosis Type1 (Recklinghausen's disease). Neurologic and cognitive assessment with sibling controls. *American Journal of Diseases of Children*, 143:833-837.
- Elston RC (1992). Segregation and linkage analysis. *Animal Genetics*, 23:59-62.
- Erlich HA, Gelfand D, Sninsky JJ (1991). Recent advances in the polymerase chain reaction. *Science*, 252:1643-1651.

- Evans DGR, Huson SM, Donnai D. *et al.* (1992). A clinical study of type 2 neurofibromatosis. *Quarterly Journal of Medicine*, 84:603-618.
- Fahsold R, Rott HD, Claussen U, Schmalenberger B (1991). Tuberous sclerosis in a child with de novo translocation t(3,12) (p26.3,q23.3). *Clinical Genetics*, 40:326-328.
- Falk CT, Rubinstein P (1987). Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Annals of Human Genetics*, 51:227-233.
- Filipek P (1996). Brief report: neuroimaging in autism: the state of the science 1995. *Journal of Autism and Developmental Disorders*, 26:211-215.
- Finnegan JA, Quarrington B (1979). Pre-, peri-, and neonatal factors and infantile autism. *Journal of Child Psychology and Psychiatry*, 20:119-128.
- Fisher RA (1921). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, A202:309-68.
- Flejter WL, Bennet-Baker PE, Ghazziuddin M, McDonald M, Sheldon S, Gorski JL (1996). Cytogenetic and molecular analysis of inv dupl(15) chromosomes observed in two patients with autistic disorder and mental retardation. *American Journal of Medical Genetics*, 61:182-187.
- Folling A (1934). Phenylpyruvic acid as a metabolic anomaly in connection with imbecility. *Zeitschrift für Physiologische Chemie*, 224:169-176.
- Folstein SE, Bisson E, Santangelo SL, Piven J (1998). Finding specific genes that cause autism: a combination of approaches will be needed to maximize power. *Journal of Autism and Developmental Disorders*, 28:439-445.
- Folstein S, Rutter M (1977). Infantile autism: a genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry*, 18:297-321.
- Folstein SE, Piven J (1991). Etiology of autism: Genetic influences. *Pediatrics*, 87:767-773.
- Folstein SE, Rutter ML (1988). Autism: familial aggregation and genetic implications. *Journal of Autism and Developmental Disorders*, 18:3-30.
- Fombonne E (1992). Diagnostic assessment in a sample of autistic and developmentally impaired adolescents. *Journal of Autism and Developmental Disorders*, 22:563-581.
- Fombonne E (1998). *Epidemiological Surveys of Autism*. In: Volkmar F (ed), *Autism and Pervasive Developmental Disorders*. Cambridge University Press, Cambridge, England.
- Fombonne E (1999). The epidemiology of autism: a review. *Psychological Medicine*, 29:769-786.
- Fombonne E, du Mazaubrun C (1992). Prevalence of infantile autism in four French regions. *Social Psychiatry and Psychiatric Epidemiology*, 27:203-210.
- Fombonne E, Du Mazaubrun C, Cans C, Grandjean H (1997). Autism and associated medical disorders in a French epidemiological survey. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36:1561-1569.

- Freedman R, Adler LE, Leonard S (1999). Alternative phenotypes for the complex genetics of schizophrenia. *Biological Psychiatry*, 45:551-558.
- Friedman E (1969). The 'autistic syndrome' and phenylketonuria. *Schizophrenia*, 1:249-261.
- Frith U (1989). *Autism and 'Theory of Mind'*. In: Gillberg C (ed), *Diagnosis and Treatment of Autism*. Plenum Press, New York, pp 33-52.
- Frith U (1991). *Autism and Asperger Syndrome*. Cambridge University Press, Cambridge.
- Fujimoto A, Reddy KS, Spinks R (1998). Interstitial deletion of chromosome 4, del(4)(q12q21.1), in a mentally retarded boy with a piebald trait, due to maternal insertion, ins (8;4). *American Journal of Medical Genetics*, 75:78-81.
- Fulker DW, Cherny SS, Cardon LR (1995). Multipoint interval mapping of quantitative trait loci, using sib pairs. *American Journal of Human Genetics*, 56:1224-1233.
- Gaffney GR, Tsai LY (1987). Magnetic resonance imaging of high level autism. *Journal of Autism and Developmental Disorders*, 17:433-438.
- Gaffney GR, Tsai LY, Kuperman S, Minchin S (1987). Cerebellar structure in autism. *American Journal of Diseases in Children*, 141:1330-1332.
- Garber HJ, Ritvo ER (1992). Magnetic resonance imaging of the posterior fossa in autistic adults. *American Journal of Psychiatry*, 149:245-247.
- Gath A, Gumley D (1986). Behaviour problems in retarded children with special reference to Down's Syndrome. *British Journal of Psychiatry*, 149:156-161.
- Ghaziuddin M (1997). Autism in Down's syndrome: family history correlates. *Journal of Intellectual and Disability Research*, 41:87-91.
- Ghaziuddin M, Tsai LY, Eilers L, Ghaziuddin N (1992a). Brief report: autism and herpes simplex encephalitis. *Journal of Autism and Developmental Disorders*, 22:107-113.
- Ghaziuddin M, Tsai LY, Ghaziuddin N (1992b). Autism in Down's syndrome: presentation and diagnosis. *Journal of Intellectual and Disability Research*, 36:449-456.
- Gillberg C (1984). Infantile autism and other childhood psychoses in a Swedish urban region. Epidemiological aspects. *Journal of Child Psychology and Psychiatry*, 25:35-43.
- Gillberg C (1989). Asperger syndrome in 23 Swedish children. *Developmental Medicine and Child Neurology*, 31:520-531.
- Gillberg C (1990). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 31:99-119.
- Gillberg C (1991). The treatment of epilepsy in autism. *Journal of Autism and Developmental Disorders*, 21:61-77.
- Gillberg C (1992a). Subgroups in autism: are their behavioural phenotypes typical of underlying medical conditions? *Journal of Intellectual Disability Research*, 36:201-214.

- Gillberg C (1993). Autism and related behaviours. *Journal of Intellectual and Disability Research*, 37:343-372.
- Gillberg C (1995). *The Prevalence of Autism and Autism Spectrum Disorders*. In: Koot FCVaHM (ed), *The Epidemiology of Child and Adolescent Psychopathology*, vol 11. Oxford University Press, Oxford, pp 227-257.
- Gillberg C (1998). Chromosomal disorders and autism. *Journal of Autism and Developmental Disorders*, 28:415-425.
- Gillberg C, Coleman M (1992). *The Biology of the Autistic Syndromes*, 2nd edn. MacKeith Press, London.
- Gillberg C, Coleman M (1996). Autism and medical disorders: a review of the literature. *Developmental Medicine and Child Neurology*, 38:191-202.
- Gillberg C, Forsell C (1984). Childhood psychosis and neurofibromatosis--more than a coincidence. *Journal of Autism and Developmental Disorders*, 14:1-8.
- Gillberg C, Gillberg IC (1983). Infantile autism: a total population study of reduced optimality in the pre-, peri- and neonatal period. *Journal of Autism and Developmental Disorders*, 13:153-166.
- Gillberg C, Gillberg IC, Steffenburg S (1992b). Siblings and parents of children with autism: a controlled population-based study. *Developmental Medicine and Child Neurology*, 34:389-398.
- Gillberg C, Persson E, Grufman M, Themner U (1986). Psychiatric disorders in mildly and severely mentally retarded urban children and adolescents: epidemiological aspects. *British Journal of Psychiatry*, 149:68-74.
- Gillberg C, Schaumann H, Steffenburg S (1991a). Is autism more common now than 10 years ago? *British Journal of Psychiatry*, 158:403-409.
- Gillberg C, Steffenburg S, Jakobsson G (1987). Children with Kanner-type autism or Asperger syndrome. *Developmental Medicine and Child Neurology*, 29:641-649.
- Gillberg C, Steffenburg S, Wahlstrom J, Gillberg IC, Sjostedt A, Martinsson T, Liedgren S, Eeg-Olofsson O (1991b). Autism associated with marker chromosome. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30:489-494.
- Gillberg C, Svennerholm L (1987). CSF monoamines in autistic syndromes and other pervasive developmental disorders of early childhood. *British Journal of Psychiatry*, 151:89-94.
- Gillberg CL (1992b). The Emanuel Miller Memorial Lecture 1991: Autism and autistic-like conditions: subclasses among disorders of empathy. *Journal of Child Psychology and Psychiatry*, 33:813-842.
- Gillberg IC, Gillberg C, Ahlsen G (1994). Autistic behaviour and attention deficits in tuberous sclerosis: A population-based study. *Developmental Medicine and Child Neurology*, 36:50-56.
- Gillberg IC, Gillberg C, Kopp S (1992a). Hypothyroidism and autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 33:531-542.

- Goldgar DE (1990). Multipoint analysis of human quantitative genetic variation. *American Journal of Human Genetics*, 47:957-967.
- Gomez MR (1988). *Tuberous Sclerosis (2nd ed.)*. Raven Press, New York.
- Gordon CT, State RC, Nelson JE, Hamburger SD, Rapoport JL (1993). A double-blind comparison of clomipramine, desipramine, and placebo in the treatment of autistic disorder. *Archives of Genetic Psychiatry*, 50:441-447.
- Gurling H (1986). Candidate genes and favored loci: Strategies for molecular genetic research into schizophrenia, manic depression, autism, alcoholism, and Alzheimer disease. *Psychiatric Diagnosis*, 4:289-309.
- Gutierriz GC, Smalley SL, Tanguay PE (1998). Autism in tuberous sclerosis complex. *Journal of Autism and Developmental Disorders*, 28:97-103.
- Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994). The 1993-94 Genethon human genetic linkage map. *Nature Genetics*, 7:246-339.
- Hagerman RJ (1989). *Chromosomes, Genes and Autism*. In: Gillberg C (ed), *Diagnosis and Treatment of Autism*. Plenum Press, New York, pp 105-132.
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 250:1684-1689.
- Hallmayer J, Pintado E, Lotspeich L, Spiker D, McMahon W, Petersen PB, Nicholas P, Pingree C, Kraemer HC, Wong DL, Ritvo E, Lin A, Hebert J, Cavalli-Sforza LL, Ciaranello RD (1994). Molecular analysis and test of linkage between the FMR-1 gene and infantile autism in multiplex families. *American Journal of Human Genetics*, 55:951-959.
- Hallmayer J, Spiker D, Lotspeich L, McMahon WH, Petersen PB, Nicholas P, Pingree C, Ciaranello RD (1996). Male-to-male transmission in extended pedigrees with multiple cases of autism. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 67:13-18.
- Happe FGE (1994a). Annotation: current psychological theories of autism. The "Theory of Mind" account and rival theories. *Journal of Child Psychology and Psychiatry*, 35:215-229.
- Happe F (1994b). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24:129-154.
- Hashimoto T, Tayama M, Murakawa K, Yoshimoto t, Miyazaki M, Harada M, Kuroda Y (1995). Development of the brainstem and cerebellum in autistic patients. *Journal of Autism and Developmental Disorders*, 25:1-18.
- Haslam J (1809). *Observation of Madness and Melancholy*. Hayden, London.
- Hastbacka J, Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics*, 2:204-211.

- Hauser ER, Boehnke M, Guo S-W, Risch N (1996). Affected sib pair interval mapping and exclusion for complex genetic traits: Sampling considerations. *Genetic Epidemiology*, 13:117-137.
- Hearne CM, Ghosh S, Todd JA (1992). Microsatellites for linkage analysis of genetic traits. *Trends in Genetics*, 8:288-294.
- Herauld J, Petit E, Buchler M, Martineau j, Cherpi C, Perrot a, Sauvage D, Barthelemy C, Pierre Muh J, Lelord G (1994). Lack of association between three genetic markers of brain growth factors and infantile autism. *Society of Biological Psychiatry*, 35:281-283.
- Herauld J, Petit E, Martineau J, Perrot A, Lenoir P, Cherpi C, Barthelemy C, Sauvage D, Mallet J, Muh JP, Lelord G (1995). Autism and genetics: Clinical approach and association study with two markers of HRAS gene. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 60:276-281.
- Hetzler BF, Griffin JL (1981). Infantile autism and the temporal lobe of the brain. *Journal of Autism and Developmental Disorders*, 11:317-330.
- Hinds D, Risch N (1996). *The ASPEX package: affected sib-pair mapping*. <ftp://lahmed.stanford.edu/pub/aspex>.
- Holmans P, Craddock N (1997). Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. *American Journal of Human Genetics*, 60:657-666.
- Holtttum JR, Minshew NJ, Sanders RS, Phillips NE (1992). Magnetic resonance imaging of the posterior fossa in autism. *Biological Psychiatry*, 32:1091-1101.
- Honda H, Shimizu Y, Misumi K, Niimi M, Ohashi Y (1996). Cumulative incidence and prevalence of childhood autism in children in Japan. *British Journal of Psychiatry*, 169:228-235.
- Horwitz B, Rumsey JM (1994). *Positron Emission Tomography: Implications for Cerebral Dysfunction in Autism*. In: Bauman ML, Kemper TL (eds), *The Neurobiology of Autism*. The Johns Hopkins University Press, Baltimore, :pp 102-118.
- Hoshino Y, Kumashiro H, Yashima Y, Tachibana R, Watanabe M (1982). The epidemiological study of autism in Fukushima-ken. *Folia Psychiatry Psychiatrica Neurologica Japonica*, 36:115-124.
- Hotopf M, Bolton P (1995). A case of autism associated with partial tetrasomy 15. *Journal of Autism and Developmental Disorders*, 25:41-49.
- Housman D (1995). Molecular medicine: Human DNA polymorphism. *The New England Journal of Medicine*, 332:318-320.
- Howlin P, Wing L, Gould J (1995). The recognition of autism in children with Down syndrome--implications for intervention and some speculations about pathology. *Developmental Medicine and Child Neurology*, 37:406-414.
- Hunt A (1995). Gaining new understanding of tuberous sclerosis. *Nursing Times*, 91:31-33.

- Hunt A, Shepherd C (1993). A prevalence study of autism in tuberous sclerosis. *Journal of Autism and Developmental Disorders*, 23:323-339.
- International Molecular Genetic Study of Autism Consortium (1998). A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Human Molecular Genetics*, 7:571-578.
- Ishii T, Takahashi O. *Epidemiology of Autistic Children in Toyota City, Japan*. World Child Psychiatry Conference, Dublin 1982. Unpublished.
- Jeffreys AJ, Wilson V, Thein SL, Weatherall DJ, Ponder BAJ (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics*, 39:11-24.
- Jones MB, Szatmari P (1988). Stoppage rules and genetic studies of autism. *Journal of Autism and Developmental Disorders*, 18:31-40.
- Jorde LB, Carey JC, White RL (1995). *Medical Genetics*. Mosby-Year Book Inc., USA.
- Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, McMahon WM, Petersen B, Jensen WR, Mo A (1991). Complex segregation analysis of autism. *American Journal of Human Genetics*, 49:932-938.
- Kan YW, Dozy AM (1978). Polymorphism of DNA sequence adjacent to the human B-globin structural gene: relationship to sickle mutation. *Proceedings of the National Academy of Science USA*, 75:5631-5635.
- Kanner L (1943). Autistic disturbances of affective contact. *Nervous Child*, 2:217-250.
- Kanner L, Eisenberg L (1956). Early infantile autism 1943-1955. *American Journal of Orthopsychiatry*, 26:55-65.
- Kanter WR, Eldridge R, Fabricant R, Allen JC, Koerber T (1980). Central neurofibromatosis with bilateral acoustic neuroma. Genetic, clinical and biochemical distinctions from peripheral neurofibromatosis. *Neurology*, 30:851-859.
- Kendler K (1990). Toward a scientific psychiatric nosology. *Archives of General Psychiatry*, 47:969-973.
- Kerbeshian J, Burd L, Randall T, Marstolf J, Jalal S (1990). Autism. Profound mental retardation and atypical bipolar disorders in a 33-year-old female with a deletion of 15q12. *Journal of Mental Deficiency Research*, 34:205-210.
- Klauck SM, Poustka F, Benner A, Lesch KP, Poustka A (1997). Serotonin transporter (5-HTT) gene variants associated with autism? *Human Molecular Genetics*, 6:2233-2238.
- Kleiman MD, Neff S, Rosman NP (1992). The brain in infantile autism: are posterior fossa structures abnormal? *Neurology*, 42:753-760.
- Knapp M, Seuchter S, Baur M (1994). Linkage analysis in nuclear families. 1. Optimality criteria for affected sib-pair tests. *Human Heredity*, 44:37-43.
- Knobloch H, Pasamanick B (1975). Some etiologic and prognostic factors in early infantile autism and psychosis. *Pediatrics*, 55:182-191.

- Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988). Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American Journal of Human Genetics*, 43:520-526.
- Kraepelin E, Diefendorf A (1923). *Clinical Psychiatry: A Textbook for Students and Physicians*. Macmillan Publishing Co Inc., New York.
- Krause U, Faisst G, Cordes U, Beyer J (1979). Secretory B-cell activity in insulin dependent maturity-onset diabetes. *Hormone Metabolism Research*, 11:261-265.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *American Journal of Human Genetics*, 58:1347-1363.
- Kruglyak L, Lander ES (1995). High-resolution genetic mapping of complex traits. *American Journal of Human Genetics*, 56:1212-1223.
- Kumar-Singh R, Kenna PF, Farrar GJ, Humphries P (1993). Evidence for further genetic heterogeneity in autosomal dominant retinitis pigmentosa. *Genomics*, 15:212-215.
- Lainhart JE (1997). Developmental abnormalities in autism. *Lancet*, 349:373-374.
- Lainhart JE, Piven J (1995). Diagnosis, treatment, and neurobiology of autism in children. *Current Opinion in Pediatrics*, 7:392-400.
- Lalwani AK, Brister JR, Fex J, Grundfast KM, Pikus AT, Ploplis B, San Agustin T, Skarka H, Wilcox ER (1994). A locus for Waardenburg syndrome type II maps to chromosome 1p13.3-2.1. *American Journal of Human Genetics*, 55:A14.
- Landa R, Wzorek MM, Piven J, *et al.* (1989). *Discourse Characteristics of Adult First Degree Relatives of Autistic Children*. Presented at the First World Congress on Psychiatric Genetics, Churchill College, Cambridge, England, August 3-5.
- Lander E, Kruglyak L (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, 11:241-247.
- Lander ES, Schork NJ (1994). Genetic dissection of complex traits. *Science*, 265:2037-2047.
- Landgren M, Pettersson R, Kjellman B, Gillberg C (1996). ADHD, DAMP and other neurodevelopmental/neuropsychiatric disorders in six-year-old children. Epidemiology and comorbidity. *Developmental Medicine and Child Neurology*, 38:891-906.
- Lane H (1976). *The Wild Boy of Aveyron*. Harvard University Press, Cambridge, Mass.
- Lathrop GM, Hooper AB, Huntsman JW, Ward RH (1983). Evaluating pedigree data. I. The estimation of pedigree error in the presence of marker mistyping. *American Journal of Human Genetics*, 35:241-262.
- Lathrop GM, Lalouel JM, Julier C, Ott J (1985). Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *American Journal of Human Genetics*, 37:482-98.

- Lathrop GM, Lalouel LM, Julier C, Ott J (1984). Strategies for multilocus linkage analysis in humans. *Proceedings of the National Academy of Science USA*, 81:3443-46.
- Lauritsen M, Mors O, Mortensen PB, Ewald H (1999). Infantile autism and associated autosomal chromosome abnormalities: A register-based study and a literature survey. *Journal of Child Psychology and Psychiatry*, 40:335-345.
- Lawrence J (1977). *Rheumatism in Populations*. William Heinemann Medical Books, London.
- Le Couteur A, J BA, Rutter M, Gottesman I (1989a). *An Epidemiology Based Twin Study of Autism*. Presented at the First World Congress on Psychiatric Genetics, Churchill College, Cambridge, England, August 3-5.
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan L (1989b). Autism diagnostic interview: a standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19:363-387.
- Leana-Cox J, Jenkins L, Palmer CG, Plattner R, Sheppard L, Flejter WL, Zackowski J. *et al.* (1994). Molecular cytogenetic analysis of inv dup(15) chromosomes, using probes specific for the Prader-Willi/Angelman syndrome region: clinical implications. *American Journal of Human Genetics*, 54:748-756.
- Levinson G, Gutman GA (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4:203-221.
- Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, Hayward NK, Crowe RR, Andreasen NC, Black DW, Silverman JM, Endicott J, Sharpe L, Mohs RC, Siever LJ, Walters MK, Lennon DP, Jones HL, Nertney DA, Daly MJ, Gladis M, Mowry BJ (1998). Genome scan of schizophrenia. *American Journal of Psychiatry*, 155:741-750.
- Levy S, Zoltak B, Saelens T (1988). A comparison of obstetrical records of autistic and nonautistic referrals for psychoeducational evaluations. *Journal of Autism and Developmental Disorders*, 18:573-581.
- Lopreiato JO, Wulfsberg EA (1992). A complex chromosome rearrangement in a boy with autism. *Journal of Developmental Behavioural Pediatrics*, 13:281-283.
- Lord C, Mulloy C, Wendelboe M, Schopler E (1991). Pre- and perinatal factors in high-functioning females and males with autism. *Journal of Autism and Developmental Disorders*, 21:197-209.
- Lord C, Rutter M, Godde S, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989). Autism diagnostic observation schedule: a standardized observation of communication and social behavior. *Journal of Autism and Developmental Disorders*, 19:185-212.
- 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. *Journal of Autism and Developmental Disorders*, 24:659-685.
- Lotter V (1966). Epidemiology of autistic conditions in young children. *Social Psychiatry*, 1:124-137.

- Macdonald H, Rutter M, Rios P, Bolton P (1989). *Cognitive and Social Abnormalities in the Siblings of Autistic and Down's Syndrome Probands*. Presented at the First World Congress on Psychiatric Genetics, Churchill College, Cambridge, England, August 3-5.
- Maestrini E, Marlow AJ, Weeks DE, Monaco AP (1998). Molecular genetic investigations of autism. *Journal of Autism and Developmental Disorders*, 28:427-437.
- Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnelly K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC (1996). Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nature Genetics*, 14:90-94.
- Markowitz PL (1983). Autism in a child with congenital cytomegalovirus infection. *Journal of Autism and Developmental Disorders*, 13:245-254.
- Martinsson T, Johannesson T, Vujic M, Sjostedt A, Steffenburg S, Gillberg C, Wahlstrom J (1994). Maternal origin of a supernumerary isodicentric chromosome 15-marker in infantile autism. *European Child and Adolescent Psychiatry*, 5:185-192.
- Martinsson T, Johannesson T, Vujic M, Sjostedt A, Steffenburg S, Gillberg C, Wahlstrom J (1996). Maternal origin of inv dup(15) chromosomes in infantile autism. *European Child and Adolescent Psychiatry*, 5:185-192.
- Martuz RL, Eldridge R (1988). Neurofibromatosis 2 (bilateral acoustic neurofibromatosis). *New England Journal of Medicine*, 318:684-688.
- Matsuishi T, Shiotsuki Y, Yoshimura K, Shoji H, Imuta F, Yamashita F (1987). High prevalence of infantile autism in Kurume City, Japan. *Journal of Child Neurology*, 2:268-271.
- McCarthy, Fitzgerald M, Smith MA (1984). Prevalence of childhood autism in Ireland. *Irish Medical Journal*, 77:129-130.
- McDougle CJ, Naylor ST, Cohen DJ, Volkmar FR, Heninger GR, Price LH (1996). A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder. *Archives in Genetic Psychiatry*, 53:1001-1008.
- McEvoy RE, Rogers SJ, Pennington BF (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry*, 34:563-578.
- Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Pritchard LE, Cucca F, Barnett AH, Bain SC, Todd JA (1998). A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nature Genetics*, 19:297-300.
- Meldrum BS, Bruton CJ (1992). *Epilepsy*. In: Adams JH, Duchan LW (eds), *Greenfield's Neuropathology* (5th edn). Edward Arnold, London, pp 1246-1283.
- Menold MM, Martin ER, Wolpert CM, Bass MP, Donnelly SL, Ravan SA, Zimmerman A, Vance JM, Maddox LO, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Gilbert JR, Pericak-Vance MA (1998). Failure to replicate linkage disequilibrium

- between marker GABRB3 155CA-2 and autistic disorder. *The American Journal of Human Genetics*, 63(Suppl):A300.
- Microsoft (1985-1996). *Microsoft Excel for Windows 95* (7th edn.). Microsoft Corporation, USA.
- Miladi N, Lamaout A, Kaabachi N, Helayem M, Ben Hamida M (1992). Phenylketonuria: an underlying etiology of autistic syndrome. A case report. *Journal of Child Neurology*, 7:22-23.
- Miller SA, Dykes DD, Polesky HF (1988). A simple salting-out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16:1215.
- Minshew NJ (1994). *In vivo Brain Chemistry of Autism*. In: Bauman ML, Kemper TL (eds), *The Neurobiology of Autism*. The Johns Hopkins University Press, Baltimore, :pp 86-101.
- Morton N (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, 7:277-318.
- Morton NE (1996). Logarithm of odds (lods) for linkage in complex inheritance. *Proceedings of the National Academy of Science USA*, 93:3471-3476.
- Morton N (1998). Significance levels in complex inheritance. *American Journal of Human Genetics*, 62:690-697.
- Mouridsen SE, Andersen LB, Sorensen SA, Rich B, Isager T (1992). Neurofibromatosis in infantile autism and other types of childhood psychoses. *Acta Paedopsychiatry*, 55:15-18.
- Mullis KB, Falcona FA (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155:335-350.
- Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA (1994). The British Isles survey of multiple sclerosis in twins. *Neurology*, 44:11-15.
- Murayama K, Greenwood RS, Rao KW, Aylsworth AS (1991). Neurological aspects of del(1q) syndrome. *American Journal of Medical Genetics*, 15:488-492.
- Murray JC, Buetow KW, Weber JL, Ludwigsen S, Scherphier-Haddema T, Marion F, Quillen J, Sheffield VC, Sunden S, Duyk G. *et al.* (1994). A comprehensive human linkage map with centimorgan density. *Science*, 265:2049-2054.
- Myers BA, Pueschel SM (1991). Psychiatric disorders in persons with Down Syndrome. *Journal of Nervous and Mental Disease*, 179:609-613.
- NIH/CEPH Collaborative Mapping Group (1992). A comprehensive genetic linkage map of the human genome. *Science*, 258:67-86.
- Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, Bertheas MF, Mandel JL (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, 252:1097-1102.
- Olson JM (1995). Multipoint linkage analysis using sib pairs: An interval mapping approach using dichotomous outcomes. *American Journal of Human Genetics*, 56:788-798.

