Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Molecular Genetics Investigations in Autism

Louise Gallagher

MB BCh MRCPsych

PhD Thesis

Department of Psychiatry

Trinity College Dublin

2003

Declaration

Declaration 1:

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree at this or any other university.

Signed.....

Date.....27/7/04.....

Declaration 2:

This thesis is the result of my own investigations except where otherwise stated in the text. Other sources are acknowledged in the text. A bibliography is appended.

Signed.....

Date.....27/7/04.....

Declaration 3:

I hereby give consent for this thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed.....

Date.....27/7/04.....

Summary

This research thesis involved the establishment and management of a research group in autism genetics in addition to the collection of an autism sample of parent-child trios. A well-characterised sample of individuals with autism were recruited to participate in a genetics study in autism. One hundred trios were recruited for genetic association studies. The aim was to conduct both candidate gene studies and linkage disequilibrium studies within chromosomal regions identified by genomewide linkage studies as showing putative evidence for linkage.

A candidate gene study was performed that investigated association with variants in the Homeobox A1 (HOXA1) and Homeobox B1 (HOXB1) genes and autism. This was an attempted replication of work published by Ingram et al, 2000 that reported association with these variants and an autism sample (Ingram et al. 2000a). No association was detected with these variants and the sample described here, however a non-significant trend towards increased transmission of the G allele of the HOXA1 gene was observed.

During the course of clinical recruitment, an individual with autism was identified with an apparent deletion of chromosome 2q. This interesting finding was of significance as the abnormality was located within a region on chromosome 2q showing putative evidence of linkage with autism (Buxbaum et al. 2001; IMGSAC 2001; Shao et al. 2002a). Collaboration was established with the National Centre for Medical Genetics to map the extent of this deletion. Mapping was not straight-forward and the anomaly was shown to be a translocation of chromosome 2q to chromosome 9q (46, XY, ins (9;2)(q31.1;q31.2q31.3). The region on chromosome 2q was prioritised for follow-up owing to the aforementioned findings of three genomewide linkage studies. A microsatellite-based fine mapping linkage disequilibrium experiment was performed across a 8.5Mb region on chromosome 2q that had been defined by the apparent deletion. The Transmission Disequilibrium Test (TDT) revealed the presence of linkage disequilibrium between a microsatellite marker (D2S2077) located at the centromeric end of the region and autism (χ^2=5.8, p=0.013, OR = 1.75). A second multi-allelic marker, D2S270, located at the telomeric end of the region showed

evidence of genotypewise association with autism with the extended TDT. The significant finding around D2S2077 was followed up by a linkage disequilibrium fine mapping experiment using single nucleotide polymorphisms (SNPs) across a 868kb region defined by the microsatellite markers (D2S2310 and D2S364) that flanked D2S2077. No single SNP showed significant evidence of association with autism when tested with TDT. One SNP (dbSNP: rs155149) showed non-significant increased transmission in the autism sample. This was significant when transmissions were tested with the Haplotype-Based Haplotype Relative Risk (HHRR) (χ^2=5.16, p=0.023, RR=1.48, 1.01<RR<2.16). Association testing of haplotypes was performed using TRANSMIT. Association with autism was detected with several haplotypes that contained the over-transmitted alleles of the D2S2077 marker and the SNP (rs155149). A six-marker haplotype was found to have an odds ratio of 6.71 (χ^2=12.06, p=0.0005). This haplotype was located around a known gene, Integrin Alpha 4 (ITGA4), and two of the SNPs on the haplotype were located within introns in the gene. ITGA4 is involved in cell-adhesion and is expressed in the limbic brain, a region postulated to be involved in pathological process in autism. Two further known genes, Neurogenic Differentiation Factor 1 (NEUROD1) and Sperm Specific Antigen 2, SSFA2, have been identified in the region surrounding D2S2077. NEUROD1 is also a good candidate gene for autism as it is expressed in both the limbic brain and cerebellum and mouse null mutants show evidence of pathological abnormalities in the brains that are similar to those reported in post-mortem findings in autism. SSFA2 is expressed solely in the testis, however it may influence the expression of NEUROD1 through positional effects. This region on chromsome 2q represents a susceptibility region for autism based on converging evidence from several sources, three genomewide linkage studies, identification of an individual with a cytogenetic abnormality in the region and the finding of positive association in the region with a sample of individuals with autism. Further investigations following up this important finding are ongoing.

Acknowledgements

There are many people I must thank without whom this work would never have been possible. First and foremost I would like to thank my supervisors Professor Michael Gill and Professor Michael Fitzgerald. I would particularly like to acknowledge Michael Gill as without his mentorship and encouragement over the past four years I would never have even dreamed to undertake this work.

Very importantly, this work would never have been possible were it not for all the families who participated. I would like to thank them sincerely for giving up precious time and always affording a warm welcome to the research nurses and me. Their ongoing interest and enthusiasm for research is inspiring and has served as a great motivator to us all over the past four years.

In the neuropsychiatric genetics laboratory there are many I must thank for their valuable support in particular my good friend Dr. Ziarah Hawi for reading this work and many helpful suggestions. I must not forget David Lambert for all his expert technical support (and for converting me to Macs), the stars of the autism team, Judith Conroy and Dr. Ricardo Segurado, Naomi Lowe for all the 'top tips', Dr. Derek Morris and Kevin McGhee.

On the clinical side there are innumerable people to acknowledge but a particular mention must go to my wonderful research nurses past and present, Geraldine Kearney and Eleanor Meally.

At the National Centre for Medical Genetics I would like to thank Dr. Sean Ennis, Prof. Andrew Green, Dr. Ray Stallings and Dr. David Barton. Sean has been the mainstay of this collaboration and without his patience, dedication and hard work I would never have completed this thesis.

I would also like to mention the generosity of Dr. Nollaig Byrne and Prof. Carol Fitzpatrick at the Mater who allowed me to take leave from my clinical post in the Mater Hospital to complete this thesis.

The runners, Mary, Aiveen, Ruth and Ashie represent a special category to be thanked in their role in maintaining my sanity over the past four years and for being such good friends. Ruth and Aiveen in particular have helped me out in so many little ways over the past year. I have really valued their support and encouragement.

Finally I must give credit to the real workers behind this PhD, my family. Thanks to my husband Adrian for all his patience and support and generously putting his own life on hold particularly over the last few months. Thanks to my Mum for all the practical and emotional support and to my Dad who never tired of discussing my work with me before he died and who sadly did not achieve his ambition to make it to my graduation. My parents-in law, Des and Bernie, were always willing to step in for impromptu childcare, and that was greatly appreciated. A special thanks to my precious Max for not going off his mummy despite all the weekends spent away from him and to my darling daughter Alicia for a well-behaved pregnancy during the preparation of this thesis.

Statement of Work

This work was the product of the autism genetics group, Department of Psychiatry, Trinity College Dublin. Additionally, part of the work was performed in collaboration with the National Centre for Medical Genetics, Our Lady's Hospital, Crumlin.

The clinical recruitment and collection of samples for DNA were performed largely by the author and Geraldine Kearney (research nurse) and to a lesser extent by Eleanor Meally (research nurse).

DNA extraction and quantification was performed by the author and Judith Conroy (PhD student). All genotyping and analysis for the HOXA1 and B1 genes was performed by the author. Mapping of the translocation on chromosome 2q and the associated analysis was performed by Dr. Sean Ennis, Lecturer in Medical Genetics at the National Centre for Medical Genetics. The genotyping and analysis of the microsatellite based linkage disequilibrium experiment on chromosome 2q was conducted by the author. Genotyping of single nucleotide polymorphisms on chromosome 2q was carried out commercially by the Medical Research Council Geneservice. The analysis of this data was conducted by the author.

Publications

Gallagher, L, Becker, K, Kearney, G, Dunlop, A, Stallings, R., Green, A, Fitzgerald, M, Gill, M (2003). Brief report: A case of autism associated with del(2)(q32.1q32.2) or (q32.2q32.3). *J Autism Dev Disord*;33(1):105-8.

Gallagher, L, Hawi, Z, Kearney, G, Fitzgerald, M, Gill, M (2003). No Association Between Allelic Variants of HOXA1/ HOXB1 and Autism. *J Med Genet* (in press).

Conroy, J, Meally, E, Kearney, G, Fitzgerald, M, Gill, M, Gallagher, L (2003). The Serotonin Transporter Gene And Autism: A Haplotype Analysis in an Irish Autistic Population. *Molecular Psychiatry* (in press).

Dedication

This work is dedicated to my father Jim who always encouraged me to run far and strive high. We miss him dearly but he lives on in the hearts of his family and many dear friends.

Abbreviations

5-HT	5-Hydroxytryptamine
A	adenosine
ADI-R	Autism Diagnostic Interview – Revised
ADOS-G	Autism Diagnostic Observation Schedule - Generic
BAC	bacterial artificial chromosome
bp	base pair
C	cytidine
Ca^{2+}	calcium
CI	confidence interval
cM	centimorgans
dATP	deoxy adenosine triphosphate
dCTP	deoxy cytidine triphosphate
del	deletion
df	degrees of freedom
dGTP	deoxy guanosine triphosphate
DHPLC	Denaturing High Performance Liquid Chromatography
DNA	deoxyribonucleic acid
dNTP	deoxy nucleotide triphosphate
dTTP	deoxythymidine triphosphate
DZ	dizygote
ECACC	European Collection of Cell Cultures
eTDT	extended transmission disequilibrium test
G	guanosine
GDB	Genome Database
HHRR	haplotype-based haplotype relative risk
HWE	Hardy-Weinberg equilibrium
kb	kilobase
LD	linkage disequilibrium
LOD	log of the odds

Mb	megabase
MMLS	maximum multipoint lod score
MZ	monozygotic
NCBI	National Centre for Biotechnelogy Information
NPL	Non-parametric Lod Score
OMIM	Online Mendelian Inheritance in Man
OR	odds ratio
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
SD	standard deviation
SNP	single nucleotide polymorphism
T	thymidine
TDT	transmission disequilibrium test
UTR	untranslated region
UV	ultraviolet
VABS	Vineland Adaptive Behaviour Scale

Table of Contents

Index of Tables

Page

Index of Figures

Chapter 1

Introduction

1.1 Clinical background

1.1.1 Kanner's autism

In October 1938, at the Harriet Lane Home, Leo Kanner saw an autistic child for the first time. The American-trained physician, of Austrian origin had no formal background in either paediatrics or child psychiatry. Yet based on his clinical expertise and practicality he had been entrusted by Adolf Meyer to establish the Department of Child Psychiatry in John's Hopkins University. Little did he know on that October morning that he would lay the basis of the clinical description of a condition that would intrigue modern medicine for the next sixty years.

Donald was brought to the clinic by his parents at the age of 5 years, 1 month. Before the appointment his father had sent Kanner a thirty-three page handwritten account of his son's development to date and the difficulties the family were experiencing with him. Kanner was struck by the accomplishments of the child:

"At the age of 1 year "he could hum and sing many tunes accurately." Before he was 2 years old, he had "an unusual memory for faces and names, knew the names of a great number of houses" in his hometown. "He was encouraged by the family in learning and reciting short poems, and even learned the Twenty-third Psalm and twenty-five questions and answers of the Presbyterian Catechism." The parents observed, "he was not learning to ask questions or to answer questions unless they pertained to rhymes or things of this nature, and often then he would ask no question except in single words." His enunciation was clear. He became interested in pictures and very soon knew an inordinate number of the pictures in a set of Compton's Encyclopedia. "He knew the pictures of the presidents" and knew most of the pictures of his ancestors and kinfolks on both sides of the house. He quickly learned the whole alphabet "backward as well as forward" and to count to 100." (Kanner 1943)

Despite these accomplishments the child's behaviour was clearly odd, as his father described:

"He seems to be self-satisfied. He has no apparent affection when petted. He does not observe the fact that anyone comes or goes, and never seems glad to see father or mother or any playmate. He seems almost to draw into his shell and live within himself. We once secured a most attractive little boy of the same age from an orphanage and brought him home to spend the summer with Donald, but Donald has never asked him a question nor answered a question and has never romped with him in play. He seldom comes to anyone when called but has to be picked up and carried or led wherever he ought to go."

(Kanner 1943)

Over the following five years Kanner saw a further ten children with similar disturbances in development and published his case reports in 1943 entitled, "*Autistic Disturbances of Affective Contact*". In his monograph Kanner describes the abnormalities of "autistic aloneness", "obsessiveness", "stereotypies" and "echolalia" which he believes are fundamental to the disorder, differing from childhood schizophrenia in their presence from birth. On the origins of the disorder, in 1943, he concludes:

"We must then, assume that these children have come into the world with innate inability to form the usual, biologically provided affective contact with people, just as other children come into the world with innate physical or intellectual handicaps."

1.1.2 Asperger syndrome.

Hans Asperger (1906-1980) was an Austrian paediatrician with a special interest in "psychically abnormal" children. His paper, "*Die "Autistischen Psychopathen" im Kindesalter.*" (Asperger 1944) was based on investigations of more than 400 children with "autistic psychopathy" and described a pattern of behaviours in several young boys who had normal intelligence and language development, but who also exhibited autistic-like behaviours. The abnormalities he noted were of speech, non-verbal communication and social interactions. He also noted the presence of repetitive behaviours, resistance to

change and abnormalities of motor co-ordination. Similar to some of the children described by Kanner, Asperger also noted special skills or abilities such as excellent rote memory.

Like Kanner, Asperger also hypothesised that the condition was innate, believing it to be hereditary. He observed similar abnormalities in the fathers of the children he studied.

Since all his publications were in German, his descriptions went largely unnoticed until they were popularized by Lorna Wing (Wing 1976). Asperger's Syndrome is used to describe individuals with autistic like deficits associated with normal intelligence and is included in the International Classification of Diseases, 10th Edition (ICD-10) (World Health Organization 1992) and the Diagnostic and Statistics Manual, 4th Edition, (DSM-IV) and as a separate diagnostic category from autism (see section 1.1.6)

1.1.3 Autistic children prior to the 1940s.

A puzzling observation regarding the classification of autism as a discrete diagnostic entity is that there are almost no accounts in medical literature or otherwise regarding children with these kinds of abnormalities prior to the 1940s. Several authors refer to feral children such as Victor, the Wild boy of Aveyron, (1797) and Kamala and Amala (1929) of Midnapore, India, as possible examples of unsocialised autistic children abandoned by their mothers. One possible assumption is that children with abnormal development of any kind were regarded as mentally deficient. This observation is supported by the recent publication of prevalence data from California showing increased autism prevalence corresponding with a comparable decrease in the prevalence of intellectual disability of unknown cause (Croen et al. 2002). However given that child mental health services did not exist prior to the twentieth century it seems likely that such abnormalities of childhood were largely unrecognised.

1.1.4 Autism description and classification

Since the descriptions of Kanner and Asperger, autism classification has undergone several revisions and there is still discussion as to what constitutes autism and whether

autism and Asperger's disorder are discrete entities. A discussion of the development of autism classification is beyond the scope of this work. It is now generally accepted that autism and autistic spectrum disorders exist on a continuum of which Asperger's disorder is a part (Wing and Gould 1979). The focus of this work is on genetic studies in autism, relying on the narrowly defined syndrome according to ICD-10 and DSM-IV and it is the diagnostic features and classification of this disorder that are discussed below.

1.1.5 Clinical presentation.

What is now recognized as autism presents as a neurodevelopmental disorder within the first three years with the following characteristic features:

(1) Early onset (before 3 years).
(2) Severe abnormalities of reciprocal social relatedness.
(3) Severe abnormalities of communication development.
(4) Restricted, repetitive and stereotyped patterns of behaviours, interests, activities and imagination.

Children with autism generally present with speech and language delay associated with social abnormalities, often appearing to dwell in a "world of their own". They have difficulty making friends and appear to be lacking in empathy, i.e., unable to understand the thoughts of others as being separate from their own and have little understanding of the emotions of others. They may fail to moderate their behaviour in response to social or emotional cues and therefore may behave in a socially embarrassing way. Communication impairments are broad ranging. Speech and language acquisition is usually delayed and approximately 50% of individuals with autism remain non-verbal. Speech may be characterized by abnormalities such as echolalia and neologisms, or may appear scripted, comprised often of pieces of speech that have been lifted from every day life, television, film or radio. Individuals may appear stiff owing to the apparent lack of use of gesture, restricted facial expressions and monotonous speech. Repetitive behaviours may include spinning wheels on toys repeatedly or lining up objects rather than including them in normal play. Circumscribed interests or unusual preoccupations

may exist that are performed or talked about to the exclusion of other things. Difficulties with changes in routine may cause temper tantrums without obvious cause. Motor abnormalities usually take the form of complex mannerisms such as spinning or bouncing and characteristic finger flicking. Other motor abnormalities may also exist in association with the condition such as motor dyspraxia and impaired fine motor control.

Epilepsy occurs comorbidly in about 17% of autistic cases (Fombonne 2002). It tends to be commoner where the level of brain injury is likely to be more severe, e.g. lower ranges of intellectual functioning (Rutter 1970; Olsson et al. 1988). It usually has an onset in childhood although there appears to be a further period of onset during adolescence. Thus there may be an underestimate of the true rate as many studies have not followed subjects throughout adolescence.

1.1.6 Clinical classification of autism.

Standardised classification for autism has been defined by the ICD-10 (Table 1. 1) and DSM-IV (Table 1. 2) criteria.

There is generally agreement between ICD-10 and DSM-IV concerning the nature of the clinical abnormalities that occur in autism. Both agree that abnormalities should occur across the three domains of reciprocal social interaction, communication and behaviour. However some differences also occur. In the social domain, ICD-10 included impairments in the offering or seeking of comfort when distressed, this is not included in DSM-IV. Abnormalities of communication include abnormalities in initiating and maintaining conversation, poor flexibility in language (and lack of creativity/ use of fantasy in thought processes – ICD-10). DSM-IV refers to this as stereotyped/ repetitive use of speech. ICD-10 includes a lack of emotional response to the verbal overtures of others while DSM-IV does not. DSM-IV includes delay or lack of language development. ICD-10 does not include speech delay, but refers to abnormalities in speech with respect to poor variation in the cadence of speech. ICD-10 also includes lack of emphatic or descriptive gestures to aid meaning with speech. In the restricted, repetitive and stereotyped behaviour ICD-10 includes specific attachments to unusual objects.

Both classifications require the presence of the abnormalities before the age of three. DSM-IV states that the abnormalities should not occur in the context of other developmental disorders such as Rett's disorder or Childhood Disintegrative Disorder.

Table 1. 1 ICD-10 diagnostic criteria.

DIAGNOSTIC CRITERIA FOR AUTISM DISORDER (ICD-10) (WHO 1992)

At least 8 of the 16 specified items must be fulfilled.

A. Qualitative impairments in reciprocal social interaction, as manifested by at least three of the following five:

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.
3. Rarely seeking and using other people for comfort and affection at times of stress or distress and/or offering comfort and affection to others when they are showing distress or unhappiness.
4. Lack of shared enjoyment in terms of vicarious pleasure in other peoples' happiness and/or spontaneous seeking to share their own enjoyment through joint involvement with others.
5. Lack of socio-emotional reciprocity.

B. Qualitative impairments in communication:

1. Lack of social usage of whatever language skills are present.
2. Impairment in make-believe and social imitative play.
3. Poor synchrony and lack of reciprocity in conversational interchange.
4. Poor flexibility in language expression and a relative lack of creativity and fantasy in thought processes.
5. Lack of emotional response to other peoples' verbal and non-verbal overtures.
6. Impaired use of variations in cadence or emphasis to reflect communicative modulation.
7. Lack of accompanying gesture to provide emphasis or aid meaning in spoken communication.

C. Restricted, repetitive and stereotyped patterns of behaviour, interests and activities, as manifested by ate least two of the following six:

1. Encompassing preoccupation with stereotyped and restricted patterns of interest.
2. Specific attachments to unusual objects.
3. Apparently compulsive adherence to specific, non-functional routines or rituals.
4. Stereotyped and repetitive motor mannerisms.
5. Preoccupations with part-objects or non-functional elements of play material.
6. Distress over changes in small, non-functional details of the environment.

D. Developmental abnormalities must have been present in the first three years for the diagnosis to be made.

Table 1. 2 DSM-IV diagnostic criteria for autism.

DSM-IV Criteria, Pervasive Developmental Disorders

299.00 Autistic Disorder

A. A total of six (or more) items from (1), (2), and (3), with at least two from (1), and one each from (2) and (3):

(1) Qualitative impairment in social interaction, as manifested by at least two of the following:

(a) Marked impairment in the use of multiple nonverbal behaviors, such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction.

(b) Failure to develop peer relationships appropriate to developmental level.

(c) A lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest).

(d) Lack of social or emotional reciprocity.

(2) Qualitative impairments in communication, as manifested by at least one of the following:

(a) Delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime).

(b) In individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others.

(c) Stereotyped and repetitive use of language or idiosyncratic language.

(d) Lack of varied, spontaneous make-believe play or social imitative play appropriate to developmental level.

(3) Restricted, repetitive, and stereotyped patterns of behavior, interests, and activities as manifested by at least one of the following:

(a) Encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus.

(b) Apparently inflexible adherence to specific, nonfunctional routines or rituals.

(c) Stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting or complex whole-body movements).

(d) Persistent precoccupation with parts of objects.

B. Delays or abnormal functioning in at least one of the following areas, with onset prior to age 3 years: (1) social interaction, (2) language as used in social communication, or (3) symbolic or imaginative play.

C. The disturbance is not better accounted for by Rett disorder or childhood disintegrative disorder.

1.1.7 Diagnosis of autism for genetic studies

Defining a behavioural disorder such as autism for genetic research is complicated by several other factors that must be considered when recruiting individuals to a genetic study. Firstly, there are several other child psychiatric conditions that present with similar clinical features to autism. Thus a careful differential diagnosis must be considered in order to exclude other disorders. Secondly, the traits being measured are not discrete traits. Many occur within the normal population across a spectrum, e.g. sociability or rigidty. In order to minimise heterogeneity within the phenotype a rigorous research diagnosis must be obtained, measuring traits and determining a cut-off for diagnosis. Typical clinical assessments in autism are not sufficiently rigorous for defining a research phenotype (Le Couteur et al. 1989). While such instruments detect children on the autistic spectrum the majority do not provide a cut-off between children with autism and the autistic spectrum.

1.1.8 Differential diagnosis of autism.

A number of differential diagnoses for autism should be considered when an autism diagnosis is suspected. Both ICD-10 and DSM-IV include broader categories of autism diagnosis. In ICD-10 this is 'Atypical Autism', and in DSM-IV this is represented by Pervasive Developmental Disorder - not otherwise specified (PDD-NOS). 'Other Pervasive Developmental Disorders' in ICD-10 represents Rett disorder and childhood disintegrative disorder. Asperger Disorder is a further diagnostic category consisting of abnormalities in social interaction, communication and restricted or circumscribed areas of interest. This diagnosis is restricted to individuals of normal intellectual functioning and in DSM-IV it is specified that there should be no history of speech and language disorder. Additionally, for this diagnosis to be made, the presentation should not meet criteria for an autistic spectrum disorder. What is clear from the above discussion is that the disorders classified within the autistic spectrum/ Asperger Disorder represent a group of related disorders.

Other distinct child psychiatric disorders have similar presentation to autism and must be considered in the differential diagnosis. These are illustrated in Table 1.3. In addition

several medical conditions present with autistic like features and are also discussed below (section 1.2.1)

1.1.9 Development of diagnostic tools for a research diagnosis of autism

Standardised diagnostic instruments are increasingly used in psychological and psychiatric research to overcome issues of heterogeneity. Cohen et al, 1987, made the observation that the steps leading from clinical data to diagnosis need to be as explicit and operationalised as possible. Operationalised diagnoses allow for comparison between research studies and overcome localized differences in clinical diagnosis. For molecular genetic studies in complex disorders the need for standardized diagnoses cannot be understated. Phenocopies should be detected where possible by the application of appropriate exclusion criteria.

A number of questionnaires and assessments are commonly used in the clinical setting to diagnose autism but these are not sufficiently rigorous for use in research. The Autism-Diagnostic-Interview Revised (Le Couteur et al. 1989) and the Autism Diagnostic Observation Schedule – Generic (Lord et al. 2000) have been developed for use in autism research and are used almost exclusively in autism genetics investigations to define the phenotype. These tools are described in chapter 3 (section 3.3.5).

1.2 Aetiological theories in autism

The aetiology of autism remains elusive. Asperger made the assertion that the condition was heritable. The role of genetic factors in autism is discussed in detail below (section 1.3). In efforts to understand the aetiology of autism much interest has been devoted to the associated medical conditions, neuropsychological findings, neuropathological findings and the role of possible environmental factors. These are discussed briefly here.