- Olsson I, Steffenburg S, Gillberg C (1988). Epilepsy in autism and autistic-like conditions: a population-based study. *Archives of Neurology*, 45:666-668.
- Ornitz EM (1978). *Biological Homogeneity or Heterogeneity*. In: Rutter M, Schopler E (eds), *Autism: A Reappraisal of Concepts and Treatment*. Plenum Press, New York, pp 243-250.
- Ott J (1991). *Analysis of Human Genetic Linkage*. The Johns Hopkins University Press, Baltimore, London.
- Ott J (1992). *Analysis of Human Genetic Linkage* (2nd edn.). The Johns Hopkins University Press, Baltimore, London.
- Ozonoff S, Pennington B, Rogers S (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, 32:1081-1105.
- PE Applied Biosystems (1993-96). *Genotyper 2.0 Software*. Perkin-Elmer Corporation, USA.
- PE Applied Biosystems (1994). *ABI PRISM™ Linkage Mapping Set*. Perkin-Elmer Corporation, USA.
- PE Applied Biosystems (1996). *310 GeneScan™ Analysis 2.1, Macintosh and Apple 3.1.1*. Perkin-Elmer Corporation, USA.
- Penrose LS (1935). The detection of autosomal linkage in data which consist of pairs of brothers and sisters or unspecified parentage. *Annals of Eugenics*, 6:133-38.
- Pericak-Vance MA, Haines JL (1995). Genetic susceptibility to Alzheimer disease. *Trends in Genetics*, 11:504-508.
- Pericak-Vance MA, Wolpert CM, Menold MM, Bass MP, DeLong GR, Beaty LM, Zimmerman A, Potter N, Gilbert JR, Vance JM, Wright HH, Abramson RK, Cuccaro ML (1997). Linkage evidence supports the involvement of chromosome 15 in autistic disorder (AUT). *American Journal of Human Genetics*, 61(Suppl):A40.
- Pericak-Vance MA, Wolpert CM, Menold MM, Bass MP, Hauser ER, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh M, Aylsworth AS, Vance JM, DeLong GR, Wright HH, Abramson RK, Gilbert JG, Cuccaro ML (1998). Chromosome 7 and autistic disorder (AD). *The American Journal of Human Genetics*, 63(Suppl):A16.
- Perkin-Elmer (1996). *377 DNA Sequencer User's Manual*. Applied Biosystems Inc (a division of Perkin Elmer Corporation), USA.
- Petit E, Herault J, Raynaud M, Cherpi C, Perrot A, Barthelemy C, Lelord G, Muh J-P (1996). X chromosome and infantile autism. *Biological Psychiatry*, 40:457-464.
- Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, van Malldergerme L, Penet C, Feingold J, Brice A, Leboyer M (1999). Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. *Human Molecular Genetics*, 8:805-812.
- 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. *American Journal of Human Genetics*, 57:717-726.
- Piven J (1997). The biological basis of autism. *Current Opinion in Neurobiology*, 7:708-712.
- Piven J, Arndt S, Bailey J, Havercamp S, Andreasen NC, Palmer P (1995). A MRI study of brain size in autism. *American Journal of Psychiatry*, 152:1145-1149.
- Piven J, Berthier ML, Starkstein SE, Nehme E, Pearlson G, Folstein S (1990). Magnetic resonance imaging evidence for a defect of cerebral cortical development in autism. *American Journal of Psychiatry*, 147:734-739.
- Piven J, Chase GA, Landa R, Wzorek M, Gayle J, Cloud D, Folstein S (1991b). Psychiatric disorders in the parents of autistic individuals. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30:471-478.
- Piven J, Gayle BS, Landa R, Wzorek M, Folstein S (1991a). The prevalence of fragile X in a sample of autistic individuals diagnosed using a standardized interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, 3:825-830.
- Piven J, Gayle J, Chase GA, Fink B, Landa R, Wzorek MM, Folstein SE (1989a). A family history study of neuropsychiatric disorders in the adult siblings of autistic individuals. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29:177-183.
- Piven J, Landa R, Wzorek M. *et al.* (1989b) *Personality Characteristics and Psychiatric Disorders in the Relatives of Autistic Individuals*. Presented at the First World Congress on Psychiatric Genetics, Churchill College, Cambridge, England August, 3-5.
- Piven J, Nehme E, Simon J, Barta P, Pearlson G, Folstein SE (1992). Magnetic resonance imaging in autism: measurement of the cerebellum, pons, and fourth ventricle. *Biological Psychiatry*, 31:491-504.
- Piven J, Simon J, Chase GA, Wzorek M, Landa R, Gayle J, Folstein S (1993). The etiology of autism: pre-, peri- and neonatal factors. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32:1256-1263.
- Plioply AV, Greaves A, Kazemi K, Silverman E (1994). Immunoglobulin reactivity in autism and Rett's syndrome. *Developmental Brain Dysmorphology*, 7:12-16.
- Podolin PL, Denny P, Armitage N, Lord CJ, Hill NJ, Levy ER, Peterson LB, Todd JA, Wicker LS, Lyons PA (1998). Localization of two insulin-dependent diabetes (Idd) genes to the Idd10 region on mouse chromosome 3. *Mammal Genome*, 9:283-286.
- Pueschel SE, Herman R, Groden G (1985). Brief report: Screening children with autism for fragile X syndrome and phenylketonuria. *Journal of Autism and Developmental Disorders*, 15:335-343.
- Ramos FJ, Emanuel BS, Spinner NB (1989). Expression of a common fragile site at Xq27.2 under conditions of thymidylate stress in patients referred for fragile-X analysis and controls. *American Journal of Human Genetics*, Abstract 1.218.
- Raymond GV, Bauman ML, Kemper TL (1996). Hippocampus in autism: a Golgi analysis. *Acta Neuropathology*, 91:117-119.

- Riikonen R, Simell O (1990). Tuberous sclerosis and infantile autism spasms. *Developmental Medicine and Child Neurology*, 32:203-209.
- Rineer S, Finucane B, Simon EW (1998). Autistic symptoms among children and young adults with isodicentric chromosome 15. *American Journal of Medical Genetics*, 81:428-433.
- Risch N (1990a). Linkage strategies for genetically complex traits. I. Multilocus models. *American Journal of Human Genetics*, 46:222-228.
- Risch N (1990b). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *American Journal of Human Genetics*, 46:229-241.
- Risch N (1993). Exclusion mapping for complex diseases. *American Journal of Human Genetics*, 53 (Suppl):A185.
- Risch N (unpublished). Personal Communication.
- Risch N, Botstein D (1996). A manic depressive history. *Nature Genetics*, 12:351-353.
- Risch N, Claus E, Giuffra L (1989). *Linkage and Mode of Inheritance in Complex Traits*. In: Elston RC, Spence MA, Hodge SE, MacCluer JW (eds), Progress in clinical and biological research (vol 329): Multipoint mapping and linkage based upon affected pedigree members. Liss, New York, pp 183-188.
- Risch N, Ghosh S, Todd JA (1993). Statistical evaluation of multiple locus linkage data in experimental species and relevance to human studies: application to murine and human IDDM. *American Journal of Human Genetics*, 53:702-714.
- Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science*, 273:1516-1517.
- Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, McCague P, Dimiceli S, Pitts T, Nguyen L, Yang J, Harper C, Thorpe D, Vermeer S, Young H, Hebert J, Lin A, Ferguson J, Chiotti C, Wiese-Slater S, Rogers T, Salmon B, Nicholas P, Peterson PB, Pingree C, McMahon W, Wong DL, Cavalli-Sforza LL, Kraemer HC, Myers RM (1999). A genomic screen of autism: evidence for a multilocus etiology. *American Journal of Human Genetics*, 65:493-507.
- Risch N, Zhang H (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*, 268:1584-1589.
- Ritvo ER, Freeman BJ, Mason-Brother A, Mo A, Ritvo AM (1985a). Concordance for the syndrome of autism in 40 pairs of afflicted twins. *American Journal of Psychiatry*, 142:74-77.
- Ritvo ER, Jorde LB, Mason-Brothers A, Freeman BJ, Pingree C, Jones MB, McMahon WM, Peterson PB, Jenson WR, Mo A (1989). The UCLA-University of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counseling. *American Journal of Psychiatry*, 146:1032-1036.
- Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, Jenson WR, McMahon WM, Petersen PB, Jorde LB, Mo A, Ritvo A (1990). The UCLA-University of Utah epidemiologic survey of autism: the etiologic role of rare diseases. *American Journal of Psychiatry*, 147:1614-1621.

- Ritvo ER, Spence MA, Feeman BJ, Mason-Brothers A, Mo A, Marazita ML (1985b). Evidence for autosomal recessive inheritance in 46 families with multiple incidences of autism. *American Journal of Psychiatry*, 142:187-192.
- Robinson W, Binkert F, Gine R, Vazquez C, Muller W, Rosenkranz W, Schinzel A (1993). Clinical and molecular analysis of five inv dup(15) patients. *European Journal of Human Genetics*, 1:37-50.
- Rodier PM, Ingram JL, Tisdale B, Nelson MS, Roman J (1996). Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. *Journal of Comparative Neurology*, 370:247-261.
- Rogers T, Kalaydjieva L, Hallmayer J, Petersen PB, Nicholas P, Pingree C, McMahon W, Spiker D, Lotspeich L, Kraemer H, McCague P, Dimicelli S, Nouri N, Peachy T, Yang J, Hinds D, Risch N, Myers RM (1999). Exclusion of linkage to the HLA region in ninety multiplex sibships with autism. *Journal of Autism and Developmental Disorders*, 29:195-201.
- Rouleau GA, Merel P, Lutchman M, Sanson M, Zucman J, Marineau C, Hoang-Xuan K, Demczuk S, Desmaze C, Plougastel B, Pulst SM, Lenoir G, Bijlsma E, Fashold RF, Dumanski J, Jong PD, Parry D, Eldrige R, Aurias A, Delattre O, Thomas G (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature*, 363:515-521.
- Rouleau GA, Seizinger BR, Wertelecki W, Haines JL, Superneau DW, Martuza RL, Gusella JF (1989). Precise mapping of the neurofibromatosis type 2 locus on chromosome 22. *Cytogenetics and Cell Genetics*, 51:1070.
- Rouleau GA, Wertelecki W, Haines JL, Hobbs WJ, Trofatter JA, Seizinger BR, Martuza RL, Superneau DW, Conneally PM, Gusella JF (1987). Genetic linkage of bilateral acoustic neurofibromatosis to a DNA marker on chromosome 22. *Nature*, 329:246-248.
- Rousseau F, Heitz D, Biancalana V, Blumenfeld S, Kretz C, Boue J, Tommerup N, Van Der Hagen C, DeLozier-Blanchet C, Croquette MF, Gilgenkrantz S, Jalbert P, Voelckel MA, Oberle I, Mandel JL (1991). Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *New England Journal of Medicine*, 325:1673-1681.
- Royle NJ, Clarkson Re, Wong Z, Jeffreys AJ (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics*, 3:352-360.
- Rutter M (1978). *Diagnosis and Definition*. In: Rutter M, Schopler E (eds), *A Reappraisal of Concepts and Treatment*. Plenum Press, New York :pp 1-25.
- Rutter M (1979). *Language, Cognition and Autism*. Raven, New York.
- Rutter M (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry*, 24:513-531.
- Rutter M, Bailey A, Bolton P, Le Couteur A (1994). Autism and known medical conditions: myth and substance. *Journal of Child Psychology and Psychiatry*, 35:311-322.
- Rutter M, Macdonald H, Couteur AL, Harrington R, Bolton P, Bailey A (1990). Genetic factors in child psychiatric disorders-II. Empirical findings. *Journal of Child Psychology and Psychiatry*, 31:39-83.

- Sadovnick AD, Baird PA, Ward RH (1988). Multiple Sclerosis: Updated risks for relatives. *American Journal of Medical Genetics*, 29:533-541.
- Sahley TL, Panksepp J (1987). Brain opioids and autism - an updated analysis of possible linkages. *Journal of Autism and Developmental Disorders*, 17:201-216.
- Saiki R, Scharf S, Falcona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of B-Globin genomic sequences and restriction site analysis for diagnosis of Sickle Cell Anemia. *Science*, 230:1350.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239:487-491.
- Salmon B, Hallmayer J, Rogers T, Kalaydjieva L, Petersen PB, Nicholas P, Pingree C, McMahon W, Spiker D, Lotspeich L, Kraemer H, McCague P, Dimicelli 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. *American Journal of Medical Genetics*, 88:551-556.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: a Laboratory Manual* (2nd edn). Cold Spring Harbor Laboratory Press.
- Sapolsky RJ, Hsie L, Berno A, Ghandour G, Mittmann M, Fan JB (1999). High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genetic Analyst*, 14:187-92.
- Schinzel A (1990). Autistic disorder and additional inv dup(15)(pter-q13) chromosome. *American Journal of Medical Genetics*, 35:447-448.
- Schinzel A, Brecevic L, Bernasconi F, Binkert F, Berthet F, Wuilloud A, Robinson WP (1994). Intrachromosomal triplication of 15q11-q13. *Journal of Medical Genetics*, 31:798-803.
- Schork NJ, Weder AB, Schork MA, Bassett DR, Julius S (1990). Disease entities, mixed multi-normal distributions, and the role of the hyperkinetic state in the pathogenesis of hypertension. *Statistical Medicine*, 9:301-314.
- Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D, Stevenson RE (1998). Autism and maternally derived aberrations of chromosome 15q. *American Journal of Medical Genetics*, 76:327-336.
- Schwab SG, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein RP, Ackenheil M, Lerer B, Risch N (1995). Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nature Genetics*, 11:325-327.
- Seizinger BR, Martuza RL, Gusella JF (1986). Loss of genes on chromosome 22 in tumorigenesis of human acoustic neuroma. *Nature*, 322:644-647.
- Seizinger BR, Rouleau G, Lane AH, Ozelius LJ, Faryniarz AG, Iannazzi J, Hobbs W, Roy JC, Falcone B, Huson S, et al. (1987). DNA linkage analysis in Von Recklinghausen neurofibromatosis. *Journal of Medical Genetics*, 24:529-530.

- Sheffield VC, Pierpoint ME, Nishimura D, Beck JS, Burns TL, Berg MA, Stone EM, Patil SR, Lauer RM (1997). Identification of a complex congenital heart defect susceptibility locus by using DNA pooling and shared segment analysis. *Human Molecular Genetics*, 60: 979-992.
- Singh VK, Warren RP, Odell JD, Warren WL, Cole P (1993). Antibodies to myelin basic protein in children with autistic behaviour. *Brain, Behavior and Immunology*, 7:97-103.
- Skuse D, 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*, 387:705-708.
- Smalley SL (1991). Genetic influences in autism. *Psychiatric Clinics of North America*, 1:125-139.
- Smalley SL (1997). Genetic influences in childhood-onset psychiatric disorders: autism and attention-deficit/hyperactivity disorder. *American Journal of Human Genetics*, 60: 1276-1282.
- Smalley SL, Asarnow RF, Spence MA (1988). Autism and genetics. *Archives of General Psychiatry*, 45:953-961.
- Smalley SL, Tanguay PE, Smith M, Gutierrez G (1992). Autism and tuberous sclerosis. *Journal of Autism and Developmental Disorders*, 22:339-355.
- Spence MA, Ritvo ER, Marazita ML, Funderburk SJ, Sparkes RS, Freeman BJ (1985). Gene mapping studies with the syndrome of autism. *Behavior Genetics*, 15:1-13.
- Spielman RS, McGinnis RE, Ewens WJ (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, 52:506-516.
- Spiker DK, Lotspeich L, Kraemer HC, Hallmayer J, McMahon W, Peterson PB, Nicholas P, Pingree C, Wiese-Slater S, Chiotti C, Wong DL, Dimicelli S, Ritvo E, Cavalli-Sforza LL, Ciaranello RD (1994). The genetics of autism: characteristics of affected and unaffected children from 37 multiplex families. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 54:27-35.
- Steffenburg S (1991). Neuropsychiatric assessment of children with autism: a population based study. *Developmental Medicine and Child Neurology*, 33:495-511.
- Steffenburg S, Gillberg C (1986). Autism and autistic-like conditions in Swedish rural and urban areas: a population study. *British Journal of Psychiatry*, 149:81-87.
- 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. *Journal of Child Psychology and Psychiatry*, 30:583-600.
- Stein Z, Susser M (1981). Methods of epidemiology. *Journal of the American Academy of Child Psychiatry*, 20:444-461.
- Steinhausen HC, Gobel D, Breinlinger M, Wohlleben B (1986). A community survey of infantile autism. *Journal of the American Academy of Child Psychiatry*, 25:186-189.

- Stubbs EG (1978). Autistic symptoms in a child with congenital cytomegalovirus infection. *Journal of Autism and Child Schizophrenia*, 8:37-43.
- Stubbs EG, Ash E, Williams CPS (1984). Autism and congenital cytomegalovirus. *Journal of Autism and Developmental Disorders*, 14:183-189.
- Stubbs EG, Crawford ML (1977). Depressed lymphocyte responsiveness in autistic children. *Journal of Autism and Child Schizophrenia*, 7:49-55.
- Suarez BK, Hampe CL, Van Eerdewegh P (1994). *Problems of Replication Linkage Claims in Psychiatry*. In: Gershon ES, Cloninger CR (eds), *Genetic Approaches to Mental Disorders*. American Psychiatric Press, Washington, DC, pp 23-46.
- Sugiyama T, Abe T (1989). The prevalence of autism in Nagoya, Japan: a total population study. *Journal of Autism and Developmental Disorders*, 19:87-96.
- Sutherland GR, Baker E (1990). The common fragile site in band q27 of the human X chromosome is not coincident with the fragile X. *Clinical Genetics*, 37:167-172.
- Szatmari P, Jones MB (1991). IQ and the genetics of autism. *Journal of Child Psychology and Psychiatry*, 32:897-908.
- Szatmari P, Jones MB, Fisman S, Tuff L, Bartolucci G, Mahoney WJ, Bryson SE (1995). Parents and collateral relatives of children with Pervasive Developmental Disorders: A family history study. *American Journal of Medical Genetics*, 60:282-289.
- Tanoue Y, Oda S, Asano F, Kaqashima K (1988). Epidemiology of infantile autism on Southern Ibarake, Japan: differences in prevalence rates in birth cohorts. *Journal of Autism and Developmental Disorders*, 18:155-166.
- Teng J, Siegmund D (1997). Combining information within and between pedigrees for mapping complex traits. *American Journal of Human Genetics*, 60:979-992.
- Terwilliger J, Shannon W, Lathrop G, Nolan J, Goldin L, Chase G, Weeks DE (1997). True and false positive peaks in genomewide scans: applications of length-based sampling to linkage mapping. *American Journal of Human Genetics*, 61:430-438.
- Thompson MW, McInnes RR, Willard HG (1991). *Thompson and Thompson Genetics in Medicine*. Saunders Company, USA.
- Thomson G (1995). Mapping disease genes: family-based association studies. *American Journal of Human Genetics*, 57:487-498.
- Todd RD (1986). Pervasive developmental disorders and immunological tolerance. *Psychiatric Development*, 4:147-165.
- Torrey EF, Hersh SP, McCabe KD (1975). Early childhood psychosis and bleeding during pregnancy: a prospective study of gravid women and their offspring. *Journal of Autism and Childhood Schizophrenia*, 5:287-297.
- Treffert DA (1970). Epidemiology of infantile autism. *Archives of General Psychiatry*, 22:431-438.

- Trofatter JA, MacCollin Mm, Rutter JL, Murrell JR, Duyao MP, Parry DM, Eldridge R, Kley N, Menon AG, Pulaski K, Hasse VH, Ambrose CM, Martuza EL, MacDonald ME, Seizinger BR, Short MP, Buckler AJ, Gusella JF (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell*, 72:791-800.
- Tsai L, Stewart MA (1983). Etiological implication of maternal age and birth order in infantile autism. *Journal of Autism and Developmental Disorders*, 13:57-65.
- Tsai L, Stewart MA, August G (1981). Implication of sex differences in the familial transmission of infantile autism. *Journal of Autism and Developmental Disorders*, 11:165-173.
- Tynan K (1998). *New Tools for Accelerating Human Genetic Research*. European School of Medical Genetics, 11th Course, Genoa, Italy.
- Vaillant G (1962). John Haslam on early infantile autism. *American Journal of Psychiatry*, 119:376.
- Venter A, Lord C, Schopler E (1992). A follow-up study of high-functioning autistic children. *Journal of Child Psychology and Psychiatry*, 33:489-507.
- Ventura M, Gibaud A, Le Pendu J, Hillaire D, Gerard G, Vitrac D, Oriol R (1988). Use of a simple method for the Epstein-Barr virus transformation of lymphocytes from members of large families of Reunion Island. *Human Heredity*, 38:36-43.
- Verkerk AJMH, 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*, 65:905-914.
- Vieland VJ, Hodge SE (1995). Inherent intractability of the ascertainment problem for pedigree data: a general likelihood framework. *American Journal of Human Genetics*, 56:33-43.
- Vieland VJ, Hodge SE (1996). The problem of ascertainment for linkage analysis. *American Journal of Human Genetics*, 58:1072-1084.
- Virtaranta-Knowles K, Sistonen P, Nevanlinna HR (1991). A population genetic study in Finland: comparison of the Finnish- and Swedish-speaking populations. *Human Heredity*, 41:248-264.
- Vogt P (1990). Potential genetic factors of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Human Genetics*, 84:301-336.
- Volkmar FR, Klin A, Siegel B, Szatmari P, Lord C, Campbell M, Freeman BJ, Cicchetti DV, Rutter M, Kline W, Buitelaar J, Hattab Y, Fombonne E, Fuentes J, Werry J, Stone W, Kerbeshian J, Hoshino Y, Bregman J, Loveland K, Szymanski L, Towbin K (1994). Field trial for autistic disorder in DSM-IV. *American Journal of Psychiatry*, 151:1361-1367.
- Volkmar FR, Nelson DS (1990). Seizure disorders in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29:127-129.