1.2.1 Medical disorders associated with autism

Autism is observed in association with several medical conditions with known neurological pathologies (Gillberg et al. 1991). In order to avoid the presence of

Table 1. 3 Differential diagnosis in autism

Autism differential diagnosis
Exclude:
Hearing impairment
Known medical conditions
Intellectual disability of unknown cause
Rett disorder
Landau-Kleffner aphasia
Speech and language disorders
Selective mutism
Consider:
Childhood disintegrative disorder
Childhood schizophrenia
Other psychoses with onset in childhood
Disorders with shared features:
ADHD
OCD
Depression
Attachment disorders

phenocopies in the autistic sample that was recruited to this study it was essential to exclude any known medical causes. The procedure for doing so is discussed in section 3.3.6. The more common disorders detected in association with autism are briefly discussed below. All of these are disorders affecting neurodevelopment and the observaton of autistic features in association with these conditions suggests that pathological neurodevelopmental processes underlie the abnormalities observed in autism.

1.2.1.1 Phenylketonuria (PKU).

PKU is an inborn error of metabolism resulting from a deficiency of phenylalanine hydroxylase and characterized by mental retardation. It has a prevalence of 1 per 5, 000 in the Irish population. There are now wide scale screening and detection programmes for this condition in the developed world and the clinical consequences of untreated PKU are observed less often. However in the past, patients with untreated PKU presented with autistic symptomatology. This observation was reported by Benda in 1960 (Benda 1960). The symptoms did not persist when the condition was controlled with a phenylalanine-free diet. Similarly the associated intellectual disability associated with PKU has been largely eradicated through dietary measures. The underlying pathological mechanism appears to be related to the toxic effects of phenylalanine in the brain. The association with autistic symptoms in untreated PKU would suggest that some of these adversely affected regions may be involved in autism.

1.2.1.2 Tuberous sclerosis (TS).

This is a genetically heterogenous group of neurocutaneous disorder characterised by benign hamartomas and abnormalities of skin and CNS. Mental retardation and seizures are commonly associated. Two genes, TSC1 and TSC2 on chromosomes 9q34 and 16p13.3, have been implicated in the aetiology (Fryer et al. 1987; Haines et al. 1991; Povey et al. 1994). Autistic symptomatology has been described in 17-58% of cases (Smalley et al. 1992). The incidence of TS in presenting cases of autism/ autism spectrum disorders ranges between 0.4-3% (Eliason 1988; Ritvo et al. 1990);(Wing and Gould 1979; Steffenburg 1991). The association between the two disorders has two important

implications. The first is that tuberous sclerosis has CNS pathology and autism by association may have abnormalities in similar brain regions. Secondly, TS is a genetic condition and it is possible that autism may be associated with abnormalities in genes involved in tuberous sclerosis. The latter implication has not been supported by scientific investigation making it unlikely that the genes involved in TS are involved in autism.

1.2.1.3 Neurofibromatosis (NF1).

Neurofibromatosis is a neurocutaneous disorder characterized particularly by cafe-au-lait spots and fibromatous tumors of the skin. Like Tuberous Sclerosis it presents with tumors of the CNS, skin and other organs. Two distinct forms exist, Von Recklingshausen's disease (type I) and bilateral acoustic neurofibromatosis (type II). Both are inherited as autosomal dominant disorders. The gene involved in NF1 is on chromosome 17q and the gene for NF2 is on chromosome 22q. Intellectual disability occurs in 10-20% of cases. Impulsivity and social difficulties are well-documented behavioural manifestations (Eliason, 1988). Several studies have reported an increased incidence of autism in association with NF1 (Gillberg and Forsell 1984);(Steffenburg 1991). The condition may go undiagnosed in childhood and should be excluded in the assessment of an individual with autistic symtpomatology.

1.2.1.4 Rett syndrome

This is a neurodevelopmental disorder with a known genetic aetiology. It typically affects girls and should be considered particularly in the differential diagnosis of girls with autism. There is typically a period of normal development in the first 6-18 months followed by a progressive loss of speech and purposeful hand movements. These are usually accompanied by social withdrawal. Profound intellectual disability is associated and acquired microcephaly occurs between 5 months and 4 years of age. Stereotypic hand movements develop, typically in the form of midline 'hand wringing', clapping, mouthing and washing/rubbing automatisms. Gait apraxia and truncal apraxia/ ataxia develop between 1-4 years and some children become non-ambulatory. Language is usually absent. Seizures occur in about 70% of children. Hyperventilation and breath-holding episodes are common. Dystonia and choreo-athetoid movements may occur and

other associated features include growth retardation, spasticity, peripheral vasomotor disturbances, scoliosis and hypotrophic small feet. The genetic abnormality underlying classical Rett syndrome was identified as a mutation within Methyl-C_pG-binding protein 2 (MeCP2) gene (Amir et al. 1999). It has since been shown that different mutations occurring within the gene may be associated with variable phenotypes such as the preserved speech variant and the male Rett phenotype (Schanen 2001). Genetic association studies between autism and the classical Rett mutation have not identified an association. However given the variation of mutations occurring within the gene ongoing studies are merited (Weaving et al. 2003).

1.2.1.5 Congenital rubella.

The congenital rubella syndrome is a constellation of physical and developmental abnormalities that present following exposure to rubella in utero. The incidence of this has decreased dramatically with widespread vaccination programs of both infants and school age girls. The association between autism and congenital rubella was first highlighted by (Chess 1971). Out of 243 children with congenital rubella who were evaluated, 18 were identified as having autism. Given the occurrence of both intellectual disability and autism in the condition it lends support to the neurodevelopmental aetiology of autism.

1.2.2 Neuropsychological abnormalities

Increased understanding of the underlying neuropsychological abnormalities occurring in autism is important with respect to the identification of potential brain regions that are affected in the disorder. While an in-depth discussion is not possible the main theories are mentioned briefly.

Deficits in Theory of Mind (ToM), or the ability to represent the mental states of others, theoretically explain some of the social impairment associated with autism (Baron-Cohen et al. 1997). Repetitive, stereotyped and restricted behaviours of autism are regarded as a manifestation of executive dysfunction, since it is executive functions that allows our

behaviour to be planned, adaptable, appropriate and relevant (Rinehart et al. 2001). Weak Central Coherence, or the tendency to derive higher-level or global meaning from local detail has also been proposed (Happe 1996; Happe 1999).

Neuroanatomical hypotheses of autism arising from these neuropsychological abnormalities include abnormalities of:

(1) The frontostiatal sytem (McAlonan et al. 2002)

(2) The Limbic system (Kemper and Bauman 1998) (impaired emotional behaviour and social interactions)

(3) The fronto-cerebellar system (deficits in attention switching that may contribute to deficits in joint attention) (Courchesne et al. 1994; Harris et al. 1999)

1.2.3 Neuropathological abnormalities

Studies of neuropathological abnormalities in autism are limited however, they are important to the understanding of brain pathophysiology. Increased head circumference in autism is a well recognized finding (Woodhouse et al. 1996) and is further supported by post-mortem findings of increased brain size and relative increase in the size of the cerebellum (Bailey et al. 1998) although, this study reported a paradoxical decrease in the number of cerebellar Purkinje cells. Other neuropathologic changes reported include macroencephaly, increased neuronal packing and decreased cell size in the limbic system, agenesis of the superior olive, dysgenesis of the facial nucleus, hypoplasia of the brainstem and posterior cerebellum, and increased neuron-packing density of the medial, cortical and central nuclei of the amygdala and the medial septum (Courchesne 1997). Abnormalities in organization of the cortical minicolumn, representing the fundamental subunit of vertical cortical organization, may underlie the pathology of autism and result in altered thalamocortical connections, cortical disinhibition, and dysfunction of the arousal-modulating system of the brain (Casanova et al. 2002).

Structural neuroimaging studies have shown increased average brain volume; decreased grey matter volumes in the limbic system (Abell et al. 1999); reduced neuron numbers in

the vermis of the cerebellum and gross structural changes in cerebellum (Sparks et al. 2002) and the parietal lobes (Giedd et al. 1997). Poorer neuronal integrity in prefrontal areas has been suggested (Murphy et al. 2002), as have concurrent abnormalities in the frontal cortex and cerebellum (Carper and Courchesne 2000). A longitudinal functional magnetic resonance imaging (fMRI) study suggests an accelerated period of growth during childhood compared with typically developing controls (Courchesne et al. 2001).

1.2.4 Environmental factors and autism.

There are several documented examples of an association between environmental factors and the development of autism. Increased rates of autism have been reported in association with thalidomide exposure in utero, particulary early in the first trimester (Stromland et al. 1994). The insult was estimated to occur between days 20-24 post-conception. Rodier 1996, hypothesized that autism could be caused by a defect of neural tube closure as closure occurs around days 20-24. Autism has also been observed following exposure to valproic acid (Christianson et al. 1994);(Moore et al. 2000). Valproic acid also interrupts neural tube closure. A rat model exposed to valproic acid showed reductions in the number of cranial nerve nuclei in addition to malformed ears (Rodier et al. 1996). There have been reports of higher rates of anatomic and physiological dysfunction of the cranial nerves in individuals with autism (Miller et al. 1998). These observations led to the investigation of the HOXA1/B1 genes in association with autism, based on the observation that mice knock-outs for these genes demonstrated some of the same abnormalities observed in the valproic acid-exposed mice (Ingram et al. 2000b).

Other environmental pathogens that have been mooted as having a possible causative association with autism include Measles (Furlano et al. 2001) with the attendant unsubstantiated controversy regarding the MMR vaccine (Fombonne and Chakrabarti 2001).

There are many case reports in the literature of autism occurring with other disorders with CNS pathology and a genetic aetiology, a detailed discussion of which is beyond the

scope of this work. However these observations lend support to the aetiological theory of autism as a neurodevelopmental disorder with a probable genetic aetiology. The evidence supporting this theory is discussed in the next section.

1.3 Genetic studies in autism

While the aetiologies of autism remain unclear, hereditary factors have been suspected since the time of the early descriptions of the condition. As discussed above, Asperger noted the presence of similar abnormalities in the fathers of the boys he saw with Asperger syndrome. However, during the 1950s parents, particularly mothers, were unfairly blamed for the development of their child's autism. In 1964 Bernard Rimland, a parent and professional, published *Infantile Autism: The Syndrome and its Implication for a Neural Theory of Behaviour.* He put forward an alternative theory for autism aetiology, i.e. that autism might have a biological basis rather than the psychotherapeutic theories that had been widely held at the time. Despite this, parents of autistic individuals were to endure almost another decade of criticism as to their parenting practices (reviewed by DeMyer et al 1981). This began to change by the mid-seventies when a growing interest within the scientific community regarding the role of genetic factors started to emerge.

Several observations led to this increased interest. These included the observations that autism appeared to cluster within families. These observations were subsequently confirmed by the findings of clinical genetic studies, i.e. twin and family studies. A further factor, discussed in this section, was the observation that autism was frequently associated with chromosomal anomalies such as abnormal karyotypes and Fragile X (Cohen et al. 1991).

1.3.1 Family studies

Clinical genetic studies in autism were conducted in several countries in the 1970s and 80s. Family studies were undertaken to ascertain the rates of autism in first-degree relatives of affected individuals. In a review of the literature Smalley reported sibling recurrence rates of 3% (Smalley et al. 1988). Bolton found a recurrence rate of 2.9% in siblings of individuals with autism, and evidence for 'lesser variants' of the disorder in

12.4% and 20.4% of siblings (Bolton et al. 1994). In contrast there was a marked fall off in the incidence among 2nd degree relatives. Recurrence risk estimates as high as 8.6%, (7% where the first autistic child was male, and if a female, 14.5%) have been reported (Ritvo et al. 1990). Based on this data, Jorde calculated a sibling recurrence risk of 4.5% (Jorde et al. 1991). One study also found higher rates of social and communication deficits and stereotyped behaviours in first-degree relatives within multiple incidence families (Piven and Palmer 1997). Increased rates of a broader phenotype in relatives have been confirmed (Pickles et al. 2000). Factors that seemed to predict the presence of autistic symptoms in relatives of verbal autistic individuals were severity of the autism in the index case and the birth optimality score (i.e measure of the degree of obstetric difficulties), while in non-verbal individuals, the degree of relative and parental status were the relevant factors.

An increase in the rates of other psychiatric disorders in the first-degree relatives has also been identified, perhaps suggesting a partly shared commonality in the aetiology. Bolton et al, 1998 reported increases in motor tics, obsessive-compulsive disorder (OCD) and affective disorders in relatives of autistic probands (Bolton et al. 1994). The individuals with OCD were more likely to exhibit autistic-like social and communication impairments. Relatives of autistic individuals are also reported to have high rates of major depression and social phobia, not associated with the broad autism phenotype (Piven and Palmer 1999). This was not attributed to the increased stress associated with raising an autistic child.