- Wadsby M, Lindehammar H, Eeg-Olofsson O (1989). *Neurofibromatosis in Childhood: Neuropsychological Aspects*. Neurofibromatosis 2, :pp 251-260.
- Wakabayashi S (1979). A case of infantile autism associated with Down's syndrome. *Journal of Autism and Developmental Disorders*, 9:31-36.
- Warren R, Singh V, Averett R, Odell J, Maciulis A, Burger R, Daniels W, Warren W (1996b). Immunogenetics studies in autism and related disorders. *Molecular and Chemical Neuropathology*, 28:77-81.
- Warren R, Singh V, Cole P, Odell JD, Pingree C, Warren L, DeWitt C, McCullough M (1992). Possible association of the extended MHC haplotype B44-SC30-DR4 with autism. *Immunogenetics*, 36:203-207.
- Warren RP, Margaretten NC, Pace NC, Foster A (1986). Immune abnormalities in patients with autism. *Journal of Autism and Developmental Disorders*, 16:189-197.
- Warren RP, Odell JD, Warren L, Burger RA, Maciulis A, Daniels WW, Torres AR (1996a). Strong association of the third hypervariable region of HLA-DRB1 with autism. *Journal of Neuroimmunology*, 67:97-102.
- Warren RP, Singh VK, Cole P, Odell JD, Pingree CB, Warren WL, White E (1991). Increased frequency of the null allele at the complement C4b locus in autism. *Clinical Experimental Immunology*, 83:438-440.
- Webb EVJ, Lobo S, Hervas A, Scourfield J, Fraser WI (1997). The changing prevalence of autistic disorder in a Welsh health district. *Developmental Medicine and Child Neurology*, 39:150-152.
- Weber JL (1990a). Informativeness of human (dC-dA)_n-(dG-dT)_n polymorphisms. *Genomics*, 7:524-530.
- Weber JL (1990b). Human DNA polymorphisms and methods of analysis. *Current Opinions in Biotechnology*, 1:166-171.
- Weber JL, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, 44:388-396.
- Weeks DE, Lathrop GM (1995). Polygenic disease: methods for mapping complex disease traits. *Trends in Genetics*, 11:513-519.
- Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseiz G, Lathrop M (1992). A second-generation linkage map of the human genome. *Nature*, 359:794-801.
- Weissman M, Merikangas KR, John J, Wickramaratne P, Prusoff BA, Kidd KK (1986). Family-genetic studies of psychiatric disorder. *Archives of General Psychiatry*, 43:1104-1116.
- Weizman A, Weizman R, Szekely GA, Wijzenbeek H, Livini E (1982). Abnormal immune response to brain tissue antigen in the syndrome of autism. *American Journal of Psychiatry*, 139:1462-1465.

- Willard HF, Waye JS, Skolnick MH, Schwartz CE, Powers VE, England SB (1986). Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific alpha-satellite DNA probes: implications for development of centromere-based genetic linkage maps. *Proceedings of the National Academy of Science USA*, 83:5611-5615.
- Williams SR, Sheper JM, Berger JS, Risch N, Goglin-Ewens KL, Spielman RS (1994). Systematic search for markers linked to insulin-dependent diabetes on chromosome 17. *American Journal of Human Genetics*, 55 (Suppl):A28.
- Wing L (1981). Language, social and cognitive impairments in autism and severe mental retardation. *Journal of Autism and Developmental Disorders*, 11:31-44.
- Wing L (1993). The definition and prevalence of autism: A review. *European Child and Adolescent Psychiatry*, 2:61-74.
- Wing L, Gould J (1979). Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9:11-29.
- Wing L, Yeates SR, Brierley LM, Gould J (1976). The prevalence of early childhood autism: Comparison of administrative and epidemiological studies. *Psychological Medicine*, 6:89-100.
- Wintero AK, Fredholm M, Dybdahl Thomsen P (1992). Variable (dG-dT)_n - (dC-dA)_n sequences in the porcine genome. *Genomics*, 12:281-288.
- Wolff S, Narayan S, Moyes B (1988). Personality characteristics of parents of autistic children. *Journal of Child Psychology and Psychiatry*, 29:143-153.
- Woodhouse W, Bailey A, Rutter M, Bolton P, Baird G, Le Couteur A (1996). Head circumference in autism and other pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 37:665-671.
- World Health Organization (1990). *International Classification of Diseases (ICD-10)*. World Health Organization, Geneva.
- World Health Organization (1992). *Draft of Chapter 5: Categories F00-F99, Mental and Behavioural Developmental Disorders*. In: *International Classification of Diseases*, 10th ed. World Health Organization, Geneva.
- World Health Organization (1993). *International Classification of Diseases (ICD-10)*. World Health Organization, Geneva.
- Zappella M (1993). Autism and hypomelanosis of Ito in twins. *Developmental Medicine and Child Neurology*, 35:826-832.

APPENDICES

APPENDIX 6.1: Microsatellite Information and Analysis Conditions

Chrom.	Microsatellite	Hetero.	Polymorph.	Dist. (cM)	Primer Dil. (F+R+dH2O)	[Primer]	[MgCl] (mM)	DNA (ul)	Taq	Thermocycler	PCR Prog	PCR Dil.	Allele Detect. Meth.
1	D1S551	0.73	tetranuc		4uM	0.05uM	1.75	2	G	PE 9600	TD	20ul dH2O: 1.5ul prod	Fluoro: 310
	D1S1631	0.81	trinuc	25.1	4uM	0.04uM	1.75	2	G	PE 9600	TD	20ul dH2O: 1.5ul prod	Fluoro: 310
	D1S534	0.94	tetranuc	13.0	4uM	0.04uM	1.75	2	G	PE 9600	TD	20ul dH2O: 1.5ul prod	Fluoro: 310
	D1S3462	0.43	trinuc		4uM	0.11uM	1.75	2	G	PE 9600	TD	8ul dH2O: 1.5ul prod	Fluoro: 310
	D1S235	0.69	dinuc	8.0	4uM	0.14uM	1.75	2	G	PE 9600	TD	8ul dH2O: 1.5ul prod	Fluoro: 310
	D1S1609	0.93	tetranuc	25.0	4uM	0.07uM	1.75	2	G	PE 9600	TD	20ul dH2O: 1.5ul prod	Fluoro: 310
6	D6S285	0.77	dinuc		0.67uM	0.13uM	1.50	1	S	MJ Research		no dil: 5.0ul prod	Fluoro: 373 and 310
	D6S1029	0.63	trinuc	5.9	100ng/ul	20ng/ul	1.50	1	G	MJ Research		no dil: 5.0ul prod	Fluoro: 373 and 310
	D6S276	0.84	dinuc	4.2	0.67uM	0.13uM	1.50	1	S	MJ Research		no dil: 5.0ul prod	Fluoro: 373 and 310
	D6S1014	0.76	trinuc	0.6	4uM	1.60uM	1.50	1	S	MJ Research		no dil: 5.0ul prod	Fluoro: 373 and 310
	D6S439	0.60	dinuc	1.4	100ng/ul	20ng/ul	1.50	1	G	MJ Research		no dil: 4.0ul prod	Fluoro: 373 and 310
	D6S497	?	tetranuc	0.1	100ng/ul	20ng/ul	1.50	1	G	MJ Research		no dil: 4.0ul prod	Fluoro: 373 and 310
	D6S1019	?	tetranuc	5.7	1.33uM	0.27uM	1.50	1	S	MJ Research		no dil: 4.0ul prod	Fluoro: 373 and 310
	D6S426	0.86	dinuc	4.1	4uM	2.00uM	1.50	1	S	MJ Research		no dil: 5.0ul prod	Fluoro: 373 and 310
7	D7S524	0.75	dinuc		100ng/ul	31ng/ul	1.75	2	S	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
	D7S527	0.76	dinuc	12.4	100ng/ul	19ng/ul	1.75	2	S	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
	D7S486	0.81	dinuc	15.1	100ng/ul	25ng/ul	2.50	2	S	PE 9600	standard	no dil: 1.5ul prod	Fluoro: 310
	D7S530	0.79	dinuc	6.4	100ng/ul	13ng/ul	1.75	2	S	PE 9600	standard	no dil: 1.5ul prod	Fluoro: 310
	D7S684	0.81	dinuc	16.6	100ng/ul	20ng/ul	2.50	2	S	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
	D7S2513	0.74	dinuc	3.4	100ng/ul	20ng/ul	1.75	2	S	PE 9600	standard	no dil: 1.5ul prod	Fluoro: 310
	D7S483	0.83	dinuc	10.3	100ng/ul	13ng/ul	2.50	2	S	PE 9600	standard	no dil: 1.5ul prod	Fluoro: 310

10	D10S1432	0.81	tetranuc		4uM	0.54uM	3.12	2	S	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
	D10S606	0.77	dinuc	2.9	100ng/ul	20ng/ul	2.08	2	S	PE 9600	standard	no dil: 5.0ul prod	Fluoro: 310
	D10S2327	0.69	tetranuc	6.9	4uM	0.40uM	1.56	2	G	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
15	D15S541	0.72	dinuc		5uM	1.25uM	2.00	2	S	MJ Research		no dil: 2.5ul prod	Radioactive (P32)
	D15S128	0.78	dinuc	6.1	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S156	0.54	dinuc	8.3	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S165	0.79	dinuc	9.6	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S118	0.76	dinuc	11.5	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S214	0.40	dinuc	8.7	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S132	0.76	dinuc	6.2	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S209	0.78	dinuc	19.5	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S117	0.78	dinuc	3.1	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S159	0.62	dinuc	9.4	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S153	0.87	dinuc	3.6	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S131	0.83	dinuc	8.6	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S152	0.79	dinuc	7.6	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S205	0.88	dinuc	0.1	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S158	0.75	dinuc	7.1	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S127	0.86	dinuc	0.7	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S130	0.66	dinuc	12.2	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S212	0.71	dinuc	11.7	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
	D15S120	0.73	dinuc	3.2	5uM	0.05uM	1.50	2	S	PE 9600	standard	20ul dH2O: 1.5ul prod	Fluoro: 373
17	D17S849	0.68	dinuc		5uM	1.26uM	2.50	2	G	PE 9600	standard	10ul dH2O: 2.0ul prod	Fluoro: 310
	D17S938	0.77	dinuc	20.3	5uM	1.92uM	2.50	2	G	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
	D17S945	0.86	dinuc	5.5	5uM	0.92uM	2.50	2	G	PE 9600	standard	no dil: 2.0ul prod	Fluoro: 310
18	D18S474	0.82	dinuc		5uM	1.24uM	2.50	2	G	PE 9600	standard	10ul dH2O: 2.0ul prod	Fluoro: 310
	D18S64	0.75	dinuc	15.6	5uM	1.24uM	2.50	2	G	PE 9600	standard	10ul dH2O: 2.0ul prod	Fluoro: 310

	D18S68	0.80	dinuc	12.7	5uM	0.60uM	2.50	2	G	PE 9600	standard	10ul dH2O: 2.0ul prod	Fluoro: 310
	D18S61	0.88	dinuc	6.9	5uM	0.32uM	2.50	2	G	PE 9600	standard	10ul dH2O: 2.0ul prod	Fluoro: 310
19	D19S209	0.77	dinuc		5uM	0.15uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S216	0.76	dinuc	8.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S221	0.86	dinuc	18.0	5uM	0.25uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S226	0.85	dinuc	6.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S414	0.78	dinuc	13.0	5uM	0.09uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S220	0.84	dinuc	10.0	5uM	0.13uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S420	0.79	dinuc	6.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S418	0.66	dinuc	26.0	5uM	0.07uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D19S210	0.74	dinuc	8.0	5uM	0.13uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
20	D20S117	0.84	dinuc		5uM	0.11uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S95	0.83	dinuc	14.0	5uM	0.23uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S115	0.66	dinuc	5.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S189	0.75	dinuc	9.0	5uM	0.09uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S186	0.86	dinuc	5.0	5uM	0.09uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S118	0.82	dinuc	12.0	5uM	0.11uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S195	0.81	dinuc	13.0	5uM	0.25uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S107	0.80	dinuc	7.0	5uM	0.09uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S119	0.82	dinuc	8.0	5uM	0.08uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S178	0.83	dinuc	5.0	5uM	0.07uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S196	0.81	dinuc	9.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S100	0.76	dinuc	9.0	5uM	0.06uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S173	0.67	dinuc	12.0	5uM	0.15uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
	D20S171	0.78	dinuc	10.0	5uM	0.15uM	2.50	2	G	PE 9600	standard	20ul dH2O: 2.0ul prod	Fluoro: 310
21	D21S258	0.87	dinuc		5uM	1.25uM	2.0	2	S	MJ Research		no dil: 2.5ul prod	Radioactive (P32)
	D21S1256	0.65	dinuc	5.8	5uM	1.70uM	1.50	2	S	PE 9600	TD	no dil: 1.0ul prod	Fluoro: 310

[illegible]

APPENDIX 6.2: SOLUTION AND BUFFER PROTOCOLS

(NOTE: all solution protocols are taken from Sambrook *et al.*, 1989)

12% FCS/RPMI 1640

Powdered Medium (RPMI 1640)	32.72g
Sodium Bicarbonate	4g
Fetal Bovine Serum	240ml
Penicillin/Streptomycin	40ml

adjust pH to 7.2, make up to a final volume of 2 litres with dH₂O, sterilise media by membrane filtration through a 0.2 μ m filter under positive pressure, aseptically dispense into sterile bottles and store in the dark at 2-8°C.

RCLB

1M Tris (pH 7.6)	10ml
5M NaCl	2ml
1M MgCl ₂	5ml

make up to a final volume of 1 litre with dH₂O and store at room temp.

Final concentration of reagents was 10mM of Tris, 10mM of NaCl and 5mM of MgCl₂

WCLB

1M Tris (pH 7.6)	10ml
0.5M EDTA, di-Sodium Salt	20ml
5M NaCl	10ml
0% SDS	20ml

make up to a final volume of 1 litre with dH₂O and store at room temp.

Final concentration of reagents was 10mM of Tris, 10mM of EDTA, 50mM of NaCl and 0.2%SDS.

50xTAE BUFFER (for 1L)

Tris Base	242g
Glacial Acetic Acid	57.4ml
Na ₂ EDTA.2H ₂ O	37.2g

make up to a final volume of 1 litre with dH₂O and store at room temp.

10xTBE BUFFER (for 1L)

Tris Base	108g
Boric Acid	55g
Na ₂ EDTA.2H ₂ O	8.3g

made up to 1l with dH₂O

The working solution (1xTBE) consists of 89mM Tris-base, 89mM Boric acid, 2mM EDTA, pH ~8.3 at ambient temperatures.

FICOLL LOADING BUFFER (6x)

50mM Tris HCl, pH 8.0

EDTA 1% (w/v)

Sarcosyl 7.5% (w/v)

Ficoll 400 (ave. mol. Wt. = 400 000)

0.05% bromophenol blue

0.05% xylene cyanol

6% POLYACRYLAMIDE-UREA GEL SOLUTION FOR USE IN THE ABI PRISM™ 377 GENETIC ANALYSER (fluorescent detection)

40g urea

12ml 40% acrylamide stock soln. (per 100ml contains acylamide 38g and bis-acrylamide 2g)

23ml dH₂O

Stir the ingredients together while gently warming in a water bath until the urea crystals begin to dissolve and the flask is slightly warm when touched. Continue to stir the soln. with no additional heating until all the urea has dissolved.

Filter the acrylamide soln. through a 0.2um cellulose acetate filter in a cup filter unit under vacuum, and then degas for five minutes.

Transfer the acrylamide soln. to a 100ml graduated cylinder containing 8ml of filtered 10x TBE buffer.

Adjust the volume to 80ml with dH₂O.

6% POLYACRYLAMIDE DENATURING GEL SOLUTION (19:1) FOR USE WITH THE SE1600 POKER FACE II SEQUENCER (radioactive detection)

Acrylamide	57g
Bis-acrylamide	3g
Urea	420.5g

Dissolve the acrylamide, bis-acrylamide and urea in 850ml of dH₂O and adjust the pH to 7.0.

or

Liqui-Mix 19 150ml

Urea 420.5g

Dissolve the urea in 600ml of dH₂O and add Liqui-Mix-19.

Add 5g of mixed bead resin (5g/L) to deionise the soln. and stir for 30min.

Filter and make up to 900ml with dH₂O.

Add 100ml of 10xTBE.

Degas for 30min.

Wrap bottle in foil and store at 4°C.

POLYACRYLAMIDE GEL FIXING SOLUTION

10% methanol / 10% acetic acid

100ml of methanol

100ml of acetic acid

make up to a final volume of 1l with dH₂O and store at room temp.

APPENDIX 6.3: PUBLISHED PAPERS

1. **Rogers, T.**, Kalaydjieva, L., Hallmayer, J., Petersen, P. B., Nicholas, P., Pingree, C., McMahon, W., Spiker, D., Lotspeich, L., Kraemer, H., McCague, P., Dimicelli, S., Nouri, N., Peachy, T., Yang, J., Hinds, D., Risch, N., Myers, R. M. (1999). Exclusion of linkage to the HLA region in ninety multiplex sibships with autism. *Journal of Autism and Developmental Disorders*, 29 (3): 195-201.
2. Salmon, B., Hallmayer, J., **Rogers, T.**, Kalaydjieva, L., Petersen, P. B., Nicholas, P., Pingree, C., McMahon, W., Spiker, D., Lotspeich, L., Kraemer, H., McCague, P., Dimicelli, S., Nouri, N., Pitts, T., Yang, J., Hinds, D., Myers, R. M., Risch, N. (1999). Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. *American Journal of Medical Genetics*, 88 (5): 551-556.
3. Risch, N., Spiker, D., Lotspeich, L., Nouri, N., Hinds, D., Hallmayer, J., Kalaydjieva, L., McCague, P., Dimiceli, S., Pitts, T., Nguyen, L., Yang, J., Harper, C., Thorpe, D., Vermeer, S., Young, H., Hebert, J., Lin, A., Ferguson, J., Chiotti, C., Wiese-Slater, S., **Rogers, T.**, Salmon, B., Nicholas, P., Peterson, P. B., Pingree, C., McMahon, W., Wong, D. L., Cavalli-Sforza, L. L., Kraemer, H. C., Myers, R. M. (1999). A genomic screen of autism: evidence for a multilocus etiology. *American Journal of Human Genetics*, 65 (2): 493-507.

Exclusion of Linkage to the HLA Region in Ninety Multiplex Sibships with Autism

Tamara Rogers,¹ Luba Kalaydjieva,¹ Joachim Hallmayer,² P. Brent Petersen,³ Peter Nicholas,³ Carmen Pingree,³ William M. McMahon,³ Donna Spiker,^{4,6} Linda Lotspeich,⁴ Helena Kraemer,⁴ Patty McCague,⁴ Sue Dimiceli,⁴ Nassim Nouri,⁵ Tawna Peachy,⁵ Joan Yang,⁵ David Hinds,⁵ Neil Risch⁵, and Richard M. Myers⁵

Several studies have suggested a role for the histocompatibility complex of loci (HLA) in the genetic susceptibility to autism. We have tested this hypothesis by linkage analysis using genetic marker loci in the HLA region on chromosome 6p in multiplex families with autism. We have examined sharing of alleles identical by descent in 97 affected sib pairs from 90 families. Results demonstrate no deviation from the null expectation of 50% sharing of alleles in this region; in fact, for most marker loci, the observed sharing was less than 50%. Thus, it is unlikely that loci in this region contribute to the genetic etiology of autism to any significant extent in our families.

KEY WORDS: Histocompatibility; HLA region; autism-affected sib pairs.

INTRODUCTION

Autism is a chronic developmental disorder characterized by behavioral, communication, and social abnormalities that manifest in the first years of life. Although the etiology of autism is unknown, there is general agreement that it results from early neurodevelopmental dysfunction.

Several studies suggest that immune imbalances might contribute to the abnormal development of the brain. Two of the possible explanations for this observation are a viral etiology and an autoimmune etiology. Clinical studies have reported numerous prenatal infections in autistic patients, most of which are viral: rubella

(Chess, 1977), encephalitis (Deykin & MacMahon, 1979; Markowitz, 1983), syphilis (Stubbs, 1978; Stubbs, Crawford, Burger, & Vanderbark, 1977), toxoplasmosis (Todd, 1986), and herpes simplex (Ghaziuddin, Tsai, Eilers, & Ghaziuddin, 1992). Other investigators have also found evidence for a deficiency in T-cell-mediated immunity (Warren, Margaretten, Pace, & Foster, 1986), depressed lymphocyte responsiveness (Stubbs, *et al.*, 1977), and antibodies against brain antigens (Weizman, Weizman, Szekely, Wijnenbeek, & Livini, 1982; Plioplys, Greaves, Kazemi, & Silverman, 1994; Singh, Warren, Odell, Warren & Cole, 1993) in autistic patients. It has been proposed that these autoimmune and/or viral processes in some way affect the nervous system and alter central nervous system (CNS) activity, which lead to the deficits observed in autistic children (Van Gent, Heijnen, & Trefferes, 1997).

The genetic composition of a person plays a critical role in the immune system's ability to respond to many antigens. Some of the genes that mediate immune response are located in the human leukocyte antigen (HLA) complex on chromosome 6. Investigators have reported associations between alleles of the HLA

¹ Centre for Human Genetics, Edith Cowan University, Perth, Australia.

² Centre for Clinical Research in Neuropsychiatry, Graylands Hospital/University of Western Australia, Perth, Australia.

³ Department of Psychiatry, University of Utah, Salt Lake City, Utah.

⁴ Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California.

⁵ Department of Genetics, Stanford University School of Medicine, Stanford, California.

⁶ Address all correspondence to Donna Spiker, Department of Psychiatry, Stanford

complex and autism when compared to normal controls. These autism-associated alleles include the null allele of the C4B gene (Warren *et al.*, 1991), the extended haplotype B44-S30-DR4 (Daniels *et al.*, 1995; Warren *et al.*, 1992;), the third hypervariable region of HLA-DRB1 (Warren, Odell, *et al.*, 1996) and two alleles of the DR beta 1 gene (Warren, Singh, *et al.*, 1996). The third hypervariable region of HLA-DRB1 showed the strongest association with autism (relative risk of 19.8 for allele 1 and 4.2 for allele 3, respectively; Warren, Odell, *et al.*, 1996).

Family and twin studies strongly suggest a genetic etiology for autism (Bolton & Rutter, 1990). Family studies demonstrate that the risk of autism among siblings of probands with autism is 3–5%, which is 50 to 100 times greater than in the general population (Bolton, *et al.*, 1994; Smalley, 1991). Twin studies have demonstrated that the concordance rate for monozygotic twins is approximately 25-fold higher than for dizygotic twins (Bailey *et al.*, 1995; Folstein & Rutter, 1977; Ritvo, Freeman, Mason-Brothers, Mo, & Ritvo, 1985; Steffenburg *et al.*, 1989). These data strongly suggest the interaction of multiple genetic loci in the etiology of autism. Thus, the reports of associations between specific HLA alleles and autism suggest that this region may harbor one such contribution. Therefore, the HLA region on chromosome 6p deserves further exploration, to determine the extent to which such associations might contribute to the genetic etiology of autism.

To test the hypothesis that a susceptibility gene in the HLA region contributes to the etiology of autism, we employed multipoint sib-pair analysis (Risch, 1990, 1993). This method detects how frequently two affected sibs share the same chromosomal region using the principle of identity by descent (ibd). Sib-pair analysis is very useful in studying autism because it is independent of both the affected status of parents and the mode of inheritance. The power of the sib-pair method to detect linkage depends on the risk ratio λ_s . The risk ratio λ_s is the increase in the risk of the disorder in sibs of an affected individual, compared with the general population prevalence. Assumptions about genetic model parameters, such as number of alleles at a given locus, gene frequencies, and penetrances are not required in sib-pair analysis (Risch, 1990). In the case of autism, the λ_s (increase in risk to siblings) is equal to 75. The value of 75 is the ratio of recurrence risk (3%) over the prevalence of autism in the general population (estimated to be 1/2500; Smalley, 1991). Because the λ_s value of 75 is high, affected sib-pair analysis may be a powerful tool

for mapping susceptibility genes in autism, depending on how many individual genes determine the total value of λ_s .

Because our sample is based on multiplex sibships, we can test for linkage, or the sharing of alleles, in the HLA region among siblings with autism. In this fashion, we have attempted to determine the extent to which genes in the HLA region contribute to the etiology of autism. As part of a larger general genome screen, 97 sib pairs from 90 families were analyzed for marker loci in the HLA region by using the multipoint sib-pair method to examine this question.