1.3.2 Twin studies

The first twin study of a cohort of 21 twin pairs was published by Folstein and Rutter in 1977 (Folstein and Rutter 1977). Concordance rates of 36% in MZ twins and 0% in DZ twins were found for the narrow definition of autism. Although the concordance rates were low, the marked differences in concordances between mono- and dizygotic twins were suggestive of a role for genetic factors in the disorder. One of the drawbacks to this study was the lack of operationalised criteria for the diagnosis of autism. This was addressed in a more extensive follow up study published in 1995 by Bailey et al. This

was an extension of Folstein and Rutter's original twin study using operationalised criteria as defined by the ADI-R and ADOS-G diagnostic tools. The findings of this study showed MZ:DZ concordances of 60%:0%, i.e. 60% of the monozygotic twins were both autistic compared with 0% of the dizygotic twins. The marked discrepancies in concordance rates between the two supports the role of genetic factors in the aetiology since monozygotic twins share all of their genetic information while dizygotic twins, like siblings share only 50% of their genetic information. A twin study published by Steffenburg et al in 1991, based on extensive twin registers from Norway, Sweden and Denmark showed concordance ratios of 91%:30% which also differed in the rates between MZ and DZ twins. The higher concordances in this population were possibly observed due to broader diagnostic criteria.

1.3.3 Mode of inheritance

The family studies suggest that the disorder is familial and taken together with twin studies, the data are strongly suggestive that autism is heritable. Autism has a population prevalence of 5 –10 per 10, 000. Estimates of population prevalence rates have increased over the past decade with improved detection and possibly improved distinction between autism and generic learning disability (Fombonne 1999; Croen et al. 2002). Prevalence estimates from studies conducted prior to this were 3-5/10, 000. Based on a sibling recurrence rate of 3% and a population prevalence of 4 per 10, 000, the recurrence risk in siblings (λ_s) is 75, i.e, siblings are 75 times more at risk of developing the disorder. Using these figures the heritability of the disorder has been estimated at 91-93% using a multi-liability threshold model (Falconer 1965; Bailey et al. 1995). This represents the highest heritability for any neuropsychiatric disorder. Unfortunately these figures have not been recalculated based on changing prevalence figures. Given a prevalence of 10 per 10, 000 and a sibling recurrence rate of 3% the λ_s would fall to 30, which would reduce the heritability estimates, although it seems clear that the disorder has a substantial heritable component.

The mode of inheritance is not clear but seems most compatible with a polygenetic disorder. The observations of marked differences in concordance rates between MZ and

DZ twins taken alongside the marked fall off between first and second degree relatives are also compatible with a gene of moderate effect in association with several genes of minor effect (Pickles et al. 1995). However it has also been suggested that up to 15 genes may be involved (Risch et al. 1999). Risch emphasised that the potential role of epigenetic factors cannot be ignored. Epigenetic mechanisms influence gene expression but these mechanisms are not directly attributable to the DNA sequence of the gene. This can involve a variety of differing mechanisms such as allelic exclusion, X-chromosome inactivation, long-range control by chromatin structure and cell position-dependent short-range signalling. Allelic exclusion in the form of imprinting occurs across the Prader-Willi/ Angelmann critical region. This region has been of interest in autism because of reports of chromosomal rearrangements (Trisomy 15q11-q13) and subsequent reports of association in the region (discussed in more detail below). The role of other epigenetic factors in relation to autism has yet to be elucidated.

1.3.4 Genetic disorders associated with autism

Features of autism occur in some medical disorders with a known genetic aetiology. These include disorders such as Phenylketonuria and Tuberous Sclerosis among others (discussed in sections 1.2.1.1 and 1.2.1.2). The discussion of the Fragile X syndrome and cytogenetic abnormalities occurs here owing to the contribution of each to molecular genetic studies in autism.

1.3.5 Fragile X

The Fragile X anomaly was first demonstrated by Sutherland who showed the learning disability syndrome to occur in association with a fragile site on the X chromosome when suspended in a folate medium (Sutherland 1977). Individuals with fragile X have a typical phenotype of large ears, macro-orchidism and intellectual disability. It is caused by a series of unstable expanding CGG repeats in non-coding regions on the X chromosome. Two sites of expansion have been identified. The Fragile X A site (FRAXA) occurs at Xq27.3 in the 5'-untranslated region (UTR) of the fragile X mental retardation site-1 (FMR-1) gene. Highly expanded numbers of repeats result in loss of gene function by abolition of transcription. Between 6 and 54 repeats are stable. Repeat

numbers in excess of 200 are unstable and have the capacity to expand further (Kremer et al. 1991).

The clinical observation of autism occurring in association with the Fragile X anomaly has been well established (Cohen et al. 1991). Initial studies suggested that autism occurred in up to 30% of individuals with this anomaly. The current estimates are that the core autistic syndrome occurs in 5-10% of individuals while a significant proportion of those with Fragile X demonstrate autistic traits (Hjalgrim et al. 1998). The gene involved in Fragile X syndrome was identified as FMR-1. Studies of FMR-1 variants in autism did not demonstrate the presence of an association (Klauck et al. 1997b);(Meyer et al. 1998).

1.3.6 Cytogenetic abnormalities

Abnormalities of chromosomal structure are widely reported in association with autism, with at least one reported on almost every chromosome to date. These include autosomal aneuploidies, deletions, duplications, translocations, ring chromosomes, inversions and supernumerary marker chromosomes. Chromosome abnormalities have been reported to occur at a rate of 3% in autism (Fombonne 1999) and cytogenetic testing is recommended in particular where autism co-occurs with dysmorphology.

1.3.6.1 Chromosome 15q

The commonest cytogenetic abnormality associated with autism is a trisomy of chromosome 15q11-q13 (Gillberg et al. 1991; Robinson et al. 1993; Baker et al. 1994); (Bundey et al. 1994; Leana-Cox et al. 1994; Schinzel et al. 1994; Crolla et al. 1995; Hotopf and Bolton 1995; Flejter et al. 1996; Cook et al. 1997a). This appears to be related to an autistic phenotype when the supernumerary chromosome is of maternal origin (Browne et al. 1997; Cook et al. 1997a).

The 15q11-q13 region is imprinted and includes the Prader-Willi/ Angelmann Critical Region (PWACR). Prader-Willi syndrome is a neuroendocrine disorder associated with psychomotor retardation, intellectual disability, obesity and short stature and is caused by a deficiency of gene(s) on the paternal chromosome 15 that arise either as a result of an

interstitial deletion on the paternal chromosome 15q or as maternal uniparental disomy of chromosome 15. The two genes implicated are the Small Nuclear Ribonucleoprotein Polypeptide N (SNRPN) gene and the necdin gene. The SNRPN is expressed predominantly in the CNS and appears to have a role in mRNA processing although the function of this class of gene is poorly understood. The necdin gene is a growth suppressor expressed in virtually all postmitotic neurons in the brain. Both SNRPN and necdin are maternally imprinted and thus only expressed on the paternally inherited chromosome in the foetal brain, heart and adult brain (Reed and Leff 1994); (MacDonald and Wevrick 1997). Angelmann syndrome is a neurologic syndrome associated with intellectual disability, absence of speech and language development, 'puppet-like' ataxic gait, and characteristic dysmorphology. It is caused by a deficiency in the ubiquitin-protein-ligase E3A gene (UBE3A) on the maternal chromosome 15. UBE3A encodes for E3 ubiquitin protein ligase (E6-AP), which transfers ubiquitin molecules to target proteins enabling their degradation by the proteosome complex. It is expressed exclusively on the maternal chromosome in the brain (Vu and Hoffman 1997). Deficiencies in the gene may arise through maternal interstitial deletions, paternal uniparental disomy or a single imprinting gene defect in UBE3A.

Due to findings on Ch15q in relation to autism, association studies have been conducted across the region. (Cook et al. 1998) identified linkage disequilibrium around an allele of microsatellite 155CA-2 marker in the GABRB3 gene, which encodes for a sub-unit of the GABAa receptor. This finding was not replicated by Martin et al, 2000 but they found evidence for linkage disequilibrium with a marker, GABRB3, approximately 60 kb beyond the 3' end of beta3-subunit gene (Martin et al. 2000). Further studies in this region have shown evidence for association around UBE3A (Nurmi et al. 2001). A further novel gene, ATPase Class V Type 10C (ATP10C), postulated to be a calcium transporting ATPase, has been mapped 200kb distal to UBE3A and is also preferentially maternally expressed (Herzing et al. 2001). No association with variants in this gene and autism have been described to date (Kim et al. 2002b). One of the problems in conducting association studies in this region is increased recombination frequency (Bass et al. 2000) as this breaks down the linkage disequilibrium relationships between markers

in the region. A discussion of linkage disequilibrium and genetic association is considered in more detail in Chapter 2.

1.3.6.2 Other cytogenetic abnormalities

Given the results of genomewide linkage findings in autism (discussed below) the occurrence of anomalies on two other chromosomes are discussed here. Chromosome 7q has been identified by several genomewide linkage studies as a putative linked region. A translocation between chromosomes 7 and 13 (Ch7q31 t(7;13)(q31.2;q21)) revealed the disruption of a tumor suppressor gene, suppressor of tumorigenicity 7 (RAY1) (Vincent et al. 2000). The translocation had been transmitted by an unaffected mother. Twenty-seven unrelated individuals with autism were screened for mutations within the gene but none were detected.

A further breakpoint on chromosome 7 at 7q11.2 was reported in a monozygotic twin pair concordant for autism carrying a translocation between chromosomes 7 and 20 (t(7;20) (q11.2; p11.2)) (Sultana et al. 2002). A novel gene AUTS2 was identified at the breakpoint. The gene is 1.2 Mb in length and has 19 exons and the predicted protein has 1295 amino acids and does not correspond to any known protein. No association was detected between two exonic polymorphisms and an autism sample. Linkage analysis of four dinucleotide repeat markers (two within and two flanking the gene) was also negative (Sultana et al. 2002).

1.3.6.3 Cytogenetic abnormalities on chromosome 2q

Chromosome 2q is also a region of interest highlighted by genomewide linkage studies (see below). There are several reports in the literature of chromosomal abnormalities arising on Ch2q. Deletions of Ch2q37 have been reported in association with autism and mild dysmorphology (Gorski et al. 1989; Lin et al. 1992; Conrad et al. 1995; Ghaziuddin and Burmeister 1999; Wolff et al. 2002).

During the course of clinical ascertainment for this research an individual was identified with an apparent deletion on chromosome 2q (del(2)(q32.2q32.2) or (q32.2q32.3)

(Gallagher et al. 2003). This anomaly was found to lie within a region on chromosome 2q showing evidence for linkage from genome-wide linkage studies (Buxbaum et al. 2001; IMGSAC 2001; Shao et al. 2002a). The individual in question had an autism spectrum disorder and very little dysmorphology. The mapping of this anomaly and subsequent association studies are discussed in Chapter 6. While there have been other reports of chromosomal deletions of 2q31q33 (Taysi et al. 1981; Al-Awadi et al. 1983; Buchanan et al. 1983; Franceschini et al. 1983; Young et al. 1983) and at least one report of a duplication of this region (Ramer et al. 1990), the region reported in this work is considerably smaller than previous reports. None of these reports refer to an association with autism in the cases although many subjects were too young at the time for a clinical diagnosis of autism to have been made. The most consistent clinical manifestations of these reports include low birth weight, cleft palate, developmental delay/ learning disability, generally in the moderate range, microcephaly and abnormalities of the digits including syndactyly and campylodactyly.

The significance of this chromosome 2q finding and autism is that there appears to be converging evidence that this region represents a susceptibility region for autism (see also section 1.3.8.1)

1.3.6.4 Cytogenetic abnormalities and gene identification

The identification of genes from cytogenetic abnormalities found in association with disorders has proven successful in the past. Classic examples of this approach to gene identification include identification of the Retinoblastoma gene (RB1) (Lee et al. 1987) the Wilms' tumor gene (WT1) (Huff et al. 1991), the dystrophin gene in Muscular Dystrophy (Zneimer et al. 1993), the UBE3A gene in Angelman Syndrome (Kishino et al. 1997) and the SNRPN/ necdin candidate genes in Prader-Willi Syndrome (Ozcelik et al. 1992). The Velo-Cardiofacial syndrome arising secondary to a deletion on chromosome 22 (22q11) has led to the identification of a possible susceptibility gene for psychosis (proline dehydrogenase –PRODH) (Jacquet et al. 2002).

1.3.7 Candidate gene studies

The identification of good candidates has been hampered by lack of knowledge of the underlying pathophysiology of the disorder. Therefore the approach has required both position-independent and position dependent strategies. Examples of the former are the investigation of genes involved in the serotonin or dopamine systems based on evidence of abnormalities in these systems in autism or the identification of the homeobox genes A1 and B1 (HOXA1/B1) based on developmental abnormalities occurring in association with autism.

Position dependent approaches have relied on the identification of candidates from regions of putative linkage highlighted by genome-wide sib-pair studies or using linkage methods in extended pedigrees (e.g. wingless-type MMTV integration site family member (WNT2), Reelin, forkhead box P2 (FOXP2), glutamate receptor, ionotropic, kainite 2 (GRIK-2)) or from breakpoints/ regions defined by cytogenetic abnormalities found in association with autism (e.g. genes in the PWACR).

1.3.7.1 Genes involved in the serotonin system

1.3.7.1.1 Serotonin and autism

Serotonin (5HT) is a ubiquitous neurotransmitter with an indole structure that is found in the peripheral and central nervous systems, blood and the enterochromaffin cells in the gastrointestinal system. (Schain and Freedman 1961) first documented elevated platelet serotonin level in individuals with autism. This is a robust finding that has been well replicated subsequently (Abramson et al. 1989; Cook 1990; Leboyer et al. 1999). There appears to be a 25-50% increase in the levels of 5HT found in autistic individuals and their relatives (Cook 1990). However, levels of 5HT and its metabolites appear to be normal in plasma and urine (Anderson 2002). This has suggested an abnormality in the platelet's 'handling' of 5HT (Anderson 2002). In particular studies have focussed on the activity of the $5HT_2$ receptor subtypes, which are the only class of receptor sub-type characterised on the platelet to date (Peroutka 1990).

The significance of hyperserotonaemia is uncertain. It is known that serotonin is found in the developing brain (Anderson 2002) and is involved in the regulation of neuronal development, i.e. neurogenesis, morphogenesis and synaptogenesis (Lauder 1990; Ramamoorthy et al. 1993; MacKenzie and Quinn 1999). Chugani et al. 1999 demonstrated developmental changes in serotonin synthesis capacity in children. She demonstrated that autistic children had a reduced serotonin synthesis capacity compared with typically developing children. Regions in the brain that are richly innervated with serotonergic neurons appear to be involved in the regulation of behaviours that are abnormal in autism, e.g. amygdala, hippocampus, nucleus accumbens, the medial orbito-frontal cortex, the anterior insular cortex and the anterior cingulated cortex (Kemper and Bauman 1998). In addition medications with a serotonergic effect have been reported to ameliorate symptoms of autism, e.g., Fluvoxamine, a selective serotonin reuptake inhibitor and Risperidone an anti-psychotic with antagonistic effects on serotonin (McDougle et al. 2003; Anderson 2002).

1.3.7.1.2 The serotonin transporter gene (5-HTT)

The 5-HTT gene (SLC6A4) is located at chromosome 17q11.2 and contains 14 exons spread over ~35kb (Ramamoorthy et al. 1993). There are three well characterized polymorphisms within the 5-HTT gene, an insertion / deletion (44bp) in the promoter region, a variable number tandem repeat (VNTR) (12, 10 and 9 repeats) in an intronic region, and a G to T transversion in a putative polyadenylation site in the 3' untranslated region (UTR) of the gene. The basal activity of the long promoter variant has been shown to be three fold higher than that of the short variant (Heils et al. 1996). The 12 repeat allele is the most common variant of the VNTR and has been shown to drive higher expression in embryonic mouse rostral hindbrain (MacKenzie and Quinn 1999).

Two studies reported preferential transmission of the long promoter variant in autism, i.e the variant was transmitted from parents to affected offspring significantly more frequently than might be expected by chance (Klauck et al. 1997a; Yirmiya et al. 2001). A subsequent study reported increased transmission of the short promoter allele in severely affected individuals, but found increased transmission of the long promoter

allele in the sample as a whole (Tordjman et al. 2001). In contrast, increased transmission of the short allele was found in a sample of 86 autistic parent/ child trios (Cook et al. 1997b). Several studies show no association with either long or short alleles (Persico et al. 2000;{Maestrini, 1999 #130;Betancur, 2002 #131}. The characterization of further polymorphisms in the serotonin transporter gene have been described {Kim, 2002 #50}. These authors also described association with several single nucleotide polymorphisms (SNPs) surrounding the VNTR.

An association study of polymorphisms in the 5HTT gene, including SNPs described by Kim et al, 2002, was conducted in the Irish autism sample described in this work. This investigation showed a trend towards preferential transmission of the short promoter allele (χ^2 = 4.5, p = 0.03) (Conroy et al. 2004). A haplotype-based approach to association testing was conducted. A number of haplotypes, especially those involving and surrounding SNP10, showed evidence of association. Transmission of a four-marker haplotype (SNP10-VNTR-SNP18-3'UTR) was the most significant (χ^2 =10.9, p= 0.005). This included the 12 repeat allele of the VNTR, which is associated with increased expression and may play a central role in the early development of the brain in affected probands. Additionally all haplotypes containing allele 1 of SNP 10 were found to show increased transmission in the autism sample. Linkage disequilibrium measurement (D') analysis (discussed in Chapter 2) demonstrated the presence of LD between markers (D' values ranged between 0.513 and 0.788 across the gene). Associated with this investigation, blood samples for platelet serotonin levels are currently being collected from affected probands and their parents to allow functional analysis of the potential effects of the polymorphisms within the gene on protein function.

1.3.7.1.3 Serotonin receptor genes

There are 15 known genes that encode serotonin receptors in humans. With the exception of the 5-Hydroxytryptamine Receptor 2A gene ($5HT_{2A}$) and the 5-Hydroxytryptamine Receptor 7 gene (5HT(7)), none of these have been extensively studied in autism.

The $5HT_{2A}$ gene is located on chromosome 13q and encodes the $5HT_{2A}$ receptor. It consists of 3 exons separated by 2 introns and spans over 20 kb (Chen et al. 1992). There is evidence for genomic imprinting of the gene, which is expressed only from the maternal allele (Kato et al. 1996). The region on chromosome 13q in which the gene is located has emerged from at least one genomewide linkage study as showing putative evidence for linkage with autism (CLSA 1999). An association study conducted between haplotypes consisting of single nucleotide polymorphisms (SNPs) in the gene and an autism sample consisting of 115 trios was negative (Veenstra-VanderWeele et al. 2002). Parent of origin effects were also investigated because of the observation of paternal imprinting of the gene but none were observed. The authors concluded that the sample had insufficient power to detect parent of origin effects because of small sample size.

The 5HT(7) gene is a serotonin receptor that belongs to the G-protein coupled receptor superfamily and is expressed in discrete parts of the limbic brain (Ruat et al. 1993). The gene has been mapped to Chromosome 10q23 and a pseudogene has been identified on chromosome 12p13 (Gelernter et al. 1995; Lassig et al. 1999). A role for the gene in autism was postulated based the antagonistic effects of Risperidone on 5HTR7. An association study in an autistic sample of 53 trios did not detect increased transmissions of a C/T polymorphism in the second intron of 5HTR7 (Lassig et al. 1999). No further published investigations of this gene have been undertaken in autism.

Given the complexity of the serotonin system and the poor understanding of the underlying mechanism of hyperserotonaemia, further studies in autism of genes involved in the sertonergic system are warranted.

1.3.7.2 Dopaminergic system genes

1.3.7.2.1 Dopamine and autism

Dopamine is involved in the control of locomotion, cognition, affect and neuroendocrine secretion in the central nervous system (CNS). Its actions are mediated by five different receptor subtypes. The role of dopamine in autism is unclear, however there are a number of clinical studies that report the amelioration of symptoms in autism in response to

dopamine antagonists, particularly aggression and self-injury (McDougle et al. 1998; Masi et al. 2001; McCracken et al. 2002).

The study of dopaminergic system genes has not as yet been widely undertaken in autism. However one study reported that dopamine-β-hydroxylase levels were elevated in a small sample of children with autism (Garnier et al. 1986) and low maternal levels of dopamine-ß-hydorxylase have been suggested as a possible modifying factor in autism severity (Robinson et al. 2001).

1.3.7.2.2 Dopamine-β-hydroxylase

This protein catalyses the conversion of dopamine into nor-epinephrine. Genetic association studies of an insertion/ deletion in the promotor of the DBH gene and autism showed an excess of del/del genotype in mothers of affected offspring (Robinson et al. 2001). The variant appears to account for approximately 40% of the variation seen in circulating Dopamine-β-hydroxylase (DBH) levels. The authors have speculated that the variant acts as a risk factor for autism, generating a sub-optimal environment in utero, by increasing the dopamine:norepinephrine ratio. The sample described in this thesis has been subject to investigations of serum DBH and variants in the gene through a collaboration with the Child Study Centre, Yale University (see chapter 3, section 3.2).

1.3.7.3 Homeobox A1 and B1 (HOXA1/B1)

These genes are paralogous homeobox genes belonging to the same class of transcription factor genes. HOXA1 and B1 are located on Ch7p15-p14.2 and 17q21-q22 respectively. Analysis of the HOXA1 gene from cDNA cloning predicts a protein of 335 amino acids (Hong et al. 1995). Mouse null mutants for these genes show abnormalities in the development of their hindbrain. The temporal expression of these genes appears to concur with the postulated timing of a CNS insult in autism (Murphy and Hill 1991). (Ingram et al. 2000a) reported an association between a variant in HOXA1 and autism. The association was not replicated in three further autistic samples (Devlin et al. 2002) (Li et al. 2002; Talebizadeh et al. 2002). An association study between the described

variants and autism in the Irish autistic sample was conducted. The results of this investigation are discussed in Chapter 5.

1.3.7.4 Wingless-type MMTV integration site family member 2 (WNT2)

This is a neurodevelopmental gene that maps to 7q31-q33. It was first cloned from a human lung cDNA library in an effort to clone a gene for cystic fibrosis. The cDNA is comprised of 2, 301bp with 5 exons (Wainwright et al. 1988). The WNT pathway relies on the dishevelled family of proteins (Dvl) to transmit its signal. Dvl knockout mice show reduced social interaction (Lijam et al. 1997). WNT-2 has been demonstrated to be involved in the development and patterning of the vertebrate central nervous system (Landesman and Sokol 1997; Uusitalo et al. 1999; Hauptmann and Gerster 2000).

It was identified as a potential candidate for autism as it maps to a putative linkage region for autism identified by the International Molecular Genetics Consortium in Autism (IMGSAC) (Wassink et al. 2001). A second linkage study also described evidence for linkage in an overlapping region (McCoy et al. 2002; Shao et al. 2002a; Shao et al. 2002b). There have also been several reports of chromosomal abnormalities in this region (Ashley-Koch et al. 1999; Vincent et al. 2000; Warburton et al. 2000). Furthermore, an expressive speech and language disorder has been mapped to 7q31 (Lai et al. 2000) which suggests that the region may contain susceptibility genes for abnormalities of speech and language development.

Wassink et al, 2001 identified two families where mutations in WNT2 segregated with speech and language difficulties (Wassink et al. 2001). In one family the mutation was a C-T transition at nucleotide 1189 in exon 5. This was a functional mutation producing Arg299Trp in the encoded protein. In the second family, a T-to-G transversion was identified at nucleotide 14 in exon 1 producing a Leu5Arg change in the encoded protein. These mutations were not identified in a sample of individuals with autism, however 2 further variants were identified in their sample, a SNP in the 3'UTR (C-T) and a further SNP (C-T) 0.5kb upstream from the WNT2. In the same study evidence for LD between the 3'UTR and an autism sample comprised of affected sib-pairs and parent/ child trios. Analysis revealed that LD was strongest within a sub-set of the sample that was

language-impaired. A subsequent study did not find evidence of association with WNT2 in an autism sample (McCoy et al. 2002).

1.3.7.5 Reelin (RELN)

RELN maps to Ch7q22 (D'Arcangelo et al. 1995), It is a large gene encoding mRNA of approximately 12kb (DeSilva et al. 1997). The gene product is a large extracellular glycoprotein secreted by the most superficial layers of the brain (Cajal-Retzius Cells in the neocortex and hippocampus, granule cells of the external granular layer of the cerebellum). It is expressed in fetal and postnatal brain as well as in liver. The expression in postnatal human brain is high in the cerebellum, a region shown to be enlarged in autism. The RELN protein directs cortical layer formation by acting on migrating cells (e.g. cortical plate neurones in cortex and cerebellar Purkinje cells) through its interaction with the adaptor protein, disabled 1 (dab1) (D'Arcangelo et al. 1999). Splice mutations are associated with an autosomal recessive form of lissencephaly (Hong et al, 2000). A 'Reeler' mouse (autosomal recessive mutation) demonstrates impaired motor co-ordination, ataxia and tremor (Tueting et al. 1999). The expression of RELN was reduced in areas of post-mortem brain (pre-frontal and temporal cortices, hippocampi, caudate nuclei and cerebelli) in subjects with schizophrenia (Impagnatiello et al. 1998). In relation to autism, evidence for linkage around marker D7S477 that maps to 7q22 has been reported (IMGSAC 2001). This region contains the RELN gene. An association was identified between autistic disorder and a polymorphic GGC repeat located 5' of the reelin gene (RELN) ATG initiator codon (Persico et al. 2001). Haplotypes formed by this polymorphism with two single-base substitutions located in a splice junction in exon 6 and within exon 50 were also found to be associated. This finding has not been widely replicated (Krebs et al. 2002; Zhang et al. 2002) . However Zhang et al, 2002 have proposed that there is an increase in transmission of the variant in autistic individuals with delayed phrase speech.