METHOD

Participants and Diagnosis

Families with two or more children with clinical diagnoses of autism or Pervasive Developmental Disorder Not Otherwise Specified (PDDNOS) were recruited. All presumed affected children were then evaluated with two standardized diagnostic instruments, the Autism Diagnostic Interview (ADI; Le Couteur *et al.*, 1989; Lord, Rutter, & Le Couteur, 1994) and the Autism Diagnostic Observation Schedule (ADOS; Lord *et al.*, 1989), to determine the research diagnosis of autism. The ADI is a semistructured interview, based on ICD-10 and DSM-IV criteria for autism, that is administered to the parents of the autistic children. The ADOS is a semistructured observation instrument used to observe the behavior of the child and corroborate the ADI information. This instrument was designed to elicit social, language, and communication behaviors. To check the reliability of diagnoses, all evaluations were videotaped and some were randomly selected and blindly assessed independently by other trained diagnosticians. To be eligible as a sib pair, affected children must have scores above the prespecified cutoff points in all four areas of the ADI (i.e., social impairment, language and communication impairment, unusual and restricted interests and routines, and age of onset prior to 3 years); also, review of the ADOS videotape by two or more diagnosticians was required to confirm the presence of significant deficits in social and communicative reciprocity consistent with a diagnosis of autism. Families were excluded if there was no consensus that all affected siblings had such significant deficits. This resulted in the exclusion of cases that would have clinical diagnoses of PDDNOS. For more details on the diagnostic protocol, see Spiker *et al.* (1994).

A total of 97 affected sib pairs from 90 families were available for the study, with a male to female ratio of 3.6:1 for the affected individuals and 0.8:1 for the unaffected individuals. The number of families with four, three, and two affected children was 2, 3, and 85, respectively. For families with n affected children, where n is greater than 2, $n-1$ sib pairs were formed by matching the first listed child (oldest) with each of the subsequent children; these pairs are fully independent. Both parents were available for genetic analysis in 67 families, and one parent in the remainder. In those families where one parent was absent, unaffected siblings were included when possible (8 families) to reconstruct missing parents for ibd tallies.

The affected siblings ranged in age at the time of the ADI administration from 2.9 to 40.9 years ($M = 9.8$ years, $SD = 7.3$, $Mdn = 7.3$, 25th–75th percentiles = 4.9–11.3 years). Children's diagnostic evaluation and school records were gathered to obtain IQ and mental age data. When available, nonverbal IQ scores were obtained from the performance subtests of the Wechsler Scales or Stanford-Binet Intelligence (4th ed.), or the full Leiter International Performance Scale or Merrill-Palmer Scale. When these scores were unavailable, scores from such tests as the Stanford-Binet Intelligence Scale (3rd ed.), Slossen Intelligence Scale, or McCarthy Scales were used, or ratio IQ scores were based on nonverbal scales from various developmental inventories (e.g., Daily Living Domain of the Vineland Adaptive Behavior Scale, the full Bayley Scales, or Developmental Profile II). The mean IQ was 66 ($SD = 28$, $Mdn = 66$, range = 15–160). One hundred eight children had scores below 70 (57%). The mean mental age was 68 months ($Mdn = 52$, $SD = 55$, range = 13–373 months). There were no families in which the IQs of all the affected siblings were below 30. In one family, one affected child had a mental age below 18 months.

Genotyping

DNA was extracted either from whole blood or immortalized lymphoblastoid cell lines by using standard proteinase K digestion and salting out procedures. Genotyping was carried out in two laboratories independently with slightly different procedures. In one laboratory, polymerase chain reactions (PCR) were carried out in 5 μ l and contained 5 ng DNA, 200 μ M dNTPs, 0.13–0.4 μ M of each primer (the forward primer had been fluorescently labeled using HEX, TET, and FAM dyes), 50 mM KCl, 10 μ M Tris-HCl, 1.5 mM $MgCl_2$, and 1.25 U Taq-Polymerase (Perkin Elmer). The

solutions were pipetted into a capillary tube (Biotech) using a programmable pipettor (Corbett Research), and processed in a thermocycler (MJ Research). After denaturation for 5 min at 94°C, 27 cycles were carried out, each cycle consisting of 30 sec at 94°C, 75 sec at 55°C, and 15 sec at 72°C. A final extension step was performed for 5 min at 72°C. After completion, 12 μ l of formamide and 0.8 μ l of GS500 Tamra standard dye were added, then the samples were denatured for 5 min at 94°C and loaded onto a Perkin Elmer 310 Capillary Sequencer. In the other laboratory, the PCR reactions were performed in a 10 μ l volume in 96-well plates and included 30 ng or DNA, 67 mM Tris-HCl pH 8.8, 200 μ M dNTPs, 16 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5–3 mM $MgCl_2$, 0.2 μ M of each primer, 0.2 U Taq polymerase (Perkin Elmer), using Perkin Elmer 9700 thermal cyclers. The forward primers for each marker were labeled as above. The reaction was first denatured at 95°C for 75 sec, then 35 cycles of 45 sec at 95°C, 45 sec at 57°C, 60 sec at 72°C were performed with a final elongation step of 7 min at 72°C. PCR products from 12 markers at a time were pooled and run on ABI Prizm 377 DNA sequencers. The products were resolved using the Applied Biosystems' GeneScan and Genotyper software packages. In both laboratories, polymorphic bands were scored and alleles were assigned to the pedigree members.

The following markers were used in the analysis: D6S285, D6S1959, D6S1029, D6S276, D6S1014, D6S497, D6S439, D6S1019, GGAA15B08, and D6S426. Location and distance of markers are given in Figure 1 and are based on a composite of the published Marshfield, CHLC, and Southampton linkage maps available on the Internet.

Unaffected sibs were also included in the genotyping of the families. Individuals performing the genotyping were blind to the diagnostic status of the subjects, although they were aware of the family structures.

Statistical Analysis

Linkage analysis was performed by using multi-point sib-pair analysis. This mapping approach provides optimal use of information by analyzing multiple markers, instead of flanking markers which are used in interval mapping (Hauser, Boehnke, Guo, & Risch, 1996; Olson, 1995). Also, the simultaneous use of multiple markers reduces the effect of recombination and allows for localization of a disease-susceptibility locus (Risch, 1990).

The method is based on the sharing of a chromosomal region by affected sib pairs. A random pair of

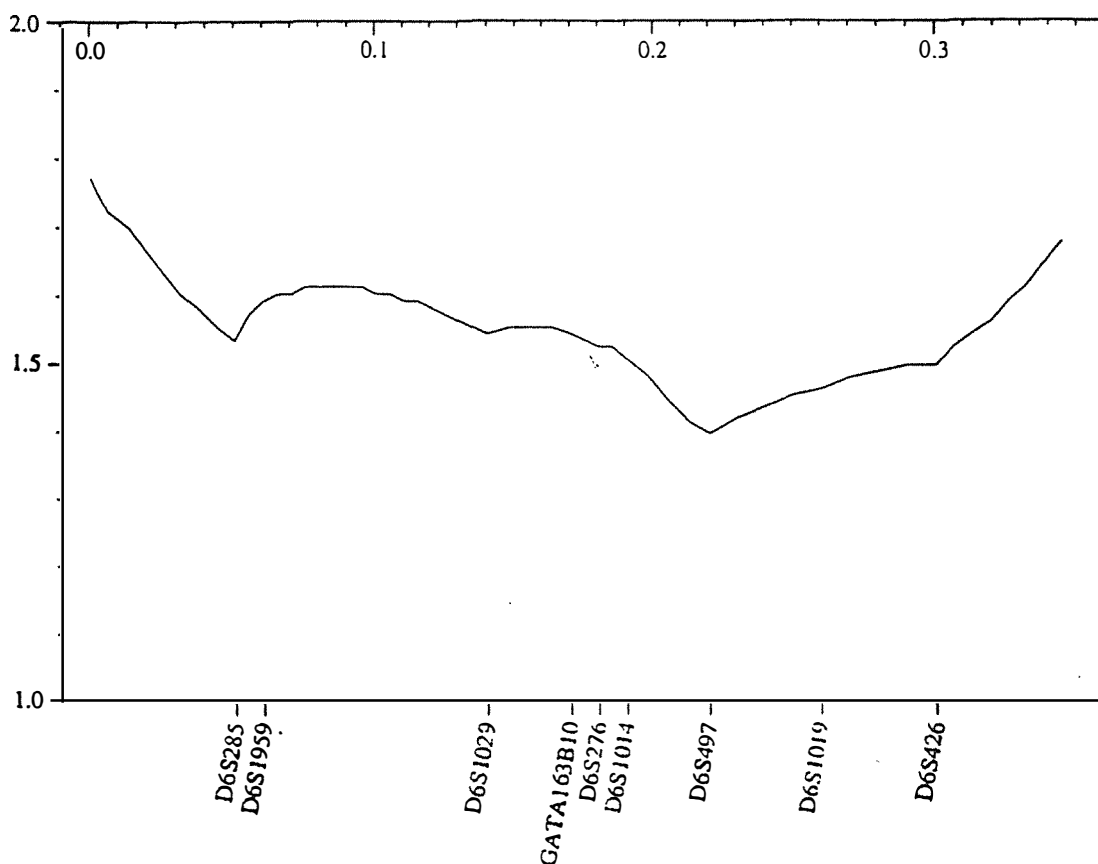


Fig. 1. Results of multipoint exclusion mapping in the HLA region. The vertical axis gives the smallest value of λ_s excluded (lod score < -2). The horizontal axis is map location on chromosome 6p measured in morgans (M), with 0 corresponding to 6p-ter. The HLA region is at .18M, near the marker D6S276.

sibs will share an allele at a genetic locus from a given parent half the time. Therefore, considering both parents simultaneously, the probability they will inherit the same allele from each of the two parents is 25%. Similarly, the probability they inherit different alleles from each parent is 25%. The probability they share an allele from one parent but not the other is 50%.

If an allele or alleles at a given locus increases the risk of autism, there will be an excess of sharing of alleles by affected sib pairs over the prior expectations for a random pair. The amount of excess depends on the frequency of the allele (or alleles), and how much the risk of autism is increased in those who carry it. One simple way to parameterize this excess is in terms of λ_s , the ratio of risk to a sib of an individual with autism compared to the population prevalence of autism due to this locus. Specifically, the probability an affected sib pair shares 0 alleles ibd at this locus, denoted z_0 , is given by $.25/\lambda_s$ (Risch, 1990). Assuming a model where the parental contributions are independent, we

can obtain the probability z_2 that the affected pair shares 2 alleles ibd (z_2) as $(1 - 1/2 \sqrt{\lambda_s})^2$, and for one allele ibd the probability z_1 is $(2 \sqrt{\lambda_s} - 1)/2 \lambda_s$.

At a given chromosomal location, the ibd probabilities, or equivalently the value of λ_s , can be estimated and its statistical significance assessed. This can be done for all genetic markers in a region simultaneously. The lod score for a sib pair at a given map location is the \log_{10} of the likelihood ratio of the marker data, given a value of $\lambda_s > 1$ versus a value of $\lambda_s = 1$. These lod scores were calculated for each sib pair at each marker location as well as at points between these markers, under the assumption of a specific λ value. The lod scores at each interval were then summed over all sib pairs to give the total lod score. If this score was < -2 , linkage was excluded for that value of λ_s ; a lod score > 3 was considered suggestive of linkage.

All calculations were carried out by using ASPEX, a set of computer programs for performing multipoint

exclusion mapping of affected sibling pair data for discrete traits (D. Hinds and N. Risch, personal communication). Two analyses were performed. In the first analysis, only ibd data were included from families with both parents typed, and in families with only one parent when an unaffected sib (or sibs) was available for reconstruction. In the second analysis, data on identity by state (ibs) were included for those families with missing parents, using likelihoods based on allele frequencies at the marker loci in these families. The results of these two analyses were similar; only the results for the ibd (first) analysis are presented.

We also performed analysis of linkage disequilibrium via the TDT test (Spielman *et al.*, 1993) in families with genotyped parents. For each of the microsatellite loci, we tested each allele with a frequency greater than 10% for excess transmission to affected children from heterozygous parents.

RESULTS

This analysis is based on allelic sharing, or ibd, between affected sibs at each of the marker loci, as well as a multipoint analysis of all the marker data considered simultaneously. For each marker, the null hypothesis value for allelic sharing, combining across the two parents, is 50%. The observed sharing at each of the markers is given in Table I. The sharing is less than 50% for all markers except one (D6S285), but does not differ significantly from 50% for any of the markers. Marker D6S285 had 56 out of a total 109 alleles shared (51%). Because the ibd in the HLA region was essentially negative (lod scores < 0), we calculated at each location on the map the smallest value of λ_s that would be excluded (lod score < -2) (see Figure 1). Values of $\lambda_s > 1.6$ can be excluded from the entire region, and values > 1.5 can

be excluded for most of the region involving the HLA loci (closest to markers D6S276 and D6S1014). Thus, we can exclude a gene of moderate effect on autism susceptibility from residing in this region.

We also tested each of the markers in the region for linkage disequilibrium using the TDT test. The results of these tests were also negative. After correction for the number of loci tested, there were no significant findings of excess allele transmission for any marker.

DISCUSSION

Specific HLA alleles have been associated with a variety of disorders. A few of the best documented associations include IDDM, multiple sclerosis, celiac disease, and Graves disease (Farid *et al.*, 1979; Haile *et al.*, 1983; Mackintosh & Asquith, 1978; Svejgaard & Ryder, 1989). Similar association studies have suggested that a susceptibility locus for autism is located in the HLA region. These association studies compared autistic subjects to normal controls to calculate relative risks for autism. In the study presented here, 90 families with at least two affected individuals were tested. Identity-by-descent information could be obtained for most families, thus allowing calculations to be independent of any gene or allele frequencies, mode of inheritance, and population stratification.

The linkage results presented here demonstrate that a moderate gene effect in the HLA region can be excluded. For a moderate gene effect to exist, λ_s values greater than 1.5 would result in positive lod scores. In this study, positive lod scores could not be found for any λ_s values greater than 1.0, thus excluding a moderate gene effect. However, linkage analysis could not formally exclude smaller gene effects. Similarly, TDT analysis did not produce any positive findings. These results are consistent with those of Spence *et al.*, (1985), who also found no evidence for linkage with HLA in 27 multiplex autism sibships.

Other results of HLA association studies in autism were based on singleton cases, whereas our analysis was based on multiplex sibships. One possible explanation for the apparent discrepancy between our results and previous studies is that the relative risks associated with alleles in the HLA region must be moderately large to produce a detectable linkage signal by sib-pair analysis. Thus, while our results do not entirely exclude a possible association between autism and alleles in the HLA region, they do indicate that if an association exists, it must be weak, and contributes little to the familial recurrence in autism.

Table I. Results of Identity by Descent Analysis of Autism Sibling Pairs

Marker	Total	% Shared	χ^2
D6S285	109	51.4	0.08
D6S1959	67	46.3	0.37
D6S1029	71	40.8	2.38
D6S276	123	47.2	0.40
D6S1014	109	48.6	0.08
D6S497	88	42.0	2.23
D6S439	82	45.1	0.78
D6S1019	91	46.2	0.54
GGAA15B08	130	45.4	1.11
D6S426	132	46.2	0.76

Alternatively, certain alleles in the HLA region may predispose to autism only for a subset of patients, perhaps those with a suspected autoimmune etiology or those exposed to recurrent infection or immunologic insult. Our families were unselected in this regard, and may have contained few with this type of etiology. However, based on our results, it is also likely that if HLA were involved in a significant way in these cases, they account for a relatively small proportion of all multiplex families.

Thus, putative immunological abnormalities found in autistic children (Plioplys et al., 1994; Singh et al., 1993; Stubbs, 1977; Warren et al., 1986; Weizman et al., 1982) are unlikely to be related to susceptibility genes in the HLA region. The immune system is connected to the brain in an integrated regulatory circuit that is only slowly beginning to be understood. An altered immune response during early childhood might produce changes in brain development. Although it appears that genes in the HLA region have little or no effect on this process, we have not excluded the possibility that genes of immunologic interest other than HLA might be involved in the pathophysiology of autism.

ACKNOWLEDGMENTS

This research has been supported by NIMH Grants MH52708, MH39437, MH00219, and MH00980; NHMRC Grant 0034328; and by grants from the Scottish Rite, the Spunk Fund, Inc., the Rebecca and Solomon Baker Fund, the APEX Foundation, the National Alliance for Research in Schizophrenia and Affective Disorders (NARSAD), the endowment fund of the Nancy Pritzker Laboratory, Stanford University; and gifts from the Autism Society of America and families and friends of individuals with autism. We are grateful to our external clinical advisors, Isabelle Rapin, Peter Szatmari, and Luke Tsai, for their expert advice on clinical aspects of this study. Finally, we express our gratitude to our partners in this research, the families who have participated in this project.

REFERENCES

- Bailey, A., LeCouteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine*, 25, 63-77.
- Bolton, P., & Rutter, M. (1990). Genetic influences in autism. *International Review of Psychiatry*, 2, 67-80.
- 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. *Journal of Child Psychology and Psychiatry*, 35, 877-900.
- Chess, S. (1977). Follow-up report on autism in congenital rubella. *Journal of Autism and Childhood Schizophrenia*, 7, 79-81.
- Daniels, W. W., Warren, R. P., Odell, J. D., Maciulis, A., Burger, R. A., Warren, W. L., & Torres, A. R. (1995). Increased frequency of the extended or ancestral Haplotype B44-SC30-DR4 in autism. *Neuropsychobiology*, 32, 120-123.
- Deykin, E. Y., & MacMahon, B. (1979). Viral exposure and autism. *American Journal of Epidemiology*, 109, 628-638.
- Farid, N. R., Sampson, L., Noel, E. P., Barnard, J. M., Mandeville, R., Larsen, B., Marshall, W. H., & Carter, N. D. (1979). A study of human leukocyte D locus related antigens in Graves' disease. *Journal of Clinical Investigation*, 63, 108-113.
- Folstein, S. E., & Rutter, M. (1977). Infantile autism: A genetic study in 21 twin pairs. *Journal of Child Psychology and Psychiatry*, 18, 297-321.
- Ghaziuddin, M., Tsai, L. Y., Eilers, L., & Ghaziuddin, N. (1992). Brief report: Autism and herpes simplex encephalitis. *Journal of Autism and Developmental Disorders*, 22, 107-113.
- Haile, R. W., Hodge, S. E., & Iselius, L. (1983). Celiac disease genetic susceptibility to multiple sclerosis: A review. *International Journal of Epidemiology*, 12, 8-16.
- Hauser, E. R., Boehnke, M., Guo, S.-W., & Risch, N. (1996). Affected sib pair interval mapping and exclusion for complex genetic traits: Sampling considerations. *Genetic Epidemiology*, 13(2), 117-137.
- Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism Diagnostic Interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19, 363-387.
- Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, 19, 185-212.
- 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. *Journal of Autism and Developmental Disorders*, 24, 659-685.
- Mackintosh, P., & Asquith, P. (1978). HLA and coeliac disease. *British Medical Bulletin*, 34, 291-294.
- Markowitz, P. L. (1983). Autism in a child with congenital cytomegalovirus infection. *Journal of Autism and Developmental Disorders*, 13, 245-254.
- Olson, J. M. (1995). Multipoint linkage analysis using sib pairs: An interval mapping approach using dichotomous outcomes. *American Journal of Human Genetics*, 56, 788-798.
- Plioplys, A. V., Greaves, A., Kazemi, K., & Silverman, E. (1994). Immunoglobulin reactivity in autism and Rett's syndrome. *Developmental Brain Dysfunctions*, 7, 12-16.
- Risch, N. (1990). Linkage studies for genetically complex traits. II. The power of affected sib pairs. *American Journal of Human Genetics*, 46, 229-241.
- Risch, N. (1993). Exclusion mapping for complex diseases. *American Journal of Human Genetics*, 53 (Suppl.), A185.
- Ritvo, E. R., Freeman, B. J., Mason-Brothers, A., Mo, A., & Ritvo, A. M. (1985). Concordance for the syndrome of autism in 40 pairs of afflicted twins. *American Journal of Psychiatry*, 142, 64-77.
- Singh, V. K., Warren, R. P., Odell, J. D., Warren, W. L., & Cole, P. (1993). Antibodies to myelin basic protein in children with autistic behavior. *Brain, Behavior and Immunology*, 7, 97-103.
- Smalley, S. L. (1991). Genetic influences in autism. *Psychiatric Clinics of North America*, 14, 125-139.

- Spence, M. A., Ritro, E. R., Marazita, M. L., Funderbunk, S. J., Sparkes, R. S., Freeman, B. J. (1985). Gene mapping studies with the syndrome of autism. *Behavioral Genetics*, 15, 1-13.
- Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, 52, 506-516.
- Spiker, D., Lotspeich, L., Kraemer, H. C., Hallmayer, J., McMahon, W., Petersen, P. B., Wong, D. L., Dimiceli, S., Ritvo, E., Cavalli-Sforza, L. L., & Ciaranello, R. D. (1994). Genetics of autism: Characteristics of affected and unaffected children from 37 multiplex families. *American Journal of Medical Genetics*, 54, 27-35.
- Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry*, 30, 405-416.
- Stubbs, E. G. (1978). Autistic symptoms in a child with congenital cytomegalovirus infection. *Journal of Autism and Childhood Schizophrenia*, 8, 37-43.
- Stubbs, E. G., Crawford, M. L., Burger, D. R., & Vanderbark, A. A. (1977). Depressed lymphocyte responsiveness in autistic children. *Journal of Autism and Childhood Schizophrenia*, 7, 49-55.
- Svejgaard, A., & Ryder, L. P., (1989). HLA and insulin-dependent diabetes: An overview. *Genetic Epidemiology*, 6, 1-14.
- Todd, R. D. (1986). Pervasive developmental disorders and immunological tolerance. *Psychiatric Developments*, 2, 147-165.
- van Gent, T., Heijnen, C. J., & Treffers, P. D. A. (1997). Autism and the immune system. *Journal of Child Psychology and Psychiatry*, 38, 337-349.
- Warren, R. P., Margaretten, N. C., Pace, N. C., & Foster, A. (1986). Immune abnormalities in patients with autism. *Journal of Autism and Developmental Disorders*, 16, 189-197.
- Warren, R. P., Singh, V. K., Cole, P., Odell, J. D., Pingree, C. B., Warren, W. L., & White, E. (1991). Increased frequency of the null allele at the complement C4B locus in autism. *Clinical Experimental Immunology*, 83, 438-440.
- Warren, R. P., Singh, V. K., Cole, P., Odell, J. D., Pingree, C. B., Warren, W. L., DeWitt, C. W., & McCullough, M. (1992). Possible association of the extended MHC haplotype B44-SC30-DR4 with autism. *Immunogenetics*, 36, 203-207.
- Warren, R. P., Odell, J. D., Warren, L., Burger, R. A., Maciulis, A., Daniels, W. W., & Torres, A. R. (1996). Strong association of the third hypervariable region of HLA-DRB1 with autism. *Journal of Neuroimmunology*, 67, 97-102.
- Warren, R. P., Singh, V. K., Averett, R. E., Odell, J. D., Maciulis, A., Burger, R. A., Daniels, W. W., & Warren, W. L. (1996). Immunogenetic studies in autism and related disorders. *Molecular Chemical Neuropathology*, 28, 77-81.
- Weizman, A., Weizman, R., Szekely, G. A., Wijsenbeek, H., & Livini, E. (1982). Abnormal immune response to brain tissue antigen in the syndrome of autism. *American Journal of Psychiatry*, 139, 1462-1465.