1.3.7.6 Forkhead Box P2 (FOXP2)

This gene is a putative transcription factor, containing a polyglutamine tract and a forkhead DNA binding domain that maps to Ch7q31 (Lai et al. 2001). It consists of 17

exons with 2 additional exons at the 5' end that are alternatively spliced. Four alternative splice variants exist with six previously unidentified exons now characterized (Bruce and Margolis 2002). The gene maps to a region identified as showing evidence for linkage (IMGSAC 1998; McCoy et al. 2002). A translocation breakpoint between exons 3b and 4 was identified in an individual with a severe expressive speech and language disorder (Lai et al. 2000). The authors demonstrated a point mutation Arg553His in the forkhead domain. No association has been identified between this gene and autism (Newbury et al. 2002; Wassink et al. 2002). A study testing linkage and association in a sample of individuals with speech and language impairment (SLI) markers with markers around FOXP2 did not find linkage or association with these markers but found an association with a marker in the CFTR locus, located approximately 3.3Mb downstream from FOXP2 and a further marker, D7S3052, located 4Mb upstream from the gene (O'Brien et al. 2003). The authors have suggested that genetic factors or regulation of common language impairment may reside around 7q31.

The evolution of FOXP2 provides an interesting insight into the mechanism of language development in humans and may ultimately help to inform candidate gene studies in other disorders of speech and language development including autism. The cDNAs encoding the FOXP2 protein have been sequenced in the chimpanzee, gorilla, orangutan, rhesus macaque, and mouse and compared with the human cDNA (Enard et al. 2002). FOXP2 is among the 5% most-conserved proteins. The chimpanzee, gorilla, and rhesus macaque FOXP2 proteins are identical and carry only 2 differences from the human protein. The orangutan carries 3 differences from humans. The authors have suggested that the human-specific change of position 325 creates a potential target site for phosphorylation by protein kinase C together with a minor change in predicted secondary structure that may affect protein function related to fine oro-facial movements, allowing for the development of spoken language in humans (Enard et al. 2002). They showed that human FOXP2 contains changes in amino acid coding and a pattern of nucleotide variation, which strongly suggests that this gene has been the target of selection during recent human evolution.

1.3.7.7 Glutamate receptor, ionotropic, kainate 2 (GRIK-2)

This gene maps to chromosome 6q21 within a region of putative linkage to autism (Phillippe et al. 1999). It is an inotropic kainate receptor belonging to the human glutamate receptor-6 kainate-preferring receptor family. It is expressed in the cerebral and cerebellar cortices during development and undergoes post-transcriptional editing (Paschen et al. 1994). Glutamate is a predominantly excitatory neurotransmitter within the brain and both glutamate and its receptors are directly involved in memory and learning (Shimizu et al. 2000). Altered glutamate levels have been demonstrated in autism with elevated levels in plasma and reduced levels in platelets compared with controls (Rolf et al. 1993). In addition it has been suggested that autism is a hypoglutamatergic condition based on the similarities seen between it and the effects of glutamate antagonists in healthy subjects (Carlsson 1998). Animal models of hypoglutamatergia show similarities with autism, such as defective habituation, restricted behavioural repertoire and inability to change the behavioural programme (Carlsson 1998). The genomic structure of the gene has been characterized (Jamain et al. 2002a). It is a large gene (670kb) and has 16 exons. There are three sites of post-transcriptional editing leading to amino acid changes at the protein level. Intron 15 contains two alternatively spliced exons which lead to differing –COOH terminal sequence and premature ending of the sequence. Using the affected sib-pair method and the transmission disequilibrium test, Jamain et al, 2002 reported evidence of both linkage and association between the gene and an autistic sample (Jamain et al. 2002a). TDT testing in affected sib-pairs showed evidence of increased maternal transmission. Mutation screening in 33 affected individuals revealed a non-synonymous SNP in a highly conserved domain of the gene. This SNP occurred in 8% of the autistic sample. The functional significance, if any, of the SNP remains unknown.

1.3.8 Linkage studies.

In an effort to localise susceptibility loci in autism genome-wide linkage studies have been undertaken. This method involves the screening of the entire genome in multiply affected families in order to identify regions of the genome showing statistical evidence of linkage. The first genome-wide linkage study was published in 1998 by the

International Molecular Genetics Consortium in Autism (IMGSAC) (IMGSAC 1998). This study identified several regions of interest with the strongest evidence occurring on chromosome 7q. There have been nine published genome-wide linkage studies published to date {IMGSAC, 1998 #43;CLSA, 1999 #16;Phillippe, 1999 #68;Risch, 1999 #73;Buxbaum, 2001 #12;Liu, 2001 #182;IMGSAC, 2001 #44;Auranen, 2002 #183;Shao, 2002 #187}. The most consistent evidence for linkage emerging from these studies is chromosomes 2q (Buxbaum et al. 2001; IMGSAC 2001; Shao et al. 2002a; Shao et al. 2002b) and 7q (IMGSAC 1998; CLSA 1999; Phillippe et al. 1999; IMGSAC 2001). Other regions achieving Lod scores in excess of 1, in both single and multiple studies include 1p (Risch et al. 1999; Auranen et al. 2002), 3p (Auranen et al. 2002), 4p (IMGSAC 1998), 5q (Liu et al. 2001), 6q (Phillippe et al. 1999; Auranen et al. 2002), 13q (CLSA 1999), 16p (IMGSAC 1998; IMGSAC 2001; Liu et al. 2001), 17q (IMGSAC 2001), 19p (Buxbaum et al. 2001) and Xqter (Liu et al. 2001; McCoy et al. 2002; Shao et al. 2002a; Shao et al. 2002b; Auranen et al. 2002). These data are shown in table 1.4.

1.3.8.1 Chromosome 2q.

Two studies reported strong evidence for linkage on 2q (Buxbaum et al. 2001; IMGSAC 2001). In the IMGSAC study (n= 152 sib-pairs) single- and multi-point analyses were conducted in the sample. Subjects in the sample were defined according to clinical presentation as 'case-type'1, 'case-type' 2 and 'case-type' 3. 'Case-type' 1 met narrow criteria for autism (ADI/ ADOS-G) and had a history of speech and language delay. 'Case-type' 2 had a clinical diagnosis of autism (one point under the cut-off of one ADI-domain), met ADOS-G criteria for pervasive developmental disorder and did not have a history of speech and language delay. Case type 3 had a clinical diagnosis of autism as defined for case-type 2 but did not meet the ADOS-G criteria for PDD-NOS. Statistical analyses were conducted in the entire sample (n=152) and in the following groups: 1/ both sib-pairs type 1; 2/ typel/ type 2 pairs; and 3/ the two groups combined. The data from the entire genome scan is illustrated in Figure 1. 1. Multipoint analysis in the entire sample generated a maximum lod score (MLS) of 3.74 at D2S2188.

Similar analyses were conducted for the sub groupings outlined above. In type1/ type 1 sib-pairs a MLS of 1.24 at D2S2188 was obtained. More sharing was observed for type 1/ type 2 pairs (MLS on ch2>1.5). The highest lod score was obtained for a group comprised of both type 1/ type 1 and type1/ type 2 pairs (MLS= 4.8 at D2S2188).

Buxbaum et al, 2001 conducted two-point analyses between microsatellite markers on chromosome 2q and autism (Buxbaum et al. 2001). Analyses under the dominant and recessive modes of inheritance were performed (penetrance set at 50%). The cohort was divided into Class I and Class II families for the purposes of the analyses. Class I families consisted of sib- or family-pairs where both were affected with a narrow definition of autism. Class II families consisted of one member with a narrow diagnosis of autism and the second with a diagnosis of borderline autism/ Asperger Disorder. Linkage analyses using non-parametric estimates of sharing (NPL) were performed using GENEHUNTER. Two-point analyses in the entire sample (Class I and Class II families) showed strongest evidence for linkage on chromosome 2q at marker D2S364 (HLOD under the dominant model = 2.25). A further marker, D2S335, located approximately 11cM towards the centromere from D2S364 showed some evidence for linkage (HLOD under dominant model = 1.2). The NPL score for the entire sample at D2S364 was 2.45.

Multipoint analyses were conducted under the same conditions. In addition the analyses were performed in subsets defined by the presence or absence of speech and language delay (absence of phrase speech >36 months). In Class I families with phrase speech delay a non-parametric lod (NPL) score of 3.29 at 175cM was identified. Marker D2S335 is located 2.9cM centromeric of marker D2S2188 and marker D2S364 is 6.8cM telomeric of it (D2S2188 is the marker with greatest evidence of linkage in the IMGSAC study).

Despite the strength of the findings from both studies, the linkage peaks are broad. The baseline width of the IMGSAC peak is approximately 75cM, with the 1 Lod unit support region ~15cM. The linkage peaks reported by Buxbaum are similarly wide and somewhat bimodal, with two markers in the region suggestive of linkage (D2S364 – NPL = 2.39 and D2S335 - NPL = 1.2).

A third report of genomewide analysis of 99 affected-sib and affected-relative pairs found statistical evidence for linkage on chromosome 2q with a NPL of 1.16 at marker D2S116 (Shao et al. 2002b). This marker is located 26.1cM and 18.5cM telomeric of markers D2S2188 and D2S364 respectively. Shao conducted further analysis on the data using the population sub-grouping described by Buxbaum et al (2001), dividing the sample into those with and without phrase speech delay (PSD) (Shao et al. 2002a). The MLS for the families with PSD increased to 2.86 for marker D2S116. A further marker, D2S2309 (at 204.6cM) that previously had a MLS of 0.57 attained a MLS of 1.58 in the PSD families.

Buxbaum and Shao suggest that a sample restricted on the basis of PSD represents a more genetically homogenous sample. Their findings support the hypothesis that grouping families on the basis of reliably measured language traits may increase the power for linkage analysis (Folstein and Mankoski 2000; Bradford et al. 2001). The results of analysis in the IMGSAC study are in conflict with this. The best MLS attained in this study was for analysis in a group that included both individuals with and without PSD. The scores attained for the families with PSD only were substantially lower. However when all families were analysed, including those with broader criteria, the MLS attained was still lower than the analysis for the type1/ type2 families. This is perhaps suggestive that reducing clinical heterogeneity may help to increase the power for linkage analysis, but the PSD sub-grouping may be unnecessarily restrictive.

The linkage signals obtained from these studies are broad and make localisation of a disease gene/ region difficult. A number of genes within the region have been proposed as candidates. The characterization of a chromosomal translocation and its implication for defining the boundaries of the locus are described in Chapter 6.

Figure 1. 1 Chromosomal regions showing evidence for linkage in the IMGSAC genomewide linkage study (IMGSAC 2001).