Absence of Linkage and Linkage Disequilibrium to Chromosome 15q11-q13 Markers in 139 Multiplex Families With Autism

Boyd Salmon,¹ Joachim Hallmayer,¹ Tamara Rogers,² Luba Kalaydjieva,² P. Brent Petersen,³ Peter Nicholas,³ Carmen Pingree,³ William McMahon,³ Donna Spiker,^{4*} Linda Lotspeich,⁴ Helena Kraemer,⁴ Patty McCague,⁴ Sue Dimiceli,⁴ Nassim Nouri,⁵ Tawna Pitts,⁵ Joan Yang,⁵ David Hinds,⁵ Richard M. Myers,⁵ and Neil Risch⁵

¹Centre for Clinical Research in Neuropsychiatry, Graylands Hospital / University of Western Australia, Perth, Australia

²Centre for Human Genetics, Edith Cowan University, Perth, Australia

³Department of Psychiatry, University of Utah, Salt Lake City, Utah

⁴Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California

⁵Department of Genetics, Stanford University School of Medicine, Stanford, California

Chromosomal region 15q11-q13 has been implicated to harbor a susceptibility gene or genes underlying autism. Evidence has been derived from the existence of cytogenetic anomalies in this region associated with autism, and the report of linkage in a modest collection of multiplex families. Most recently, linkage disequilibrium with the marker GABRB3-155CA2 in the candidate locus GABRB3, located in this region, has been reported. We searched for linkage using eight microsatellite markers located in this region of chromosome 15 in 147 affected sib-pairs from 139 multiplex autism families. We also tested for linkage disequilibrium in the same set of families with the same markers. We found no evidence for excess allele sharing (linkage) for the markers in this region. Also, we found no evidence of linkage disequilibrium, including for the locus GABRB3-155CA2. Thus, it appears that

the role of this region of chromosome 15 is minor, at best, in the majority of individuals with autism. *Am. J. Med. Genet. (Neuropsychiatr. Genet.)* 88:551–556, 1999.

© 1999 Wiley-Liss, Inc.

KEY WORDS: chromosome 15; autism

INTRODUCTION

Twin and family studies strongly support the notion that genetic factors play a significant role in the etiology of autism. The concordance rate for autism is much higher in monozygotic twins than in dizygotic twins [Bailey et al., 1995; Folstein and Rutter, 1977; Ritvo et al., 1985; Steffenburg et al., 1989], and family studies indicate that the sibling recurrence risk is 50–100 times greater than the risk for autism in the general population [reviewed in Smalley, 1997]. As in other psychiatric disorders, there is no simple mode of inheritance. Instead, the observed family pattern is compatible with a complex mode of transmission involving several or many genes. The effect of each separate gene may be modest, and interactions between genes would be expected to further complicate the genetic mechanisms involved.

Cytogenetic abnormalities have frequently been used to identify genes associated with inherited syndromes. Cytogenetic studies in autism have suggested an association with chromosomal anomalies in the 15q11-q13 region. Such cytogenetic abnormalities may indicate that a susceptibility gene (or genes) for autism is located in this region. Furthermore, symptoms of autism have been reported in individuals with Prader-Willi and Angelman syndromes [Arrieta, 1994], both of which result from alterations in the 15q11-q13 region. When only the paternal region of human chromosome

Contract grant sponsor: National Institute of Mental Health; Contract grant numbers: MH52708-03, MH39437, MH00219, MH00980-03; Contract grant sponsor: National Institutes of Health; Contract grant number: HG00348; Contract grant sponsor: NHMRC; Contract grant number: 0034328; Contract grant sponsor: Scottish Rite; Contract grant sponsor: Spunk Fund, Inc.; Contract grant sponsor: Rebecca and Solomon Baker Fund; Contract grant sponsor: APEX Foundation; Contract grant sponsor: National Alliance for Research in Schizophrenia and Affective Disorders (NARSAD); Contract grant sponsor: Nancy Pritzker Laboratory, Stanford, California; Contract grant sponsor: Autism Society of America.

*Correspondence to: Donna Spiker, Ph.D., Department of Psychiatry, Stanford University School of Medicine, Stanford, CA 94305-5719.

Received 29 October 1998; Accepted 2 March 1999

15q11-q13 is inherited, either by deletion of the maternal gene or by uniparental disomy, Prader-Willi syndrome results. Angelman syndrome results when only maternal copies of 15q11-q13 are inherited [Christian et al., 1995; Sutcliffe et al., 1997]. The most common cytogenetic abnormality reported in autism involves duplications within the 15q11-q13 region [Baker et al., 1994; Bunday et al., 1994; Cook et al., 1997a; Crolla et al., 1995; Flejter et al., 1996; Gillberg et al., 1991; Hotof and Bolton, 1995; Leana-Cox et al., 1994; Rineer et al., 1998; Robinson et al., 1993; Schinzel et al., 1994; Schroer et al., 1998]. Cook et al. [1997a] described a family with two affected children in which a maternal 15q duplication segregates with autism. The authors suggested that, although large genetic abnormalities in the 15q11-q13 region are quite uncommon in autistic patients, there may be a higher rate of small, cytologically undetectable abnormalities in this chromosomal region in some cases of autism.

Several investigators have recently examined the 15q11-q13 region for genetic linkage to autism. Cook et al. [1998] carried out linkage disequilibrium mapping of markers in this region. Evidence for linkage disequilibrium was found at the GABRB3 locus with the marker GABRB3 155CA-2, but not with the marker D15S97. No evidence of imprinting effects were detected. Most of the 140 families included in this study were singleton affected cases. Pericak-Vance et al. [1997], using sib-pair analysis with 37 autistic multiplex families, obtained a peak lod score of 2.5 for the microsatellite marker D15S156, which is located approximately 5 cM distal to the GABRB3 locus, and a lod score for GABRB3 of 1.4. A recently reported whole genome linkage scan found several areas potentially of interest, most notably one on chromosome 7 with a lod score of 2.5 [International Autism Consortium, 1998]. These researchers did not find significant linkage to autism for markers on chromosome 15. However, this genome scan included only one marker, D15S128, in the 15q11-q13 region.

In the present study, we used multipoint sib-pair analysis to test for linkage between autism and eight markers in the proximal region of chromosome 15q. We analyzed 147 affected sib-pairs from 139 families by genotyping fluorescently labeled microsatellite markers. To be included in the analyses, autistic individuals had to satisfy criteria for all three symptom areas of the Autism Diagnostic Interview (ADI) (social impairment, language and communication impairment, unusual routines, and restricted interests), plus an age-of-onset prior to 3 years. This linkage study of proximal chromosome 15q is part of a larger investigation of the genetics of autism that includes an entire genome scan, which will be reported elsewhere. Our results indicate that the proximal region of chromosome 15q, in particular markers in the 15q11-q13 region, does not show evidence of linkage to autism in these families. Furthermore, we find no evidence of linkage disequilibrium with the two GABRB3 microsatellite markers GABRB3CA or GABRB3-155CA2.

MATERIALS AND METHODS

Recruitment of Families and Diagnostic Assessment

Families were recruited for the study if they had at least two siblings with a clinical diagnosis of autism or pervasive developmental disorder-not otherwise specified (PDD-NOS). All children were subsequently assessed by using the (ADI) [LeCouteur et al., 1989; Lord et al., 1994, 1997] and the Autism Diagnosis Observation Schedule (ADOS) [Lord et al., 1989] to determine a research diagnosis of autism. The ADI is a semistructured interview of parents based on *International Classification of Diseases* (ICD-10) and *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria for the diagnosis of autism. The ADOS is a semistructured observational instrument that assesses the child by observing his/her behavior, and is used to corroborate the ADI interview. The diagnostic assessments (ADI and ADOS) were videotaped and subjected to independent reliability checks by other trained interviewers (typically the senior clinical investigators). To be considered affected in the sib-pair analysis, an individual had to satisfy the prespecified cut-off scores in all three symptom areas of the ADI (social impairment, language and communication impairment, unusual and restricted interests) and the age-of-onset had to be prior to 3 years. Review of the ADOS tape by two or more diagnosticians was used to exclude children who did not show significant impairments in social and communicative reciprocity. Families were excluded if there was no consensus that at least two affected children had such deficits consistent with a diagnosis of autism (e.g., exclusion of cases that would have a clinical diagnosis of PDD-NOS); this entailed exclusion of 33 families. We also excluded families where all affected sibs were severely retarded (IQ <30; seven families), or had significant medical conditions likely to account for the autism (six families). For more details on the diagnostic protocol, see Spiker et al. [1994].

Children's records were reviewed to obtain IQ and mental age information. When available, nonverbal IQ scores were obtained from performance subtests from the Wechsler Scales; Stanford-Binet Intelligence Scale, fourth edition; the Leiter International Performance Scale; or the Merrill-Palmer Scale. When unavailable, scores from such tests as the Stanford-Binet Intelligence Scale, third edition; Slossen Intelligence Scale; or McCarthy Scales were used. For children without these tests, mental ages and ratio IQ scores were derived from nonverbal scales of various developmental instruments (e.g., Daily Living Scale of the Vineland Adaptive Behavior Scale, the Bayley Mental Scale, Developmental Inventory II).

Genotyping

Blood was collected on all affected individuals and, if available, their parents and siblings. Lymphoblast cell lines were established and genomic DNA was extracted from peripheral blood lymphocytes or cell lines by using standard proteinase K digestion and salting out procedures.

Genotyping was carried out in two laboratories independently with slightly different procedures. In one lab, polymerase chain reaction (PCR) assays were performed in 8 microliter volume reactions containing 10 ng genomic DNA; 2.5 mM of each dNTP; 0.2–0.6 micromolar of each primer pair labeled with Fam, Hex, or Tet; 0.8 microliters of 10× buffer (Perkin-Elmer, Norwalk, CT); 1.5–2.5 mM MgCl₂; and 0.2 units AmpliTaq Gold Polymerase (Perkin Elmer). PCR assays were performed in an Applied Biosystems (Foster City, CA) 9600 Thermocycler by denaturing for 10 minutes at 94°C, then 15 cycles of 30 sec at 94°C, 15 sec at 60°C, and 15 sec at 72°C. Another 20 cycles of 30 sec at 94°C, 15 sec at 65°C, and 15 sec at 72°C were performed, and these were followed by a final extension step of 10 min at 72°C. In the other lab, PCR reactions were performed in a 10 µL volume in 96 well plates and included 30 ng of DNA, 67 mM Tris-HCl pH 8.8, 200 µM dNTPs, 16 mM (NH₄)₂SO₄, 0.01% Tween-20, 1.5–3 mM MgCl₂, 0.2 µM of each primer, 0.2 U Taq polymerase (Perkin Elmer), using Perkin Elmer 9700 thermal cyclers. The forward primers for each marker were labeled as above. The reaction was first denatured at 95°C for 75 sec, then 35 cycles of 45 sec at 95°C, 45 sec at 57°C, 60 sec at 72°C were performed with a final elongation step of 7 min at 72°C. Fluorescently labeled markers were analyzed on Applied Biosystems 373 or 377 Genetic Analyzers, and the data were further analyzed by using Applied Biosystems Genescan and Genotyper software. In both labs, polymorphic bands were scored and alleles were assigned to the pedigree members.

The following eight markers were typed: D15S128, GABRB3CA, GABRB3-155CA2, D15S822, D15S156, D15S165, D15S1232, and ACTC. These loci span approximately 25 cM of chromosome 15, in the region of 15q11-q13. Figure 1 shows the position of these markers and the distances between them.

Statistical Analyses

Multipoint sib-pair linkage analysis was used to detect or exclude linkage. Multipoint sib-pair analysis is

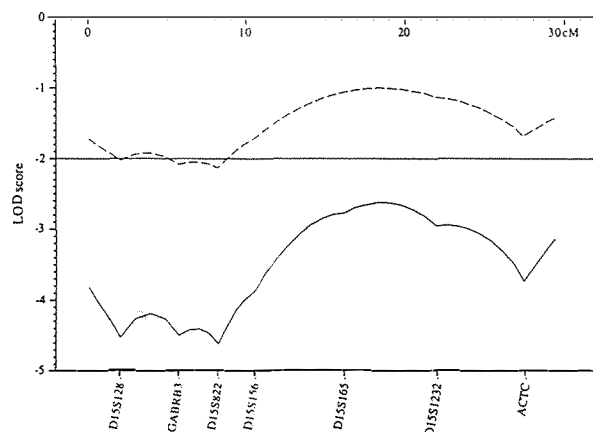


Fig. 1. Multipoint exclusion plot for an autism susceptibility locus on chromosome 15q11-13. The top scale is map distance (in centimorgans), and the Y axis is the lod score. The relative location of the various chromosome 15q11-13 markers is given along the bottom. The solid line is for $\lambda_S = 2.0$; the dashed line is for $\lambda_S = 1.5$.

based on allele sharing, or identity by descent (ibd) between affected siblings over a number of loci. At each locus, the null hypothesis is that the likelihood of affected sibs sharing an allele is 50%. Lod scores are calculated per sib-pair at points between markers and at markers by using a specific value of λ_S , which is the sibling recurrence risk ratio due to the locus in this region [Risch, 1990]. Each lod score is the log10 of the ratio of the likelihood of observing the marker data at a given value of λ_S (>1) versus the likelihood at a value of $\lambda_S = 1$. For a specific value of λ_S and marker distance, lod scores are calculated for each sib-pair. The total lod score is an overall sum for all sib-pairs. For a given value of λ_S , a lod score less than -2 is indicative of strong exclusion of linkage [Hauser et al., 1996] and values larger than 3 suggest linkage. Lod scores are calculated at multiple points over a genetic distance. The advantages of this type of analysis compared with interval mapping with flanking markers are that it allows all of the information to be used simultaneously, and it reduces the effect of recombination in defining a disease-causing locus [Risch, 1990]. For the six families with n affected sibs, where $n > 2$, $n - 1$ sib pairs were formed with one sib randomly selected. These pairs are fully independent for ibd scoring. Thus, in total, there are 147 independent affected sib pairs. The transmission disequilibrium test (TDT) [Spielman et al., 1993] was used to test for linkage disequilibrium for the tested markers. In this case, we employed two tests, one that calculates a total chi-square statistic on the full $2 \times n$ table, and the second that calculates the maximum chi-square across each tested allele; the level of significance for both tests was calculated using Monte-Carlo resampling [Lazzeroni and Lange, 1998]. Because we are dealing with multiplex sibships, all resampling was conditional on parental genotypes; the two alleles of each parent were switched with 50% probability [Martin et al., 1998]. All statistical analyses were performed by using the program package ASPEX (Hinds D, Risch N: ftp site //lahmed.stanford.edu/pub/aspeX).

RESULTS

Family Recruitment

One hundred thirty-nine families encompassing 147 independent sib-pairs were tested. Each family contained two or more affected siblings that satisfied the inclusion criteria for this study. One hundred thirty-five families had two affected sibs, four had three affected sibs, and two had four affected sibs. Among affected individuals, the male-to-female ratio was 3.4:1, and for unaffected individuals, the ratio was 0.8:1. In 115 of these families, DNA from both parents was available for testing; in the remaining 24 families, only one parent had DNA available. In those families where one parent was absent, unaffected siblings were included when possible (eight families) to reconstruct missing parents and ibd tallies.

At the time of the ADI administration, the affected siblings ranged in age from 2.8 to 40.9 years (mean = 8.7 years, SD = 6.4; 25th–75th centiles = 4.7–10.1

years). Mental age and IQ estimates were obtained based on available diagnostic evaluations and school records. The mean nonverbal IQ was 64 (SD = 27; range = 15–160). One hundred and seventy children (59%) had IQ scores below 70. The mean mental age was 61 months (SD = 48 months; range = 13–373 months). There were no families for which all affected children had IQ scores below 30. In four families, one affected child had a mental age below 18 months.

Approximately 90% of our families were Caucasian of mixed European origin. The remaining 10% were a mix of African/African-American, Asian, and Hispanic.

Multipoint Sib-Pair Analysis

The number of shared alleles between affected siblings at each locus tested is shown in Table I. None of the loci showed a significant excess of sharing. In fact, six of the loci showed less than or equal to 50% sharing, while the other two showed sharing only slightly greater than 50%. The estimate of sharing at each of these loci based on multipoint analysis is also given in Table I. Overall, the estimate of sharing in this region hovers around 50%. The multipoint lod scores for this region were also negative at all loci for all values of λ_S . We calculated the lod scores for two values of λ_S : $\lambda_S = 2.0$ and 1.5 . These results are given in Figure 1. Values of λ_S of 2.0 or larger could be strongly excluded over the entire interval, while a locus with $\lambda_S = 1.5$ could be strongly excluded from the region surrounding GABRB3.

The chromosomal region 15q11-13 is subject to genomic imprinting, and thus allelic expression might depend on sex of transmitting parent. Indeed, Cook et al. [1997b] found a duplication in this region to be expressed only when maternally inherited, and noted similar observations in other studies. Therefore, we examined our ibd data separately for alleles inherited from fathers versus mothers (Table I). Sharing across the 15q11-13 region in our families was somewhat greater for paternally versus maternally derived alleles (although the difference was not statistically significant), contrary to prediction based on the prior cytogenetic evidence.

Linkage Disequilibrium

The TDT was used to test for linkage disequilibrium at the eight marker loci, including the markers GABRB3CA, which is adjacent to the gene GABRB3, and the marker GABRB3-155CA2, within the GABRB3

gene and previously found to be in disequilibrium with autism [Cook et al., 1998]. We found no evidence of preferential allele transmission at any locus, either based on a global test of all alleles, or the allele giving maximum deviation. The results for GABRB3-155CA2 are given in Table II. Our allele frequency distribution is very close to that reported by Cook et al. [1998]. In that study, allele “103” demonstrated excess transmission to affected offspring. The corresponding allele according to our nomenclature is the “108” allele. In our families, this allele was transmitted 79 times and not transmitted 62 times, giving a nonsignificant chi square of 2.0.

DISCUSSION

By using multipoint sib-pair analysis, we excluded the existence of a gene in the region of chromosome 15q11-q13 conferring a moderately increased risk for autism in our set of 139 multiplex families. The entire region is strongly excluded for λ_S values greater than 2, while values of λ_S less than 1.5 could be excluded using the less stringent criterion of -1 . Our uniformly negative lod scores across this region of chromosome 15 for all values of λ_S do not support the presence of a susceptibility gene of strong or moderate effect.

Our results are in contrast to those from the study of Pericak-Vance et al. [1997] and Cook et al. [1998]. Pericak-Vance et al. [1997] reported evidence for linkage in 38 affected sib-pairs and nine cousin pairs on chromosome 15q11-q13. Their maximum lod score was 2.5 at D15S156, a score below the critical value of 3. The observed sharing at this marker in our families was 53.7%, and the multipoint sharing at this marker was 50.3%. The International Autism Consortium [1998] also scanned chromosome 15q (primarily marker D15S128) in 39 multiplex families and found no evidence of a major susceptibility gene for autism, consistent with our results.

The study of Cook et al. [1998] used the multiallelic transmission-disequilibrium test and revealed linkage disequilibrium between autistic disorder and a microsatellite marker in the gamma-aminobutyric acid A receptor subunit gene, GABRB3 155CA-2 (MTDT 28.63, 10 df, $P = .0014$). We tested two microsatellite markers at this locus, GABRB3CA [Mutirangura et al., 1992] and GABRB3-155CA2. We found no evidence of linkage disequilibrium with either marker.

The seemingly contradictory nature of the results between the respective studies must be interpreted in the

TABLE I. Results of Identity by Descent (Allele Sharing) Analysis of Chromosome 15q11-q13 Markers in 147 Autism Sibling Pairs

Marker	Paternal (%)	Maternal (%)	Total (%)	Multipoint sharing (%)
D15S128	52/95 (54.7)	41/96 (42.7)	93/191 (48.7)	49.4
GABRB3CA	47/84 (56.0)	39/91 (42.9)	87/177 (49.2)	49.2
GABRB3-155CA2	52/94 (55.3)	44/98 (44.9)	96/192 (50.0)	49.2
D15S822	36/67 (53.7)	31/69 (44.9)	68/138 (49.3)	50.1
D15S156	30/53 (56.5)	29/56 (51.8)	66/123 (53.7)	50.3
D15S165	23/46 (50.0)	27/56 (48.2)	50/102 (49.0)	50.6
D15S1232	38/72 (52.8)	37/70 (52.9)	77/146 (52.7)	49.2
ACTC	30/63 (47.6)	31/67 (46.3)	61/130 (46.9)	47.9

TABLE II. Parental Allele Frequency Distribution and Transmission Disequilibrium Test Results for Locus GABRB3-155CA2 for 286 Offspring With Autism

Allele size	N (%)	TR ^a	NT ^b	X ²
92	179 (35.2)	97	111	0.94
94	45 (8.9)	44	48	0.17
96	4 (0.8)	4	4	0.00
98	0 (0.0)	—	—	—
100	16 (3.1)	13	19	1.13
102	16 (3.1)	17	10	1.81
104	80 (15.7)	73	78	0.17
106	50 (9.8)	44	41	0.11
108 ^c	75 (14.8)	79	62	2.05
110	17 (3.3)	18	17	0.03
112	10 (2.0)	17	4	8.05
114	8 (1.6)	5	9	3.56
116	7 (1.4)	5	9	1.15
118	1 (0.2)	—	—	—

^aTR is transmitted alleles.^bNT is nontransmitted alleles.^cAllele "108" corresponds to allele "103" from the study of Cook et al. (1998).

context of the difficulties of detecting genes in a complex psychiatric disorder such as autism. Segregation analysis [Jorde et al., 1991] and twin studies [Bailey et al., 1995; Folstein and Rutter, 1977; Ritvo et al., 1985; Steffenburg et al., 1989;] provide strong evidence that autism is not caused by a single major gene. Because there do not appear to be genes of major effect, differences between autism linkage studies could be accounted for by statistical fluctuations suggesting linkage where none is present, or linkage is present but the gene effect is weak or applies to only a small subset of patients. If the latter is the case, very large samples will be necessary to provide compelling linkage evidence. Alternatively, if chromosome 15q-associated cases of autism could be identified by distinct clinical or other phenotypic characteristics, creating a homogeneous collection of such families might provide greater power to detect linkage in this region.

Despite our strong negative results on chromosome 15q11-q13, this region should still be viewed as an area of strong interest, given the number of reports showing association between autism and cytogenetic abnormalities in this region. However, our results clearly indicate that, for the majority of affected sib-pairs, there is no major gene in this region causing autism.

ACKNOWLEDGMENTS

This research was partially supported by families and friends of individuals with autism. We wish to express our gratitude to the families who have been our partners in this research.

REFERENCES

- Arrieta I, Lobato MN, Martinez B, Criado B. 1994. Parental origin of Robertsonian translocation (15q22q) and Prader-Willi syndrome associated with autism. *Psychiatric Genet* 4:63-65.
- Bailey A, LeCouteur 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.
- Baker P, Piven J, Schwartz S, Patil S. 1994. Duplication of chromosome 15q11-13 in two individuals with autistic disorder. *J Autism Dev Disord* 24:529-535.
- Bundey S, Hardy C, Vickers S, Kilpatrick M, Corbett J. 1994. Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. *Dev Med Child Neurol* 36:736-742.
- Christian SL, Robinson WP, Huang B, Mutirangura, Line MR, Nakao M, Surti U, Chakravarti A, Ledbetter DH. 1995. Molecular characterization of two proximal deletion breakpoint regions in Both Prader-Willi and Angelman syndrome patients. *Am J Hum Genet* 57:40-48.
- Cook Jr EH, Courchesne R, Lord C, Cox NJ, Yan S, Lincoln A, Haas R, Courchesne E, Leventhal BL. 1997a. Evidence of linkage between the serotonin transporter and autistic disorder. *Mol Psychiatry* May 2:3, 247-50.
- Cook Jr EH, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, Courchesne E. 1997b. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am J Hum Genet* 60:928-934.
- Cook Jr EH, 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 15q1-13 markers. *Am J Hum Genet* 62:1077-1083.
- Crolla J, Harvey J, Sitch F, Dennis N. 1995. Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. *Hum Genet* 95:161-170.
- Flejtner WL, Bennet-Baker PE, Ghaziuddin M, McDonald M, Sheldon S, Gorski JL. 1996. cytogenetic and molecular analysis of inv dup(15) chromosomes observed in two patients with autistic disorder and mental retardation. *Am J Medical Genet* 61:182-187.
- Folstein SE, Rutter M. 1977. Infantile autism: a genetic study in 21 twin pairs. *J Child Psychol Psychiatry* 18:297-321.
- Gillberg C, Steffenburg S, Wahlstrom J, Gillberg I, Sjostedt A, Martinsson T, Liedgren S et al. 1991. Autism associated with marker chromosome. *J Am Acad Child Adolesc Psychiatry* 30:489-494.
- Hauser ER, Boehnke M, Guo SW, Risch N. 1996. Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 13:117-137.
- Hotof M, Bolton P. 1995. A case of autism associated with partial tetrasomy 15. *J Autism Dev Disord* 25:41-49.
- International Molecular Genetic Study of Autism Consortium. 1998. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum Mol Genet* Mar 7:3, 571-578.
- Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, McMahon WM, Petersen B, Jensen WR, Mo A. 1991. Complex segregation analysis of autism. *Am J Hum Genet* 49:932-938.
- Lazzeroni LC, Lange K. 1998. A conditional inference framework for extending the transmission-disequilibrium test. *Hum Hered* 48:67-81.
- Leana-Cox J, Jenkins L, Palmer C, Plattner R, Shepard L, Flejtner W, Zackowski J, Tsien F, Schwartz S. 1994. Molecular cytogenetic analysis of inv dup(15) chromosomes, using probes specific for the Prader-Willi/Angelman syndrome region: clinical implications. *Am J Hum Genet* 54:748-756.
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J. 1989. Autism Diagnostic Interview: a standardized investigator-based instrument. *J Aut Dev Disord* 19:363-387.
- Lord C, Pickles A, McLennan J, Rutter M, Bregman J, Folstein S, Fombonne E, Leboyer M, Minshew N. 1997. Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Aut Dev Disord* 27:501-517.
- Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mahwood L, Schopler E. 1989. Autism Diagnostic Observation Schedule; a standardised observation of communicative and social behavior. *J Aut Dev Disord* 19:185-212.
- Lord C, Rutter M, LeCouteur A. 1994. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Aut Dev Disord* 24:659-685.
- Martin ER, Kaplan NL, Weir BS. 1997. Tests for linkage and association in nuclear families. *Am J Hum Genet* 61:439-448.
- Mutirangura A, Ledbetter SA, Kuwano A, Chinault AC, Ledbetter DH. 1992. Dinucleotide repeat polymorphism at the GABA_A receptor $\beta 3$ (GABRB3) locus in the Angelman/Prader-Willi region (AS/PWS) of chromosome 15. *Hum Mol Genet* 1:67.
- Pericak-Vance MA, Wolpert CM, Menold MM, Bass MP, Delong GR, Beaty LM, Zimmerman A, Potter N, Gilbert JR, Vance JM, Wright HH,

Abramson RK, Cuccaro ML. 1997. Linkage evidence supports the involvement of chromosome 15 in autistic disorder (AUT). *Am J Hum Genet* 40(Suppl A):208.

Rineer S, Finucane B, Simon EW. 1998. Autistic symptoms among children and young adults with isodicentric chromosome 15. *Am J Med Genet* 81:428-433.

Risch N. 1990. Linkage strategies for genetically complex traits II. The power of affected relative pairs. *Am J Hum Genet* 46, 229-239.

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* 142:64-77.

Robinson W, Binkert F, Gine R, Vazquez C, Muller W, Rosenkranz W, Schinzel A. 1993. Clinical and molecular analysis of five inv dup(15) patients. *Dur J Hum Genet* 1:37-50.

Schinzel AA, Brecevic L, Bernasconi F, Binkert G, Berthet F, Wuilloud A, Robinson WP. 1994. Intrachromosomal triplication of 15q11-q13. *J Med Genet* 31:798-803.

Schroer RJ, Phelan MC, Michaelis RC, Crawford ED, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D, Stevenson RE. 1998. Autism and maternally derived aberrations of chromosome 15q. *Am J Med Genet* 76:327-336.

Smalley SL. 1997. Genetic influences in childhood-onset psychiatric disorders: autism and attention-deficit/hyperactivity disorder. *Am J Hum Genet* 60:1276-1282.

Spielman R, 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.

Spiker D, Lotspeich L, Kraemer HC, Hallmayer J, McMahon W, Peterson PB, Wong DL, Dimiceli S, Ritvo E, Cavalli-Sforza LL, Ciaranello RD. 1994. genetics of autism: characteristics of affected and unaffected children from 37 multiplex families. *Am J Med Genet* 54:27-35.

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* 30:405-416.

Sutcliffe J, Jiang Y-H, Galjaard R-J, Matsura T, Fang P, Kubota T, Christian S, Bressler J, Cattanaach B, Ledbetter DH, Beaudet AL. 1997. The E6-Ap ubiquitin-protein ligase (UBE3A) gene is localised within a narrowed Angelman syndrome critical region. *Genome Res* 7:367-377.

A Genomic Screen of Autism: Evidence for a Multilocus Etiology

Neil Risch,¹ Donna Spiker,² Linda Lotspeich,² Nassim Nouri,¹ David Hinds,¹ Joachim Hallmayer,³ Luba Kalaydjieva,⁴ Patty McCague,² Sue Dimiceli,² Tawna Pitts,¹ Loan Nguyen,¹ Joan Yang,¹ Courtney Harper,¹ Danielle Thorpe,² Saritha Vermeer,² Helena Young,² Joan Hebert,¹ Alice Lin,¹ Joan Ferguson,² Carla Chiotti,² Susan Wiese-Slater,² Tamara Rogers,⁴ Boyd Salmon,⁴ Peter Nicholas,⁵ P. Brent Petersen,⁵ Carmen Pingree,⁵ William McMahon,⁵ Dona L. Wong,² L. Luca Cavalli-Sforza,¹ Helena C. Kraemer,² and Richard M. Myers¹

Departments of ¹Genetics and ²Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford; ³Centre for Clinical Research in Neuropsychiatry, Graylands Hospital/University of Western Australia, and ⁵Centre for Human Genetics, Edith Cowan University, Perth; and ⁴Department of Psychiatry, University of Utah, Salt Lake City

Summary

We have conducted a genome screen of autism, by linkage analysis in an initial set of 90 multiplex sibships, with parents, containing 97 independent affected sib pairs (ASPs), with follow-up in 49 additional multiplex sibships, containing 50 ASPs. In total, 519 markers were genotyped, including 362 for the initial screen, and an additional 157 were genotyped in the follow-up. As a control, we also included in the analysis unaffected sibs, which provided 51 discordant sib pairs (DSPs) for the initial screen and 29 for the follow-up. In the initial phase of the work, we observed increased identity by descent (IBD) in the ASPs (sharing of 51.6%) compared with the DSPs (sharing of 50.8%). The excess sharing in the ASPs could not be attributed to the effect of a small number of loci but, rather, was due to the modest increase in the entire distribution of IBD. These results are most compatible with a model specifying a large number of loci (perhaps ≥ 15) and are less compatible with models specifying ≤ 10 loci. The largest LOD score obtained in the initial scan was for a marker on chromosome 1p; this region also showed positive sharing in the replication family set, giving a maximum multipoint LOD score of 2.15 for both sets combined. Thus, there may exist a gene of moderate effect in this region. We had only modestly positive or negative linkage evidence in candidate regions identified in other studies. Our results suggest that positional cloning of susceptibility loci by linkage analysis may be a formidable task and that other approaches may be necessary.

Introduction

Autism (MIM 209850) is a pervasive neurodevelopmental disorder with symptoms usually apparent during the first 3 years of life. It is characterized by a triad of features, including limited or absent verbal communication, a lack of social reciprocity or responsiveness, and limited, stereotypical, and ritualized interests and behaviors. Autism is the most severe of the disorders now characterized as “pervasive developmental disorders” (PDD) (MIM 209850), which also include Asperger syndrome (MIM 209850) and PPD that is not otherwise specified (PDD-NOS).

Autism was first typified as a distinct clinical entity by Kanner (1943). The etiology of this syndrome has been debated ever since, with various proponents favoring behavioral, environmental, dietary, viral/immunologic, autoimmune, or genetic theories. On the basis of the prevalence studies that have been performed, the indication is that autism is generally rare, with a mean prevalence of 1/2,500 children (Smalley et al. 1988).

The infrequency of familial cases of autism probably led some early theorists to doubt an inherited component in this disorder (Hanson and Gottesman 1976). For example, for a typical rare recessive disease, multiple-incidence sibships are not unusual, since the recurrence risk in sibships is 25%. The recurrence risk of autism in sibships is far below that 25% figure. However, with the completion of several systematic family studies, it is now possible to provide a consistent sibling recurrence figure of 2%–6%, which is substantially higher than the population prevalence of 1/2,500, or 0.04% (Smalley et al. 1988; Bolton et al. 1994). The most recent summary of family studies provided an overall sibling recurrence risk of 2.2% (13/597), with a 95% confidence limit of 1.1%–3.3% (Szatmari et al. 1998). Thus, although familial cases are rare, they are substantially more frequent than would be predicted to occur by chance.

Received December 23, 1998; accepted for publication May 19, 1999; electronically published July 9, 1999.

Address for correspondence and reprints: Dr. Neil Risch, Department of Genetics, M322, Stanford University School of Medicine, Stanford, CA 94305-5120. E-mail: Risch@lahmed.stanford.edu

© 1999 by The American Society of Human Genetics. All rights reserved.
0002-9297/99/6502-0025\$02.00

Further supporting a genetic etiology in autism are the results of twin studies, which have documented a dramatically higher concordance rate in MZ twins than in DZ twins (Folstein and Rutter 1977). Summarizing the largest and most systematic of these studies (Steffenburg et al. 1989; Bailey et al. 1995), we can derive MZ-twin and DZ-twin probandwise concordances of 81% (47/58) and 0%, respectively, for autism. The latter figure is likely to be an underestimate as a result of small sample sizes, and the true DZ-twin concordance is likely to be similar to the nontwin-sib rate of 2%–6%. Using a 3% rate gives an MZ-twin concordance:DZ-twin concordance ratio (MZ:DZ concordance ratio) of ~25-fold. We note that, whereas the sib recurrence is based on numerous studies and a sizable total sibling population ($n = 597$), the MZ-twin concordance rate is based on a much smaller sample and thus is more susceptible to statistical fluctuation. Nonetheless, it is likely that the MZ:DZ concordance ratio in autism is quite large.

Autistic features have also been described in individuals with well-characterized genetic disorders, most notably fragile X syndrome (FMR1 [MIM 309550]) (Feinstein and Reiss 1998) and tuberous sclerosis (TSC1 [MIM 191100]) (Smalley 1998). In addition, case studies have been reported that show associations between autism and a variety of isolated chromosomal translocations, inversions, and deletions (Gillberg 1998). These findings provide evidence that genetic abnormalities can, in some cases, produce behaviors typical of autism, but they account for only a small percentage of cases and do not explain the familial aggregation of autism.

Autism does not follow a simple mode of inheritance. Genealogical studies suggested an increased risk primarily in close relatives (sibs) (Jorde et al. 1990), and a reported segregation analysis was most consistent with a polygenic model of inheritance (Jorde et al. 1991). These results suggest that family recurrence is unlikely to be explained by a simple, monogenic mechanism.

The first linkage study in autism was performed by Spence et al. (1985), using 20 classic blood-group and serum-protein markers on 34 multiplex sibships and assuming a recessive model. No significant linkage findings were obtained.

Recent advances in molecular biology have allowed genome surveys for disease-susceptibility loci to be performed on a much larger scale. For example, Hallmayer et al. (1996) examined 35 microsatellite markers on the X chromosome in 38 multiplex autism families. Although they excluded most of the chromosome from carrying a gene of large effect, including the fragile X region, they found modest evidence (maximum LOD score 1.24) for linkage in one region on proximal Xq.

A genome screen of the full genome, using 316 microsatellite markers on 39 multiplex families, has also been reported recently (International Molecular Genetic

Study of Autism Consortium 1998). Initial positive findings were followed up in an additional 60 families. No statistically significant findings were obtained. The most positive finding was a LOD score of 2.53, on chromosome 7q, with the next most positive finding being a LOD score of 1.97, near the telomere of chromosome 16p.

In this report we describe the results of what, to date, is the largest genome screen in autism. We have surveyed 90 multiplex sibships with 97 independent affected sib pairs (ASPs) in an initial analysis of 362 microsatellite markers. Regions with positive results, as well as other candidate regions, were then followed up with additional markers, as well as by analysis of a second set of 49 multiplex sibships containing 50 independent ASPs. A total of 519 markers was analyzed in the first set of families, and 149 of these markers were also analyzed in the second set of families.

Families And Methods

Family Recruitment and Diagnostic Assessments

Families were recruited nationwide for this study, by advertisement with local and national parent support groups and by referrals from clinicians. Initial intake criteria required that at least two siblings have a clinical diagnosis of a PDD. If both the initial telephone intake interview of parents and collected medical records were consistent with a presumptive diagnosis of autism in at least two children, then a follow-up visit by a diagnostician was arranged with the family. During these visits, the children were assessed—both by interview, of one or both parents, using the Autism Diagnostic Interview (ADI) (LeCouteur et al. 1989; Lord et al. 1994, 1997), and by observation of the children, using the Autism Diagnosis Observation Schedule (ADOS) (Lord et al. 1989)—to determine a research diagnosis of autism. The ADI is a scored, semistructured interview of parents that is based on ICD-10 and DSM-IV criteria for the diagnosis of autism. The ADOS is a semistructured instrument, which allows assessment of the child through observation of his or her behavior and which is used to corroborate the results of the ADI interview. Diagnostic assessments (by ADI and ADOS) were videotaped and subjected to independent reliability checks by other trained interviewers (typically, the senior clinical investigators). To be considered affected in the linkage analysis, an individual had to satisfy the prespecified cutoff scores in all three symptom areas of the ADI (social impairment, language and communication impairment, and unusual and restricted interests) as well as have an age at onset of <3 years. In addition, review of the ADOS tape by two or more diagnosticians was used to exclude children who did not show significant impairments in

Table 1
Distribution of Family Structures

GROUP AND NO. OF AFFECTED SIBS	NO. OF		Families
	Unaffected Sibs	Available Parents	
FS1:			
2	0	2	48
2	0	1	15
2	1	2	13
2	1	1	5
2	2	2	2
2	3	1	2
3	0	2	2
3	1	2	1
4	0	2	1
4	1	1	1
FS2:			
2	0	2	34
2	0	1	1
2	1	2	12
2	2	2	1
3	1	2	1

social and communicative reciprocity, irrespective of the results of the ADI interview (e.g., this would lead to the exclusion of some individuals who would have had a clinical diagnosis of PDD-NOS). Families were excluded from subsequent analysis if, on the basis of all available information, there was no consensus that at least two affected children had deficits consistent with a diagnosis of autism. For more details on the diagnostic protocol and reliability of videotapes of the ADI assessments, see the work of Spiker et al. (1994).

Children’s records were reviewed to obtain IQ and mental-age (MA) information. When available, nonverbal IQ scores were obtained from performance subtests of either the Wechsler Scales or the Stanford-Binet Intelligence Scale-4th Edition, from the Leiter International Performance Scale, or from the Merrill-Palmer Scale. When unavailable, scores from such tests as the Stanford-Binet Intelligence Scale-3rd Edition, Slossen Intelligence Scale, or McCarthy Scales were used. For children for whom results of these tests were not available, MAs and ratio IQ scores were derived from nonverbal scales of various developmental instruments (e.g., Daily Living Scale of the Vineland Adaptive Behavior Scale, the Bayley Mental Scale, and Developmental Inventory II).

For inclusion in the linkage analysis, we also required that at least one affected child have both an MA of ≥18 mo and a nonverbal IQ ≥30. This led to the exclusion of some additional families (N = 6) in which all children were severely impaired (MA <18 mo and IQ <30).

Two-Stage Analysis

Families were analyzed in two stages. The initial stage involved genotyping of family set 1 (FS1), which in-

cluded 90 multiplex families. These families comprise a total of 187 affected and 30 unaffected sibs; 67 families have two parents available, and the remaining 23 have one parent. The precise breakdown of family structures is given in table 1. Among the affected sibs, the male:female ratio is 3.6:1, whereas in the unaffected sibs it is 0.8:1.

Subsequent to the initial genome screen performed on FS1, a follow-up study was conducted on a second family set (FS2), which comprises a total of 49 multiplex families, ascertained in a manner similar to that used with FS1. These families contain a total of 99 affected and 15 unaffected sibs; 48 of the 49 families have both parents available, whereas the single remaining family has one parent. The precise breakdown, by family structure, is given in table 1. The male:female ratio in the affected sibs is 3.0:1, versus 0.7:1 in the unaffected sibs.

Laboratory Procedures

Blood was collected from all affected individuals and, if available, from their parents and unaffected siblings. Blood samples were drawn into Vacutainer sodium heparin (green cap) tubes. The lymphoblasts were isolated and immortalized with fresh or frozen virus stocks, by standard Epstein-Barr-virus transformation protocols (Anderson and Gusella 1984). The DNA was extracted from both whole blood and immortalized lymphoblast cell lines, by standard proteinase K-digestion and salt-ing-out procedures.

Genotyping was performed in two laboratories in-

Table 2
Markers with Highest Sharing in ASPs and DSPs in Initial Genome Screen

Marker	No. (%) Sharing	Z Score
ASPs:		
D1S1631	80/122 (65.6)	3.44
D13S779	64/101 (63.4)	2.69
D9S282	60/95 (63.2)	2.66
D1S1609	79/130 (60.8)	2.46
D3S2418	60/96 (62.5)	2.45
D17S1298	51/81 (63.0)	2.33
D15S652	79/132 (59.8)	2.26
D10S1412	53/86 (61.6)	2.16
D1S534	81/137 (59.1)	2.14
DSPs:		
GATA145D0	34/49 (69.4)	2.71
D16S403	42/63 (66.7)	2.65
D18S844	39/60 (65.0)	2.32
D16S3253	29/43 (67.4)	2.29
D7S1819	35/54 (65.3)	2.18
D4S2361	32/49 (65.3)	2.08
D11S2371	26/39 (66.7)	2.08
D4S2366	34/53 (64.2)	2.06
D20S482	31/48 (64.6)	2.02
D16S769	32/50 (64.0)	1.98

dependently, according to somewhat different procedures. In one lab, PCR was performed, by Perkin-Elmer 9700 thermal cyclers, in a 10- μ l volume in 96-well plates and included 30 ng of genomic DNA, 67 mM Tris-HCl pH 8.8, 200 μ M of each dNTP, 16 mM $(\text{NH}_4)_2\text{SO}_4$, 0.01% Tween-20, 1.5–3 mM MgCl_2 , 0.2 μ M of each primer, and 0.2 units of AmpliTaq polymerase (Perkin-Elmer). The forward primers for each marker were fluorescently labeled with FAM, HEX, or TET dyes. The reaction was first denatured for 75 s at 95°C; then, 35 cycles of 45 s at 95°C, 45 s at 57°C, and 60 s at 72°C were performed, followed by a final elongation step of 7–12 min at 72°C. PCR products were pooled from ~7–15 markers that were separable on the basis of size and dye color and were run on ABI Prizm 377 DNA sequencers. Tamara-labeled DNA internal-size standards were run in each lane, and the products were resolved by the GeneScan and Genotyper software packages from Applied Biosystems. Three CEPH control DNAs from individuals 1331-01, 1331-02, and 1347-02 were used to verify the sizes of PCR products for each marker.

In the second lab, PCR assays were performed in 8- μ l reactions containing 10 ng of genomic DNA; 2.5 mM of each dNTP; 0.2–0.6 mM of each primer pair labeled with the fluorescent dyes FAM, HEX, or TET; 0.8 μ l of 10 \times buffer (Perkin-Elmer); 1.5–2.5 mM MgCl_2 , and 0.2 units of AmpliTaq Gold Polymerase (Perkin-Elmer). PCR assays were performed in an Applied Biosystems 9600 Thermocycler by denaturation for 10 min at 94°C, 15 s at 60°C, and 15 s at 72°C. Another 20 cycles of 30 s at 94°C, 15 s at 65°C, and 15 s at 72°C were performed, and these were followed by a final extension step of 10 min at 72°C. Fluorescently labeled markers were analyzed on an Applied Biosystems 373 Genetic Analyzer, by Genescan and Genotyper software. In both labs, polymorphic bands were scored, and alleles were assigned to the pedigree members, by researchers who were blind to affection status.

Genetic Markers

Marker set 1 (MS1).—The genetic analysis was performed in two stages. The first stage consisted of a genome screen on the 90 multiplex FS1 families and used a total of 362 microsatellite markers. These were derived from version 8.0 of the Marshfield fluorescently labeled genome screening set (Center for Medical Genetics, Marshfield Medical Research Foundation) and were obtained from Research Genetics. Of these markers, 346 are autosomal, 14 are X linked, and 2 are pseudoautosomal. If a sex-averaged total autosomal map length of 3,500 cM is assumed, the average spacing between markers for the autosomal component is 10 cM; if a length of 180 cM for the X chromosome is assumed, the average spacing on that chromosome is 12 cM. Of

the intermarker intervals, 28 have an interval length >15 cM, although only 7 are >20 cM; the largest gap is 23 cM.

Marker set 2 (MS2).—Additional microsatellite markers were obtained from Research Genetics and were analyzed subsequent to the initial genome screen. The choice of markers was based on (1) saturation of regions giving positive linkage evidence in the analysis of FS1 with MS1; (2) follow-up of specific regions, on the basis of either candidate loci (chromosome regions 6p and 15q) or the genome screen of the International Molecular Genetic Study of Autism Consortium (1998) (chromosome region 7q); and (3) addition of markers in regions that had large gaps in the initial screen with MS1 and that did not have significantly negative linkage evidence in the preliminary analysis. In total, 157 additional markers were typed, which constituted MS2. All of these markers were typed in FS1; 89 of these were also typed in FS2. In addition, 60 of the original MS1 markers were genotyped in FS2 in the follow-up, for reasons (1) and (2) above. Thus, in total, 519 markers were typed in FS1, and 149 were typed in FS2.

Statistical Analysis

Linkage analysis.—Because all siblings had at least one parent available and most had both parents available, the statistical analyses were based entirely on identity by descent (IBD), obviating the need to use allele frequencies of the parents. When one parent is missing, a sib pair can be scored for IBD with the included parent, without bias, provided that the typed parent is heterozygous and that each of the two sibs inherits from the untyped parent an allele that is distinct from the alleles of the typed parent. Also, sharing from the untyped parent was scored, when possible, by reconstruction of the genotype of the missing parent, by use of one affected sib and all unaffected sibs (in the ASP analysis). If the genotype of the parent can be so reconstructed, then unbiased scoring can be performed by comparison of the affected sib with each remaining affected sib, who was not used for reconstruction.

Sib-pair analyses were based on fully independent pairs. When a family contained three or more affected sibs, independent pairs were formed between the first sib and each remaining sib (giving $n - 1$ pairs for a family with n affected sibs). In the 90 multiplex FS1 families, there are thus a total of 97 independent ASPs; in the 49 multiplex FS2 families, there are 50 independent ASPs (see table 1).

Initially, the number of alleles shared versus the number of alleles unshared, at each marker across sib pairs, was evaluated and compared to a χ^2 distribution. Under the null hypothesis of no linkage, the expectation is that sharing will be 50%. Subsequently, multipoint sib-pair

analysis was performed to extract the maximum information from the genetic-marker data. This analysis is parameterized in terms of the value λ_s , the sibling recurrence ratio, which translates into specific values for ASP IBD. For a genetic model with no dominance variance, the probability z_0 of an ASP sharing no alleles IBD is $.25/\lambda_s$, the probability z_1 of an ASP sharing one allele IBD is $.50$, and the probability z_2 of an ASP sharing two alleles IBD is $.25(2 - \frac{1}{\lambda_s})$; the overall sharing is $\gamma = .25(3 - \frac{1}{\lambda_s})$. For a multiplicative model which allows dominance variance, again $z_0 = \frac{25}{\lambda_s}$, $z_1 = (\sqrt{\lambda_s} - \frac{1}{2})/\lambda_s$, and $z_2 = (1 - \frac{1}{2}\sqrt{\lambda_s})^2$; the overall sharing is $\gamma = 1 - \frac{1}{2}\sqrt{\lambda_s}$. We use the latter model in all the multipoint analyses that we present, although, for modest gene effects, results are nearly identical in either model.

Multipoint LOD scores can be calculated for a fixed model (i.e., fixed value of λ_s); this approach allows for exclusion mapping. For a fixed value of λ_s , negative LOD scores can be obtained if the model gives a poorer representation of the marker data than does no gene effect. We use the LOD-score criterion of -2 for exclusion of a locus with a given λ_s value; this provides a conservative exclusion level (Hauser et al. 1996). Alternatively, the likelihood of the marker data can be maximized at each point, as a function of the genetic model (λ_s value), to derive a maximum LOD score (MLS) curve. This curve never drops to <0 , the value obtained when $\lambda_s = 1$.