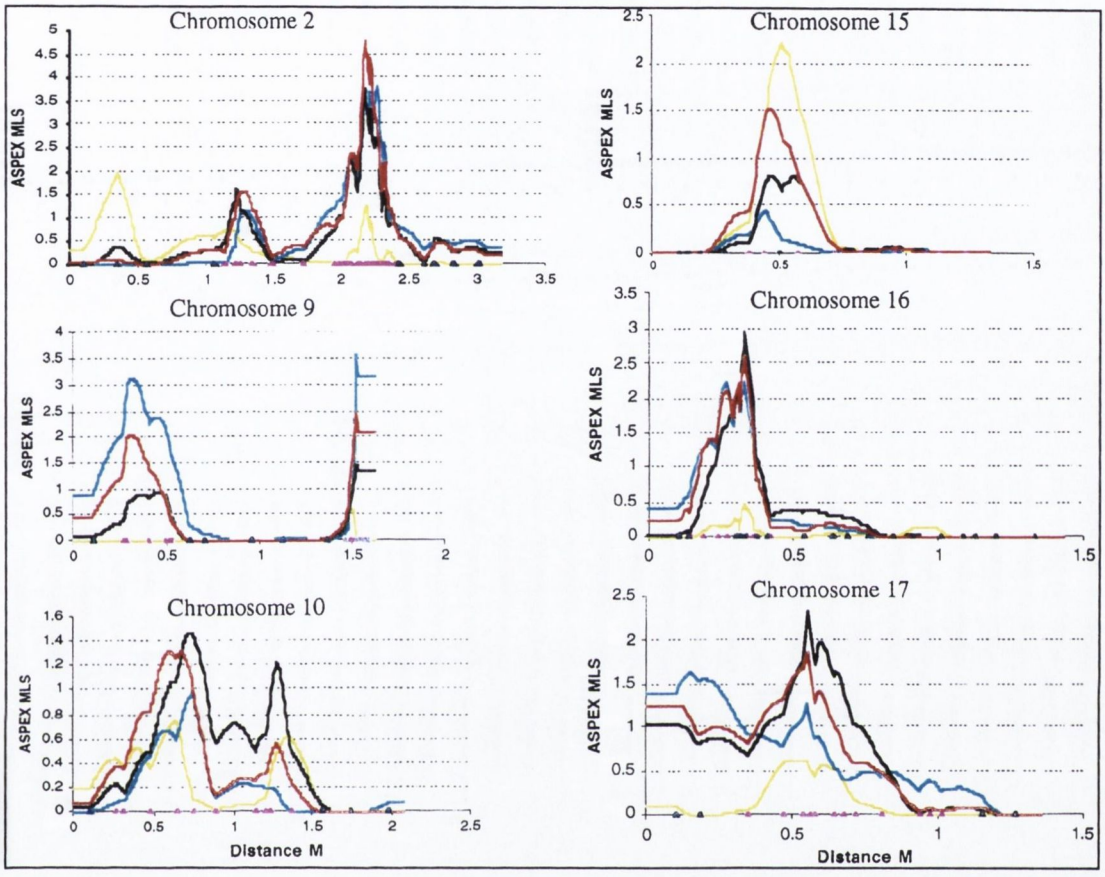
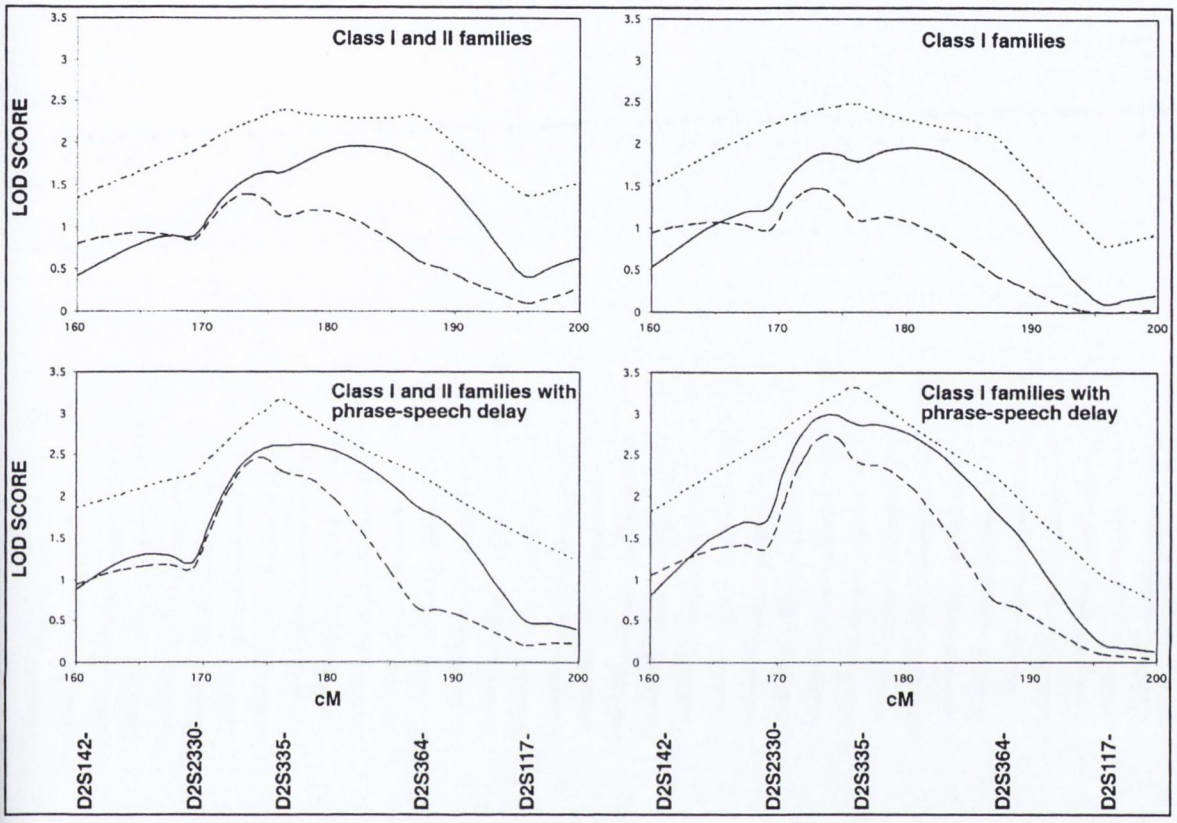
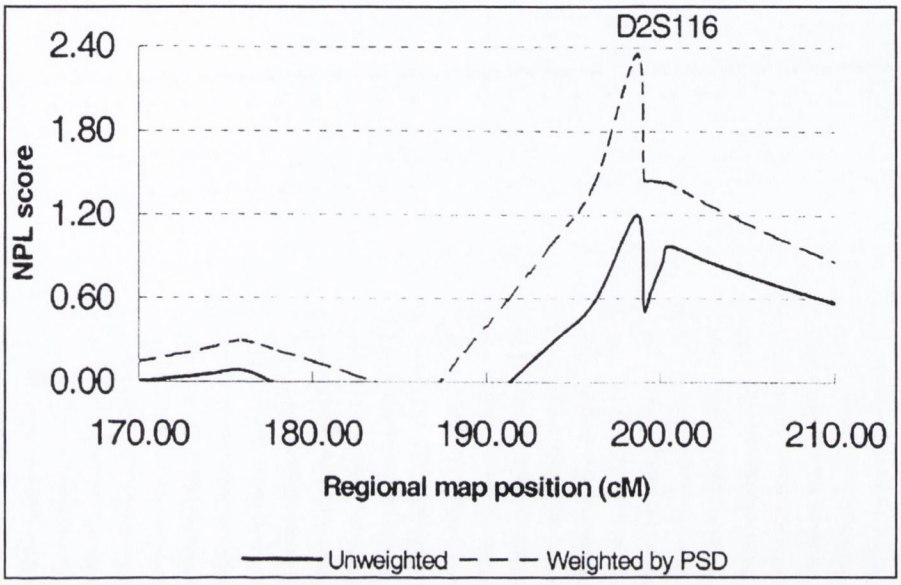


Figure 1. 2 Results of multipoint linkage analysis and autism on chromosome 2q based on presence and absence of speech delay (Buxbaum et al. 2001).



Multipoint LOD analyses under dominant (solid lines) and recessive (dashed lines) modes of inheritance were performed with penetrance set at 50% and allowing for locus heterogeneity (HLOD). Multipoint linkage analyses using nonparametric estimates of sharing were also performed using NPL (dotted lines) with GENEHUNTER {Kruglyak, 1996 #401}.

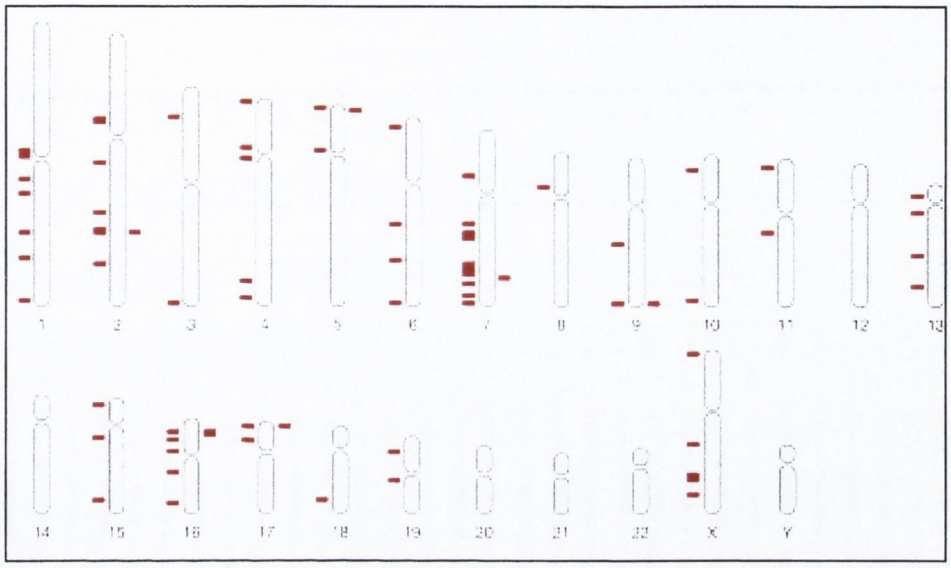
Figure 1. 3 Non-parametric linkage analysis on chromosome 2q at marker D2S116 with autism for families with and without phrase-speech delay (Shao et al. 2002a).



1.3.8.2 Chromosome 7q.

Several genome-wide sib-pair studies have identified statistical evidence of putative linkage on chromosome 7q (IMGSAC 1998; CLSA 1999; IMGSAC 2001; Shao et al. 2002b). Unlike chromosome 2q the linkage peaks from individual studies do not overlap. IMGSAC reported a MLS of 2.53 between markers D7S530 (at 134.5cM) and D7S684 (at 147.2cM) (IMGSAC 1998). In the follow up genome scan, they reported a maximum MLS of 3.20 near marker D7S477 (at 111.8cM) (IMGSAC 2001). D7S477 is approximately 22.7cM telomeric of D7S530 and 35.4cM telomeric of D7S684. The Collaborative Linkage Study in Autism (CLSA 1999) reported a maximum multipoint lod score (MMLS) of 2.2 around marker D7S1813 (at 103.6cM). This marker lies 8.2cM telomeric of D7S477. Shao et al, 2002, reported a MLS of 1.66 around D7S495 (at 144.72cM), which lies at the 3' end of the linkage region reported by (IMGSAC 1998). The region of putative linkage defined by these studies is 7q22-q31, approximately 43.6cM. Given the difficulties narrowing linkage signals mentioned above, it is obvious that the identification of candidate genes from such a broad region will not be easy. As in the analyses conducted on chromosome 2q, higher MLS scores were obtained when the analyses were restricted to families with speech and language disorder and it has been proposed that the region may harbour genes specifically for disorders of speech and language development (Folstein and Mankoski 2000; Bradford et al. 2001). The Forkhead Box P2 (FOXP2) gene discussed above maps to 7q31 and is thus at the telomeric end of this region. Other evidence suggestive of the presence of susceptibility genes for autism includes reports of cytogenetic abnormalities in the region in patients with autism (Gordon et al. 1994; Ashley-Koch et al. 1999; Vincent et al. 2000; Yan et al. 2000b). It is likely that joint analyses of the findings from the various groups would help to refine this region better. Unfortunately different marker maps were utilised and so this is not straightforward but a meta-analysis of the findings across various studies would be useful.

Figure 1. 4 Overlap of linkage signals from published genome screens (Folstein and Rosen-Sheidley 2001).



1.1.21.3 Other regions of interest.

Some of the regions of suggestive linkage are interesting with respect to the presence of potential candidate genes. For example a MMLS of 2.34 around a polymorphism in intron 2 of the 5-HTT gene was identified (IMGSAC 2001). Other regions have been identified as providing modest linkage signals in more than one study. Results of the genome scans are summarised in Table 1.4.

Table 1. 4 Data from published genome-wide affected sib-pair studies showing locations of markers where maximum lod scores were identified.

Linkage Region	IMGSAC 1998	CLSA 1999	Philippe et al 1999	Risch et al 1999	Buxbaum et al 2001	Liu et al 2001	IMGSAC 2001	Shao et al 2002	Auranen et al 2002
1p				D1S1675					D1S1675
2q					D2S364 D2S335		D2S2188	D2S116	
3p								D3S3680	D3S3038 D3S3659
3q									D3S3038- D3S369 D3S3730
4p	D4S412								
5q						D5S2494 D5S2488			
6q			D6S283						D6S1021
7q	D7S530 D7S684	D7S1813					D7S477	D7S495	
13q		D13S217- D13S1229 D13S800							
16p	D16S407- D16S3114 D16S412					D16S2619	D16S3102		
17q							HTTINT2		
19p					D19S714 D19S587- D19S601				
Xq						DXS1047		DXS6789	DXS6789 DXS7132
No. of affected sibs	N=87	N=75	N=51	N=147	N=95 (narrow defn.=49)	N=118 (narrow defn.=75)	N=152 ('strict defn.=127)	N=96	N=38

1.4 Specific contributions of current study

1.4.1 Establishment and management of a genetics study in autism

This project commenced following the award of a Wellcome Trust Mental Health Training Fellowship. The aim of the project was to establish a group of individuals with autism and their parents sampled from the Irish population for the purposes of a genetics study. One of the main objectives was to gather optimal phenotypic data using the methods as outlined in Chapter 3. The aim was to collect as many cases as possible within the scope of the available resources. As it was clear from the outset that a single investigator would be limited in their ability to recruit a large sample and conduct all the necessary laboratory work, additional funding was sought to recruit additional staff to the project. A significant proportion of time was devoted by the author to the management of the project and grant writing in addition to the clinical and laboratory aspects of the research. An important additional objective was to establish a research group that would continue to use the valuable sample to conduct genetic investigations in autism.