Linkage-disequilibrium analysis (LD).—We also examined LD, with all 519 autosomal and X-linked loci, using a transmission/disequilibrium test (TDT) (Spielman et al. 1993). For families with both parents, we scored directly transmissions to all affected children from heterozygous parents. As in the linkage analysis, we can still score children from single parents, provided that the parent is heterozygous and that the child is heterozygous for alleles different than those in the typed parent. We can also score transmissions from the missing parent when the genotype of that parent can be reconstructed on the basis of the unaffected sibs.

We examined two statistical tests—a “global χ^2 ” (GCS) test, which sums $X = (\text{observed} - \text{expected})^2 / \text{expected}$ across each allele tested, and a “maximum χ^2 ” (MCS) test, which takes the maximum value of the statistic X across all tested alleles. In this case, neither statistic has a formal χ^2 distribution when multiple sibs from the same family are tested, if linkage is present (Spielman et al. 1993). To make our test of LD independent of any linkage present, we calculated the level of significance empirically, using Monte Carlo simulation; we simply randomly inverted or not, with probability 50%, the allele label for the two alleles of heterozygous parents and recalculated the two statistics for each permutation step. The proportion of simulated statistics that exceed the observed value is the derived em-

pirical significance level. This procedure allows us to assess LD, conditional on the degree of sharing of alleles, at a given locus, by sibs from multiplex families (Martin et al. 1997; Lazzaroni and Lange 1998).

Computer program.—All statistical analysis, including the pointwise and multipoint sib-pair analyses and the TDT analyses, were performed with the ASPEX program package. We also used this package to confirm relationships among study subjects and to search for and eliminate unlikely double crossovers in the multipoint analyses, in regions that had dense markers.

Controls

In the course of this study, a number of controls were employed, to determine the degree to which potential biases might be confounding our results. After ~60 markers were run in FS1, we examined the distribution of IBD for all sib pairs, to confirm the presumed relationships, identify any mixup of samples, and identify any potential half-sibs. All families segregated appropriately according to presumed relationships (no half-sibs were identified), with the exception of two sib pairs, in which IBD was nearly 100%. These pairs were reported to us to be DZ twins, whereas our genotyping determined them to be MZ. We left these twin pairs in the analysis, both as a positive control and to determine the genotyping-error rate. The lab remained blind to the identity of these pairs.

As another control for our genetic analyses, we also included unaffected sibs in our molecular and statistical analyses. FS1 contains 30 unaffected sibs, which allowed us to construct 51 independent affected-unaffected sib pairs, or DSPs. These were derived by pairing, in each family, one unaffected sib with each affected sib. Remaining unaffected sibs were used for reconstruction of the genotypes of untyped parents. This selection scheme allowed us to tally IBD for these pairs as fully independent. Because of the typically modest recurrence risk of autism in sibships (Smalley et al. 1988), we assume that the DSPs should deviate little from the null expectation of 50% sharing—and that they thus can serve as a useful negative control.

Results

Family Recruitment

The first stage of analysis included the 90 multiplex FS1 families comprising 187 affected offspring, 30 unaffected sibs, and 157 parents. Two additional families with affected MZ twins (originally thought to be DZ) were also included. At the time of ADI administration, the age range of the affected siblings was 2.9–40.9 years (mean 9.8 years, SD 7.3 years, 25th–75th percentile 4.9–11.3 years). MA and IQ estimates were determined

on the basis of available diagnostic evaluations and school records. The mean nonverbal IQ was 66 (SD 28, range 16-160). One hundred eight subjects (57%) had IQ scores <70. The mean MA was 68 mo (SD 55 mo, range 13-373 mo). There were no families in which all affected children had IQ scores <30. In one family, one affected child had an MA of <18 mo.

The second set of study subjects, obtained with similar recruitment and inclusion/exclusion criteria, comprised a total of 99 affected children, 15 unaffected sibs, and 97 parents, from 49 families. For this group, at the time of ADI administration, the age range of affected subjects was 2.8-25.3 years (mean 6.6 years, SD 3.5 years, 25th-75th percentile 4.3-7.6 years). The mean nonverbal IQ was 61 (SD 23, range 15-124). Sixty-two subjects (63%) had IQ scores <70. Mean MA was 47 mo (SD 23 mo, range 13-150 mo). Again, there were no families in which all children had IQ scores <30, and there were three children with MA <18 mo.

The unaffected children were presumed to be developing normally, on the basis of parental report; all of these children were in regular educational programs and schools. None had any clinical diagnoses, according to parental report, nor, during telephone intake interviews, did parents report any developmental concerns. Children with a possible PDD diagnosis based on the ADI interview but who did not meet the ADI criteria for autism were categorized as being of uncertain status and were not included in this analysis. Eighty-eight percent of our sample was white (of varied European and Middle Eastern origin), 5% was black, 3% was Hispanic, and 4% was Asian. The vast majority of affected subjects were living with parents, and a small number were in residential living facilities.

FS1: Pointwise IBD Scores

A total of 346 autosomal, 14 X-linked, and 2 pseudoautosomal microsatellite loci were studied in FS1, which contained 97 independent ASPs and 51 independent DSPs. Tallying the total number of alleles shared across all 346 autosomal markers gives 19,902/38,572, or 51.6%, for the ASPs and 9,344/18,386, or 50.8%, for the DSPs. The slight increase in IBD, to >50%, in the DSPs suggests the possibility of some modest scoring bias toward allele sharing in sibships.

Of greatest interest is the right-hand tail of the observed IBD distribution in the ASPs and DSPs. For each marker, we calculated a Z score, $Z = (s - u) / \sqrt{s + u}$, where s is the number of shared alleles and u is the number of unshared alleles. Under the null hypothesis of no linkage, Z should have a normal distribution with mean 0 and variance 1. When a threshold of 1.96 ($P = .025$ for a one-sided Z test) is assumed, there are 9 loci in the ASP group that exceed this threshold and

10 loci in the DSP group that exceed this threshold, which are listed in table 2. These numbers are not different from the number expected ($.025 \times 360 = 9.0$) under the null hypothesis of no linkage. The most significant locus, in either group, was D1S1631, for the ASPs, with a sharing of 65.6% and a Z score of 3.44. The remaining loci in the ASP group were found on chromosomes 1 (two additional loci), 3, 9, 10, 13, 15, and 17. The locus on chromosome 15, D15S652, is telomeric of and not linked to the candidate region, on proximal 15q, associated with the inverted duplication found in some autistic subjects (Gillberg et al. 1991; Cook et al. 1997). Aside from locus D1S1631, there appears to be no substantial difference between the tail Z scores for the ASP group and those for the DSP group.

We also examined the full distribution of Z scores, for both the ASPs and the DSPs. These distributions are given in figure 1. For convenience, we have plotted each Z score as a function of $\Phi^{-1}[(R - \frac{1}{2})/S]$, where R is the ordered rank of that Z score, $S = 346$ is the total number of autosomal markers, and Φ is the cumulative normal-distribution function. Under the null hypothesis (no linkage), this curve should fit the line $y = x$. In figure 1, the line $y = x$ is depicted as the thin unbroken line; the thick dotted line represents DSPs, and the thick dashed line represents ASPs. As can be seen in the figure, both the ASP and DSP lines run parallel but above the line $y = x$, but more so in the case of the ASP line. This indicates that the entire distribution of Z scores for the ASPs is shifted upward, compared with that for the DSPs, and this shift is most noticeable at the left end of the distribution.

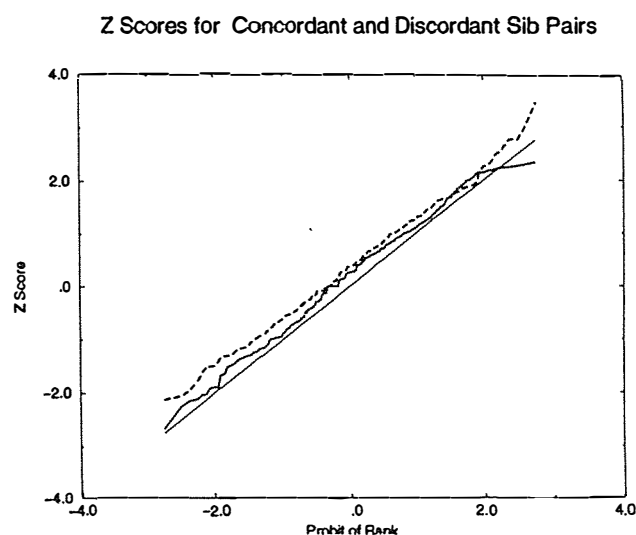


Figure 1 Observed distribution of Z scores for sharing at 346 autosomal marker loci in ASPs and DSPs, compared with the expected distribution under no linkage. The thin unbroken line represents no linkage; the thick dashed line represents sharing for ASPs; the thick dotted line represents sharing for DSPs.

bution (i.e., deficiency of negative sharing). The final "blip" at the right end of the ASP line is due to marker D1S1631.

Observed versus Expected Distribution Of IBD Scores

As discussed above, family and twin studies suggest that autism is a highly heritable disorder. However, these data alone cannot reveal the precise mode of inheritance. The very high MZ:DZ concordance ratio, taken at face value, is strongly suggestive of multiple interacting loci (epistasis), with no single locus having a substantial effect. The results of our initial genome screen would seem to support this conclusion. However, given the high recurrence-risk ratio for sibs ($\lambda_s = 75$ –100) and the high (15–25) MZ:DZ concordance ratio, one may reasonably ask what the distribution of IBD scores would look like for various genetic models consistent with those values. For the purpose of computational simplification, we assume a multiplicative epistatic model (Risch 1990) with K loci of equal effect. Given the 15–25 MZ:DZ concordance ratio, models with as few as two loci can be excluded. For a model with K loci, the i th locus will have an individual value of $\lambda_{si} = (\lambda_s)^{\frac{1}{K}}$ for $i = 1, \dots, K$ (Risch 1990). The ASP sharing y_i at each of these loci will be $y_i = 1 - .5(\lambda_{si})^{-\frac{1}{K}} = 1 - .5(\lambda_s)^{-\frac{1}{K}}$ according to a multiplicative model. For example, for $K = 3$, $y_i = 75.7\%$; for $K = 5$, $y_i = 67.5\%$; for $K = 10$, $y_i = 59.7\%$; and, for $K = 20$, $y_i = 55.1\%$. If $\lambda_s = 150$, then the corresponding y_i values for $K = 3, 5, 10$, and 20 are only slightly higher—namely, 78.3%, 69.7%, 61.1%, and 55.9%, respectively. Similar numbers can be obtained by parameterization in terms of the MZ:DZ concordance ratio, which we here denote as " λ_{MD} ." According to a multiplicative model, for K contributing loci of equal effect, the total λ_{MD} is $(\lambda_{MDi})^K$, where λ_{MDi} is the ratio contributed by the i th locus. Furthermore, the expected ASP IBD sharing for a locus with value λ_{MDi} is given simply by $y_i = \frac{1}{2} \sqrt{\lambda_{MDi}}$. Since, in a model with K loci of equal effect, $\lambda_{MDi} = (\lambda_{MD})^{\frac{1}{K}}$ for $i = 1, \dots, K$, we can calculate the expected sharing, under the assumption that $\lambda_{MD} = 25$, for $K = 3, 5, 10$, and 20 as being 85.5%, 69.0%, 58.7%, and 54.2%, respectively. It can be seen that these numbers are larger than those based on the sib risk ratio for smaller K values (i.e., <5) but that they begin to converge at $K > 5$.

Certainly our data are not consistent with a three-locus model, since we do not have three regions with sharing approaching 70%. Recombination will tend to decrease sharing away from these high values, but, given our map density, it is unlikely that the sharing would be significantly diminished. In fact, we can calculate systematically, making some simplifying assumptions, what the distribution of sharing (and associated Z scores) should be under these alternative genetic models.

Consider the distribution of Z scores derived from a genome screen. For a given marker, Z is defined, as above, as $(s - u)/\sqrt{s + u}$. Under the null hypothesis of no linkage, Z is approximately normally distributed, with mean 0 and variance 1. Now consider a genetic model of K disease loci, each of which has sharing y_i . We assume that the loci are unlinked. Assume a genetic map with a marker every w cM. On average, the distance of a disease locus to the closest marker will be $\frac{w}{4}$; the distances to remaining markers, in order of distance, are $\frac{3w}{4}, \frac{5w}{4}, \frac{7w}{4}$, etc. For a marker at recombination distance r from a disease locus, the sharing will be $y(r) = [r^2 + (1 - r)^2]y + 2r(1 - r)(1 - y)$. Thus, assuming a map function relating recombination fraction r to map distance w —for example, $r = f(w)$ —we can calculate the expected values of sharing y for markers surrounding a disease locus.

In the present analysis, we assume the Kosambi mapping function $r = f(w) = (1 - e^{-4w})/2(1 + e^{-4w})$. As in our screen, we assume a map density of 10 cM, with a total map length of 3,500 cM. For each of K putative disease loci, we estimate sharing $y(r)$ at neighboring loci at distances 2.5 cM, 7.5 cM, 12.5 cM, ..., up to 47.5 cM, using the Kosambi map function and the $y(r)$ formula given above.

For the model with K loci of equal effect, there are K loci each with value $y(r)$, where $y(r)$ ranges over the 10 values based on the 10 different marker distances. In our genome screen of ASPs, the average total number of transmissions scored at a locus was 110 (i.e., $s + u$, in the notation above). Therefore, for the K markers with expected sharing $y(r)$, we used sampling from a normal distribution with mean $y(r)$ and variance $v(r) = y(r)[1 - y(r)]/110$ to obtain the expected distribution of their observed sharing. We selected the K points from this distribution to be equally spaced to preserve the mean $y(r)$ and variance $v(r)$. For example, for $K = 5$ we chose the five points $y_1(r) = y(r) - \sqrt{2v(r)}$, $y_2(r) = y(r) - \sqrt{v(r)}/2$, $y_3(r) = y(r)$, $y_4(r) = y(r) + \sqrt{v(r)}/2$, and $y_5(r) = y(r) + \sqrt{2v(r)}$. These five points have sample mean $y(r)$ and variance $v(r)$. For the j th such point, the expected Z score was then derived by taking $Z_j(r) = 2\sqrt{110}[y_j(r) - \frac{1}{2}] = 10.49[2y_j(r) - 1]$. As an example, consider the genetic model with $\lambda_s = 75$, $K = 5$, and the third-closest marker (12.5 cM). For this case, $y = .675$, $r = .122$, $y(r) = .600$, $y_1(r) = .534$, $y_2(r) = .567$, $y_3(r) = .600$, $y_4(r) = .633$, and $y_5(r) = .666$; $Z_1(r) = .71$, $Z_2(r) = 1.41$, $Z_3(r) = 2.10$, $Z_4(r) = 2.79$, and $Z_5(r) = 3.48$.

In this fashion, Z scores for 10K markers were derived. For the remaining 346 – 10K markers, we distributed their Z scores as above but according to a normal distribution with mean 0 and variance 1. We have calculated these expected distributions, assuming that $\lambda_s = 75$ and $K = 3$ –20. The results for $K = 3, 5, 10$,

and 20 are given in figure 2. As in figure 1, we have plotted the Z scores as a function of $\Phi^{-1}[(R - \frac{1}{2})/S]$, where R is the rank of the observed Z score and S is the total number of markers (346). A normal distribution appears as a straight line ($y = x$) on this plot and, in figure 2, is depicted by the thick unbroken line. The expected curve corresponding to $K = 3$ is given by the thin dashed line, for $K = 5$ by the thin dotted line, for $K = 10$ by the thin short-dashed line, and for $K = 20$ by the thin unbroken line. The curve for the actual data is given by the thick dashed line. In calculating the Z scores for the actual data, we used 50.8% for expected sharing under the null hypothesis, instead of 50%, to allow for some potential genotyping bias, as observed in the DSPs. As can be seen in figure 2, the actual data most closely approximate the expectations for the model with $K = 20$. The fact that the thick dashed line lies above the $y = x$ line at the left end of the plot indicates that the excess sharing in ASPs compared with DSPs is due to the entire distribution being shifted to the right, rather than to a small subset of marker loci showing increased sharing. The right end of the plot, although reasonably consistent with $K = 20$, is quite deviant from the distribution expected for $K = 5$ or 10, with the exception of the final marker (D1S1631). For other values of K , the distributions lie between the curves in figure 2. For example, for $K = 15$ the distribution lies approximately midway between the line for $K = 10$ and that for $K = 20$; for $K = 13$ the distribution is somewhat closer to the line for $K = 10$.

To assess goodness of fit of each predicted distribution to the observed distribution, we calculated the statistic $W = \sum_{i=1}^{346} (Z_{i1} - Z_{i0})^2$, where Z_{i1} is the i th (rank-ordered) Z score for the expected distribution and Z_{i0} is the corresponding Z score for the observed distribution. A small value of W indicates a good fit of the observed distribution to the predicted distribution. The corresponding value of W for values of $K = 3, 5, 10, 13, 15, 17$, and 20 is 53.1, 16.7, 5.1, 3.6, 3.1, 2.8, and 2.7, respectively. Thus, models with ≤ 10 loci appear unlikely, whereas those with ≥ 15 seem more plausible. The best fit is given by the model with $K = 20$ loci, although the value of W does not increase much up to $K = 15$. Models specifying a small number of loci (i.e., one or two) with moderate effect ($\lambda_s = 2$) also give a reasonable fit (although not as good as is given by the model of 20 loci with equal effect), provided that the number of residual loci remains large (e.g., ≥ 15). Thus, our results, although consistent with a genetic basis of autism, suggest that the most likely model is one with many (≥ 15) contributing loci, with a possible gene of larger effect on chromosome 1.

In calculating the expected distributions, we have made a number of simplifying assumptions, such as equal spacing and equal polymorphism of the markers.

Observed versus Expected Z Scores for 346 Markers

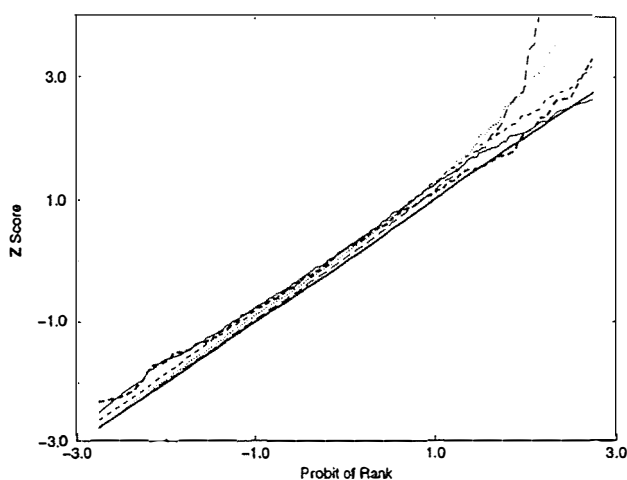


Figure 2 Observed distribution of Z scores for sharing at 346 autosomal marker loci in ASPs, compared with the expected distribution, when a model of $K = 3, 5, 10$, or 20 loci of equal effect is assumed. The thick unbroken line represents no linkage; the thick dashed line represents the observed data for ASPs; the thin long-dashed line represents $K = 3$; the thin dotted line represents $K = 5$; the thin short-dashed line represents $K = 10$; the thin unbroken line represents $K = 20$.

Most significant among these assumptions, however, is that the disease-susceptibility loci are unlinked, which will be unlikely for models with $K \geq 20$. The effect of having linked disease loci is to increase sharing locally in the region of the linked loci—and to decrease it, on average, elsewhere. Thus, the impact will be to create a distribution more similar to the case of fewer disease loci with larger effect. Therefore, our modeling is actually conservative in concluding that the observed distribution is most compatible with a large number of disease loci.

Sharing by Gender of Parent

The observation of genomewide excess sharing and a putative model of many genes of small effect accounting for the total familial effect in autism allow us to posit a further prediction. If, in fact, there are many genes scattered around the genome, then the excess of sharing that we have observed should differ when tallied separately for paternal versus maternal meioses. The reason is that the female linkage map, at 4,200 cM, is ~ 1.5 times the genetic length of the male map, at 2,800 cM. Because there is globally less recombination in males, the expected genetic distance (recombination fraction θ) to a disease-susceptibility locus in males is expected to be less than that in females, and thus the expected allele sharing at a marker linked to the disease locus is expected to be greater when one examines fathers than when one examines mothers.

Table 3
Global Allele Sharing in ASPs and DSPs, by Sex of Parent and Genetic Map-Interval Male:Female Ratio

SEX RATIO (Male:Female)	NO. (%) SHARING	
	Paternal	Maternal
ASPs:		
>1 (85 markers)	2,336/4,560 (51.2)	2,413/4,712 (51.2)
<1 (261 markers)	7,459/14,275 (52.3)	7,694/15,025 (51.2)
Total	9,795/18,835 (52.0)	10,107/19,737 (51.2)
Total (maternal + paternal)	19,902/38,572 (51.6)	
DSPs:		
>1 (85 markers)	1,110/2,160 (51.4)	1,167/2,304 (50.7)
<1 (261 markers)	3,401/6,722 (50.6)	3,666/7,200 (50.9)
Total	4,511/8,882 (50.8)	4,833/9,504 (50.9)
Total (maternal + paternal)	9,344/18,386 (50.8)	

To quantitate this prediction, define $d(0) = y - \frac{1}{2}$, where y is the sharing at a disease locus. Then, at θ distance, the value of y is $y(\theta) = [\theta^2 + (1 - \theta)^2]y + 2\theta(1 - \theta)(1 - y)$, so that $d(\theta) = y(\theta) - \frac{1}{2} = (1 - 2\theta)^2d(0)$. If we assume a uniform distribution for the map distance, w , from a random marker locus to the nearest disease locus, then, over some total map length L , the expected value of d is

$$E(d) = \frac{1}{L} \int_0^L d(0)[1 - 2\theta(w)]^2 dw ,$$

where w is some mapping function relating map distance to θ . For the Haldane mapping function $\theta(w) = (1 - e^{-2w})/2$, $E(d) = d(0)(1 - e^{-4L})/L \cong d(0)/L$, for $L > 50$ cM. Similarly, for the Kosambi mapping function $\theta(w) = (e^{4w} - 1)/2(e^{4w} + 1)$, $E(d) = d(0)[(1 + e^{2L})^{-1} - \log(1 + e^{2L}) + \log 2 - \frac{1}{2}]/L \cong .193d(0)/L$ for $L > 50$ cM. In either case, $E(d) \propto \frac{1}{L}$. Therefore, if female map distances are 1.5 times male map distances and if we let d_m represent the d value for male (paternal) meioses and let d_f represent that for female meioses, then we have $E(d_m)/E(d_f) \propto 1.5$. Therefore, when the observed sharing rate of .508 (from the discordant pairs) is used as the null value, the observed sharing for concordant pairs, .516, gives a sex-averaged value of $E(d) = .008$. Therefore, the prediction discussed above would give sex-specific $E(d)$ values of $E(d_m) = .010$ and $E(d_f) = .006$, or sex-specific sharing values of $y_m = 51.8\%$ and $y_f = 51.4\%$. If we use 50% for the null-hypothesis sharing value, then the sex-specific sharing becomes $y_m = 51.9\%$ and $y_f = 51.3\%$.

The observed total sharing for ASPs, stratified by sex of parent, is $y_m = 9,795/18,835 = 52.0\%$ and $y_f = 10,107/19,737 = 51.2\%$ (table 3), close to the prediction. For comparison, we also examined the parent-specific allele sharing for the DSPs. There was no difference

in overall allele sharing, by sex of parent, for this group (50.8% paternal, 50.9% maternal).

We can make this prediction more precise by noting that the male:female ratio of map distances is not uniform throughout the genome but, rather, is region specific. Some regions (e.g., telomeres) tend to show an excess of recombination in males. Thus, we would expect the observed excess of paternal allele sharing to be restricted to those regions where the female map distances exceed those of the male. To test this hypothesis, we identified the location of each of the 346 marker loci on the Marshfield genetic maps (Center for Medical Genetics, Marshfield Medical Research Foundation). By examining, for each locus, flanking markers up to 10 cM away (sex averaged), we determined whether the genetic length of the flanking interval was greater in males (male:female ratio >1) or in females (male:female <1). As expected, the number of markers with a male:female ratio >1 ($n = 85$) was approximately one-third the number with a male:female ratio <1 ($n = 261$). As predicted, the paternal excess of sharing in ASPs was restricted to markers lying in regions with a male:female ratio <1 (52.3% paternal sharing vs. 51.2% maternal sharing) (table 3). Again, the DSPs showed no such trend, since the paternal and maternal sharing were nearly equal in both groups (table 3).

Multipoint Sib-Pair Analysis in FS1

We performed multipoint sib-pair analysis on the data from FS1, for the 360 autosomal and X-linked markers of MS1. The results are shown in figure 3. In the figure, the MLS curve is represented by the dashed line. There were nine locations that gave an $MLS \geq 1.0$: chromosomes 1p (MLS 1.87), 1q (MLS 1.19), 7p (MLS 1.00), 11p (MLS 1.25), 13q (MLS 1.49), 15q (MLS 1.75), 17p (MLS 1.30), 18q (MLS 1.00), and 20p (MLS 1.09). The MLS on chromosome 1p was diminished compared with

the significance obtained with marker D1S1631 (table 2), because of less sharing at neighboring markers.

These results are consistent with those shown in figures 1 and 2, in that no regions revealed linkage evidence that was near significance, even when a liberal criterion of MLS 3.0 was used. Thus, it is unlikely that any genes of moderate to large effect underlie this disorder.

Follow-up in FS2

The 49 families of FS2, comprising 50 ASPs, were analyzed for a total of 149 markers, 60 of which were original (i.e., MS1) markers typed in FS1. A total of 157 additional markers (MS2) were typed in FS1; 89 of these were also typed in FS2. Many new markers were typed only in FS1, either to fill in larger gaps in the original screen or to increase marker density in certain candidate regions, in particular on chromosomes 6p and 15q. On the basis of preliminary positive results in the first stage of analysis, markers were added on chromosomes 1 (two regions), 3, 9, 10, 13, 15, and 17.

Results of multipoint analysis of the combined data from both FS1 and FS2, for all 517 autosomal and X-linked loci, are given in figure 3 (the MLS curve for the total data is represented by the unbroken line; the exclusion curve fixing $\lambda_{si} = 3.0$ is represented by the dotted line). The linkage evidence, in FS2, for markers on proximal chromosome 1p was positive, with allele sharing of 50%–66% (average 56%); in FS1, maximal sharing in multipoint analysis was observed near marker D1S1631. With the inclusion of FS2 and additional markers in this region, the MLS of 2.15 now occurs at a slightly more proximal location (near marker D1S1675). This location gives, by far, the strongest linkage evidence in our study. The next most significant result is on chromosome 17p, where the MLS peaked at 1.21, near marker D17S1876. Only two other regions had an $MLS \geq 1.0$, one each on chromosome 7p (MLS 1.01, near marker D7S2564) and 18q (MLS 1.00, near marker D18S878). All other regions with $MLS \geq 1.0$ in the initial scan were less positive in FS2, leading to a reduction in the total MLS for FS1 and FS2 combined. Notably, only chromosome 1p had substantially positive allele sharing in FS2, leading to a modest increase in the MLS for FS1 and FS2 combined.

Proximal chromosome 15q is a candidate region for an autism-susceptibility locus, because of reported inverted duplications in a number of patients (Gillberg et al. 1991; Cook et al. 1997). Furthermore, some positive linkage results were reported in this region in a small collection of multiplex families (Pericak-Vance et al. 1997). We previously have reported that, in this region of chromosome 15, there is a lack of linkage evidence in the 97 FS1 ASPs (Salmon et al., in press). The addition

of 50 ASPs from FS2 has not altered this conclusion, since our linkage evidence is still uniformly negative across the entire region (fig. 3). The positive linkage results that we obtained on this chromosome were near marker D15S1050, which is 62 cM away.

We have also previously reported a lack of linkage evidence with markers in the HLA region on chromosome 6p in FS1 (Rogers et al., in press). The data in this region remain negative with the addition of FS2.

In a previous analysis of 38 of the ASPs included in the present study, we examined markers on the X chromosome. Although we excluded most of the chromosome, including the fragile X region, from harboring a gene of large effect, a region of Xq did show modest evidence of linkage, with a LOD score of 1.24. In the present study, based on a much larger number of families, there is no evidence of linkage anywhere on the X chromosome.

A genome screen of 39 multiplex families with autism, with follow-up in an additional 60 families, found suggestive evidence of linkage on chromosomes 7q and 16p (International Molecular Genetic Study of Autism Consortium 1998). The MLS in the same region of 7q is 0.62 (for marker D7S684), on the basis of a total of 139 families and 147 ASPs. We had a slightly more positive result at marker D7S1804 (MLS 0.93), located 10.3 cM more proximal. The original study had an MLS of 2.53. Thus, if a susceptibility locus resides in this region, its effect is likely to be small, given the modest LOD scores observed.

In our initial scan (i.e., with FS1), a region of chromosome 13q had an MLS of 1.65, near marker D13S779. In our follow-up, the linkage evidence at this location was negative (46.3% sharing); however, a region of modestly increased sharing appeared somewhat distal in the same area, with an MLS of 0.68, near marker D13S800 (fig. 3). This region has also been reported in another recent genome screen (Vieland et al. 1998). Again, the modest level of linkage evidence in our collection of 147 ASPs suggests that a gene of large effect in this region is unlikely.

Using a strict exclusion criterion of -2 (Hauser et al. 1996), we were able to exclude ~95% of the genome, for a value of $\lambda_{si} = 3.0$ (fig. 3). When the less strict exclusion criterion of -1 was used, 99% of the genome could be excluded.

Positive Controls

In our study, we had initially included two twin pairs originally thought to be DZ but subsequently, through our genotyping, shown to be MZ. These subjects were left in the study as a positive control and to estimate the rate of genotyping error. The individuals performing ge-

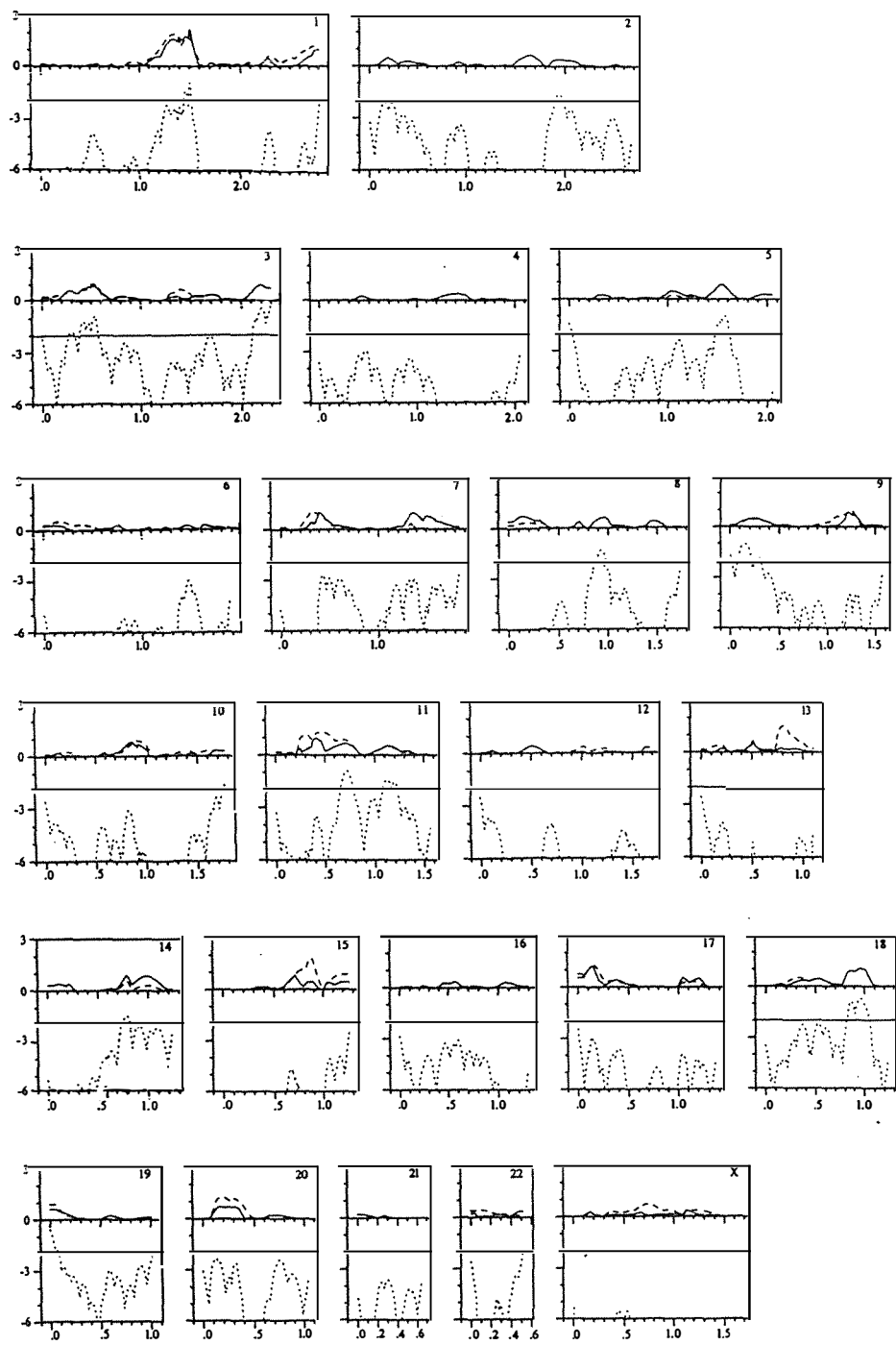


Figure 3 Multipoint sib-pair analysis of ASPs. The dashed line represents the MLS for FS1 with MS1; the unbroken line represents the MLS for FS1 + FS2, with all markers; the dotted line represents the exclusion plot for $\lambda_s = 3.0$, with all families and markers.

notyping were blinded to the status of these subjects. Considering a total of 946 loci scored for these two pairs, we found them to be genotypically discordant at 10 loci, or a rate of 1.0%. It should be noted that, in our study, families were analyzed intact, so that some genotyping

errors (Mendelian inconsistencies) could be identified. These Mendelian inconsistencies were identified and corrected (by additional genotyping, if necessary) and were not included in the error rate given above.

As a second positive control, we included the pseu-

doautosomal marker DXYS154. This marker is completely linked to sex, since it does not recombine between the X and Y chromosomes. In our study, we had a large predominance of affected males (male:female ratio 3.6:1 in FS1 and 2.8:1 in FS2). Thus, a majority of our ASPs are brother pairs, whereas a minority are sister pairs or of mixed sex. Overall, 73% of our ASPs are concordant for sex (63% males and 10% females). Therefore, we expect, at locus DXYS154, excess sharing of alleles inherited from the fathers, but not of those inherited from the mothers. In FS1 and FS2 combined, there were 58 paternal alleles shared and 22 paternal alleles not shared, giving a χ^2 of 16.2 (the LOD score equivalent of which is 3.52). By contrast, there was no difference in the sharing of maternal alleles: 42 alleles were shared and 45 alleles were not shared. The significantly increased sharing of paternal alleles at this marker demonstrates that our methodology was sufficiently sensitive to detect a real excess of sharing of this magnitude (i.e., 73%) when such an excess was present.

LD

We performed TDTs for all 517 autosomal and X-linked loci analyzed in this study. Here we report the results of the GCS test (see the Linkage-Disequilibrium analysis [LD] subsection, above). In figure 4 we have plotted the distribution of observed (empirically derived) *P* values, as a function of their rank. Under the null hypothesis of no LD, the distribution of *P* values should be uniform, and thus this plot should approximate the line $y = x$, which, in figure 4, is represented by the straight unbroken solid line; the values for affected offspring are represented by the dashed line, and those for unaffected offspring are represented as the jagged unbroken line. As can be seen in the figure, for $P < .05$ the observed and expected distributions are the same for the two types of offspring and fit the line $y = x$ well. These results are not unexpected, in that it is a priori unlikely that a disease locus would lie very close to and in LD

Discussion

We have performed a genomic screen by linkage analysis of 90 multiplex sibships with autism, including 97 independent ASPs, with 360 autosomal and X-linked markers. Regions that, in this screen, initially either (a) had positive results, (b) had gaps, or (c) were identified as specific candidate regions were subjected to follow-up with additional markers. We also genotyped a second set of 49 multiplex sibships (50 ASPs), with markers in specific target regions. In total, 519 markers were run on FS1, and 149 were run on FS2. More than 160,000 genotypes were generated.

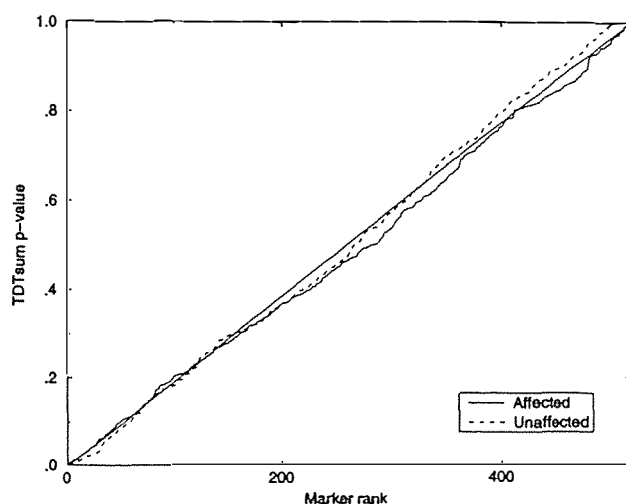


Figure 4 Observed distribution of *P* values from TDT tests for 519 markers for affected and unaffected offspring. The straight unbroken line represents expectation under no LD; the dashed line represents unaffected offspring; the jagged unbroken line represents affected offspring.

Analysis of allelic sharing for the initial 360 markers in FS1 was most consistent with a multigenic inheritance pattern, in that a global excess of allele sharing was observed in the ASPs versus the DSPs, but not specifically for a small number of loci. The overall distribution of allele sharing was most consistent with a model of ≥ 15 susceptibility loci.

These results are not inconsistent with the observed recurrence risks in nuclear families and twins. The very high (25-fold) MZ:DZ concordance ratio is indicative of at least several interacting loci and, potentially, of many such loci. Multilocus inheritance does not preclude the possibility of one or a few loci with larger effects; there is a limit, however, on how large these effects can be, given our modest linkage findings. A previous study of family recurrence risks in autism found a model with 3 interactive loci as providing the best fit to the data, with a range of 2-10 loci (Pickles et al. 1995). Our results clearly exclude a model with as few as 3 loci and are also inconsistent with a model with as few as 10 loci.

Our results also do not preclude the possibility of etiologic heterogeneity, with a subset of cases having a simpler (e.g., Mendelian) etiology. However, as explained above, such a group would need to be relatively minor, given both our results and the large MZ:DZ concordance ratio.

Our most significant findings were for proximal chromosome 1p. The most significant increase in sharing in our initial screen was for marker D1S1631, with 66% sharing and a *Z* score of 3.44 (table 2). The multipoint

analysis of FS1 brought this score down to an MLS of 1.88; however, this region was also the most positive in our follow-up and, in 147 ASPs, gave a total MLS of 2.15 and sharing of 60%. We note, however, that this result falls short of formal statistical significance for a genome screen (Lander and Kruglyak 1995). The MLS location after follow-up was somewhat more proximal than that in the initial screen, near marker D1S1675.

No other region had an MLS >1.3 in our follow-up. Our next most significant finding was for chromosome 17p, where the MLS was 1.21. We had modestly positive LOD scores in regions identified in other genome screens—chromosomes 7q and 13q. Our results for chromosome 15q, in the Prader-Willi region, which, on the basis of cytogenetic evidence, is a candidate region for an autism-susceptibility locus, were uniformly negative, consistent with our earlier results (Salmon et al., in press). However, in our initial screen, we did identify another region, more distal on chromosome 15q, that had an MLS of 1.75; but, in the follow-up, the MLS in this region decreased to 0.81.

We employed several controls in this study, to identify potential biases. First, we analyzed unaffected siblings in our multiplex sibships, to create a comparison group for our linkage and LD results. Second, we kept two pairs of MZ twins, originally thought to be DZ, in the genotyping set, as a positive control for our genotyping, as well as to estimate the error rate in genotyping. Finally, we included a pseudoautosomal marker completely linked to sex, as a positive control, reflecting the excess of sex-concordant (primarily brother) pairs in our study. As predicted, this marker showed a clear, statistically significant paternal excess of sharing (MLS 3.52). These controls provide confidence in the validity of our conclusions.

We also took a conservative approach to phenotyping. All study subjects were evaluated by both ADI assessment and ADOS assessment, and all questionable cases were excluded from analysis. Furthermore, to reduce potential heterogeneity, we excluded families in which all affected children had extremely low IQ, since this group may have distinct etiologies related to their severe-to-profound mental retardation. In total, we excluded 45 families from further analysis. The primary reason for the family exclusions was that not all affected sibs had a strict diagnosis of autism, as defined by agreement of all diagnosticians. As a result, the families in our study included few broadly defined, questionable, or mild cases (e.g., clinical diagnoses of either PDD-NOS or Asperger syndrome), even though most of the children in these excluded families did exhibit symptoms and behaviors indicative of the autism spectrum, some at the very high end of the spectrum and some at the very low end.

We employed these conservative diagnostic rules for inclusion within the study, which were applied equally to all affected subjects in a family, for two reasons. First, we wanted to enhance our chances of detecting any linkage, by creating a more homogeneous set of families. Second, we wanted to gain confidence that any negative results that we might obtain would not be due to the use of a too-broad diagnostic spectrum whose genetic basis is still uncertain (even though a number of recent family studies of autism probands suggest the presence of a broad range of mild behavioral symptoms and possibly related disorders in these families [Bailey et al. 1998]). Although our exclusion of “mild” or autism “spectrum” cases might be questioned by some who argue strongly for their inclusion within genetic studies, the only negative effect that this exclusion would have on our analysis would be to reduce the sample size. A homogeneous group of strictly defined affected subjects should maximize the chances of detection of linkage, for any plausible genetic model.

Although our genomic-screen results are largely negative, in that we did not identify a chromosomal region with significant linkage evidence, we have not formally excluded the possibility of one or a few disease-predisposing loci of moderate effect. On the basis of our results, the most likely location for such a locus or loci is on chromosome 1p—and, possibly, on chromosome 17p. We are currently examining these regions more closely. However, given the lack of strong linkage evidence in this large collection of families, positional cloning of any susceptibility genes in this disorder may be a difficult undertaking, and other approaches, such as LD studies of candidate genes and/or genomewide association studies may be required.

Acknowledgments

This research has been supported by National Institute of Mental Health grants MH52708, MH39437, MH00219, and MH00980; National Institutes of Health HG00348; National Health Medical Research Council grant 0034328; and by grants from the Scottish Rite, the Spunk Fund, Inc., the Rebecca and Solomon Baker Fund, the APEX Foundation, the National Alliance for Research in Schizophrenia and Affective Disorders, the endowment fund of the Nancy Pritzker Laboratory (Stanford); and by gifts from the Autism Society of America, the Janet M. Grace Pervasive Developmental Disorders Fund, and families and friends of individuals with autism. We are indebted to Drs. Magda Campbell, Isabelle Rapin, Peter Szatmari, and Luke Tsai for their expert clinical advice as study advisors, and we wish to thank the parent support groups and our clinician colleagues who referred families to this project. Our gratitude goes to the families with autism who have been our research partners in this endeavor. We also

wish to acknowledge the pioneering foresight of Dr. Roland Ciaranello, who led this study before his untimely death.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

ASPEX program package, <ftp://lahmed.stanford.edu/pub/aspeX>

Center for Medical Genetics, Marshfield Medical Research Foundation, <http://www.marshmed.org/genetics> (for genetic markers)

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/omim> (for TSC1 [MIM 191100]; autism, PDD, and Asperger syndrome [MIM 209850]; and FMR1 [MIM 309550])

References

- Anderson MA, Gusella JF (1984) Use of cyclosporin A in establishing Epstein-Barr virus-transformed lymphoblastoid cell lines. *In Vitro* 11:856–858
- Bailey A, LeCouteur 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, Palferman S, Heavey L, LeCouteur A (1998) Autism: the phenotype in relatives. *J Autism Dev Disord* 28:369–392
- Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, et al (1994) A case-control family history study of autism. *J Child Psychol Psychiatry* 35:877–900
- Cook EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, et al (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am J Hum Genet* 60:928–934
- Feinstein C, Reiss AL (1998) Autism: the point of view from fragile X studies. *J Autism Dev Disord* 28:393–405
- Folstein S, Rutter M (1977) Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 18:297–321
- Gillberg C (1998) Chromosomal disorders and autism. *J Autism Dev Disord* 28:415–425
- Gillberg C, Steffenburg S, Wahlstrom J, Gillberg I, Sjostedt A, Martinsson T, Liedgren S, et al (1991) Autism associated with marker chromosome. *J Am Acad Child Adolesc Psychiatry* 30:489–494
- Hallmayer J, Hebert JM, Spiker D, Lotspeich L, McMahon WM, Petersen PB, Nicholas P, et al (1996) Autism and the X chromosome. *Arch Gen Psychiatry* 53:985–989
- Hanson DR, Gottesman I (1976) The genetics, if any, of infantile autism and childhood schizophrenia. *J Autism Child Schizophr* 6:209–234
- Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 13:117–137
- International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence of linkage to a region of chromosome 7q. *Hum Mol Genet* 7:571–578
- Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, McMahon WM, et al (1991) Complex segregation analysis of autism. *Am J Hum Genet* 49:932–938
- Jorde LB, Mason-Brothers A, Wasdman R, Ritvo ER, Freeman BJ, Pingree C, McMahon WM, et al (1990) The UCLA-University of Utah epidemiologic survey of autism: genealogical analysis of familial aggregation. *Am J Med Genet* 36:85–88
- Kanner L (1943) Autistic disturbances of affective contact. *Nerv Child* 2:217–250
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Lazzeroni LC, Lange K (1998) A conditional inference framework for extending the transmission/disequilibrium test. *Hum Hered* 48:67–81
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J (1989) Autism Diagnostic Interview: a standardized investigator-based instrument. *J Autism Dev Disord* 19:363–387
- Lord C, Pickles A, McLennan J, Rutter M, Bregman J, Folstein S, Fombonne E, et al (1997) Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Autism Dev Disord* 27:501–517
- Lord C, Rutter M, Good S, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* 19:185–212
- Lord C, Rutter M, LeCouteur 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* 24:659–685
- Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. *Am J Hum Genet* 61:439–448
- Pericak-Vance MA, Wolpert CM, Menold MM, Bass MP, DeLong GR, Beatty LM, Zimmerman A, et al (1997) Linkage evidence supports the involvement of chromosome 15 in autistic disorder (AUT). *Am J Hum Genet Suppl* 61:A40
- Pickles A, Bolton P, Macdonald H, Bailey A, LeCouteur A, Sim C-H, 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* 57:717–726
- Risch N (1990) Linkage strategies for genetically complex diseases. I. Multilocus models. *Am J Hum Genet* 46:222–228
- Rogers T, Kalaydjieva L, Hallmayer J, Petersen PB, Nicholas P, Pingree C, McMahon W, et al. Exclusion of linkage to the HLA region in ninety multiplex sibships with autism. *J Autism Dev Disord* (in press)
- Salmon B, Rogers T, Kalaydjieva L, Hallmayer J, Petersen PB, Nicholas P, Pingree C, et al. Absence of linkage to chromosome 15q11-q13 markers in ninety multiplex families with autism. *Am J Med Genet* (in press)
- Smalley SL (1998) Autism and tuberous sclerosis. *J Autism Dev Disord* 28:407–414
- Smalley SL, Asarnow RF, Spence MA (1988) Autism and genetics: a decade of research. *Arch Gen Psychiatry* 45:953–961

- Spence MA, Ritvo ER, Marazita ML, Funderburk SJ, Sparkes RS, Freeman BJ (1985). Gene mapping studies with the syndrome of autism. *Behav Genet* 15:1–13
- 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* 52:506–516
- Spiker DK, Lotspeich L, Kraemer HC, Hallmayer J, McMahon W, Peterson PB, Nicholas P, et al (1994) The genetics of autism: characteristics of affected and unaffected children from 37 multiplex families. *Am J Med Genet* 54:27–35
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg L, Jakobsson G, Bohman M (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry* 30:405–416
- Szatmari P, Jones MB, Zwaigenbaum L, MacLean JE (1998) Genetics of autism: overview and new directions. *J Autism Dev Disord* 28:351–368
- Vieland V, for the Collaborative Linkage Study of Autism (CLSA) (1998) Results of a genomic screen for autism include strong evidence of linkage to chromosome 13. *Am J Hum Genet Suppl* 63:A16