1.4.2 Funding applications

In order to meet this objective, additional funding was required. During the three years of the project the author was responsible for grant applications to maximize the resources of the study. Project grants were successfully awarded from the Health Research Board in Dublin and the National Alliance for Autism Research (USA). This funded the consecutive employment of two research nurses on the clinical aspects of the project in addition to some clinical running costs. One of these nurses worked with the author on the sample collection for the sample reported here. The second was employed more recently to recruit a follow-up clinical sample and this work is currently in progress (although as indicated in the Statement of Work, this nurse collected missing data from the subjects recruited in the initial sample collection). A PhD student and Post-doctoral researcher were employed in the laboratory to work on other aspects of the genetics of autism and to follow-up the findings of this study. The author was responsible for the co-ordination of both the clinical and laboratory sides of the project in addition to completing laboratory work. This also involved the supervision and training of clinical researchers and co-supervision of laboratory personnel.

1.4.3 Molecular genetic investigations in autism

This study was designed to examine, using an association design, regions of linkage identified by other groups. Initially it was proposed to conduct a linkage disequilibrium experiment and association analysis across chromosome 7, as this was the region of interest that emerged from earlier linkage studies. However, in the course of the ascertainment and assessment of cases, a chromosomal abnormality was identified in a single case. In collaboration with colleagues at the National Centre for Medical Genetics in Our Ladies Hospital for Sick Children, it was decided to map the abnormality on the molecular level and concentrate molecular genetic studies within this region and its breakpoints.

A linkage disequilibrium study of variants within the HOXA1 and HOXB1 genes was undertaken prior to the investigations on chromosome 2q. This was the first molecular investigation undertaken in the sample following the extraction of DNA from blood and buccal swab samples. This is discussed in Chapter 5.

Methodological considerations specific to this type of study are considered in the next chapter.

Chapter 2

Methodological considerations in genetic studies of complex disorders

2.1 Introduction

As discussed in Chapter 1, finding susceptibility genes in complex disorders is complicated by factors such as diagnostic heterogeneity and phenocopies. A lack of understanding of the underlying pathophysiology of the condition hampers the identification of candidate genes. The background to minimising diagnostic heterogeneity in the sample has already been discussed. This chapter will discuss the strategies utilised in the identification of candidate genes and regions in complex genetic disorders and some of the important underlying factors that influence such studies.

2.2 Positional cloning methods of gene identification

Where gene identification proves difficult positional cloning methods are adopted in the hunt for susceptibility genes. These utilise the combined approaches of linkage analysis (using family data) and association analysis (using population data) and allows the systematic search of the genome for disease-causing genes without *a priori* knowledge of disease pathogenesis.

The work described here involved linkage disequilibrium mapping of a candidate gene and region for autism using multi-allelic microsatellite markers and single nucleotide polymorphisms (SNPs). As discussed in chapter one (section 1.3.8.1) there has been converging evidence for a susceptibility region for autism on chromosome 2q emerging from genomewide linkage studies. It is this region that has been the subject of fine mapping in this work. There follows a brief introduction of linkage methods and linkage and association methods in relation to positional cloning of susceptibility genes in complex disorders. This will be followed by a discussion on the factors that influence linkage disequilibrium (LD) and how that impacts on mapping a complex disease gene.

2.2.1 Linkage analysis

This method seeks to identify chromosomal regions likely to contain disease loci by scanning the genome for one or more genetic markers that do not assort independently with disease status, i.e departure from Mendel's Second Law (McGuffin 1994). Where alleles of a polymorphic marker and disease locus are transmitted together more frequently than might be expected by chance, they may be inferred as being linked, and are in close physical proximity on a chromosome. Linkage studies look for the co-segregation of marker and disease within affected related individuals. It is most effective when there is a close correlation between the phenotype and the inferred genotype at an unknown disease locus that is being tested for co-segregation with markers. Therefore it is effectively used to map disease loci in Mendelian disorders (Risch 2000). Linkage studies in multiply affected families measures the strength of evidence in favour of linkage as the lod score. This is the logarithm (base 10) of the ratio of the likelihood of the observed genotypes given a recombination fraction less than 0.5 compared with the likelihood under non-linkage, i.e. with the recombination fraction equal to 0.5. Traditionally a lod of 3 or more is taken as "significant" evidence for linkage {Ott, 1999 #420}. In addition to phenotypic heterogeneity and phenocopies other factors influencing the less than 100% correlation between the phenotype and the inferred genotype include genetic heterogeneity, low penetrance, and possibly high frequency of susceptibility alleles. In addition, traditional linkage methods require that the mode of inheritance should be specified in the analysis, however this is not usually known for complex genetic disorders. Non-parametric methods using the affected sib-pair and affected pedigree methods do not rely on a specified mode of inheritance (Weeks and Lathrop, 1999). The degree of allele sharing at a specified locus is assessed between related affected individuals. Hypothetically, if a disease predisposing mutation exists in a high proportion of affected families then affected individuals within each family will share alleles linked to the disease-causing locus more often than would be expected by chance.

Linkage analysis has been successful in identifying disease-causing alleles in Medelian disorders, e.g. Cystic Fibrosis, Huntingtons Disease. Identification of susceptibility genes in complex diseases has not been so easy using this approach. For the purposes of

positional cloning there is large variation between linkage peaks relative to the physical location of disease genes (Roberts et al. 1999). Therefore other strategies may need to be employed to identify susceptibility genes within regions showing evidence of linkage.

2.2.2 Association studies.

These are population-based studies comparing genetic variation in a disease group (cases) versus a non-disease group (controls). Studies may focus on known candidate genes or regions. Association studies rely on the phenomenon of linkage disequilibrium (LD) and are referred to as LD mapping. LD is the non-independence within the population of alleles at two separate loci. Polymorphisms (both microsatellites and SNPs) are genotyped within a candidate region in anticipation that some are in LD with the disease-causing variant. Studies can be performed in case-control or trio-based samples. As trio-based association mapping forms the basis of this work, a discussion will follow outlining the principles of LD, the factors influencing LD, comparisons of study designs and how underlying LD structures may be used to inform LD mapping.

2.2.3 Linkage disequilibrium

Linkage disequilibrium (LD) may be defined as the non-random association between two loci. It occurs where a marker allele lies close to a disease susceptibility allele and these loci are inherited together over many generations. Thus the same allele is detected in affected individuals in multiple unrelated families. For example when a disease-causing mutation arises then all variants on the chromosome are transiently associated with the new variant. But this relationship decays relatively quickly as a result of recombination (section 2.2.3.1) and the rate of decay is directly related to the distance separating the loci. With time and an increasing number of recombination events, only those loci that are close to the new variant remain on a stable haplotype and these are said to be in linkage disequilibrium. LD relationships within the genome are influenced by several factors, discussed below.

2.2.3.1 Recombination

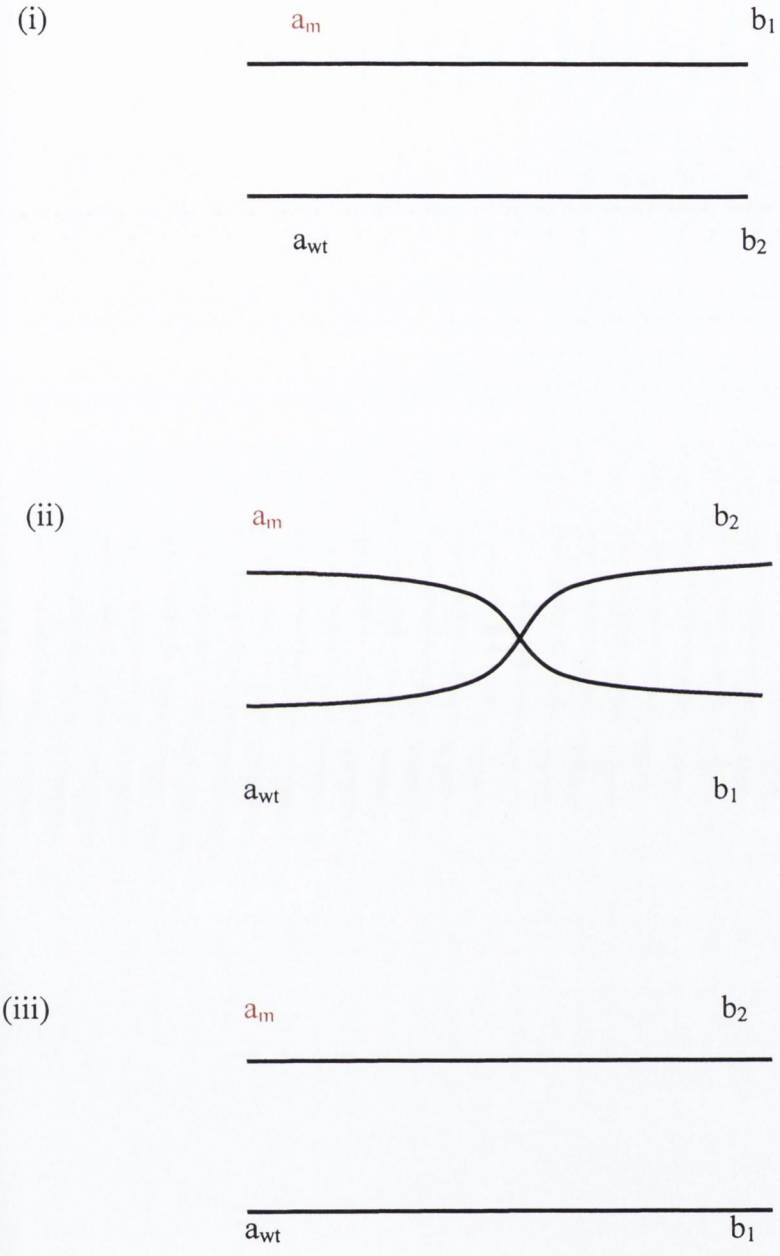
Recombination events usually arise as a consequence of crossing-over between neighbouring pairs of chromosomes. The frequency of recombination events between two loci effects LD between them. The number of recombination events is determined by 1/ the recombination fraction between the loci and 2/ the number of doubly heterozygous crossing over events that have occurred between the current population and the initial mutant chromosome. Recombination rates vary widely across the genome, and thus association is affected accordingly. Generally recombination is directly related to the distance between loci. Thus, the further apart loci are, the greater the chance is for recombination to occur and weaken the LD.

Length of time since the mutation arose affects LD. Recombination occurs on average once per chromosome for each meiosis. Therefore the greater the number of meioses (or generations) since the mutation arose, the greater the chance that a crossing-over event will have occurred between them.

To illustrate how recombination affects association between loci, consider the following example:

A mutation a_m arises at locus A, close to neighbouring locus B, with two alleles b_1 and b_2. The mutation occurs on the same chromosome carrying the b_1 allele. For a short period of time alleles a_m and b_1 are associated. Assuming that the mutation is disease causing then the b_1 allele will be transmitted with the disease. However if a crossing-over event occurs between the loci, the disease causing mutation will come to lie on the same chromosome as allele b_2, thus a_m is no longer associated with b_1 and transmission of both alleles b_1 and b_2 will be observed with the disease. This is illustrated in Figure 2.1.

Figure 2. 1 Recombination event leading to changes in LD relationship.



The extent of LD is inversely proportional to the local recombination rate. It has been proposed that the genome is arranged into discrete haplotypic blocks of low recombination frequency separated by recombination 'hot spots' (Daly et al. 2001), (Dawson et al. 2002; Gabriel et al. 2002). Knowledge of the underlying haplotypic structure in a region under study is useful in understanding the LD relationship between genetic markers and in the identification of high-risk haplotypes within the region. As there is little variation in LD within haplotypes, genotyping of a single SNP on the haplotype may be almost as powerful in detection of LD between haplotypes and disease carrying haplotypes (Johnson et al. 2001). This method of haplotype-tagging is more cost effective due to a reduction in the amount of genotyping required. Haplotypes are likely to vary in their size in differing populations. It has been demonstrated that haplotypes in older populations are smaller as there has been greater length of time for the haplotypes to be eroded by recombination. Younger populations, e.g., Asian and European, have larger haplotypes (Gabriel et al. 2002). The extent of LD in the Irish population is unknown and there is no published data concerning underlying haplotypic structure. Kendler et al., 1999, have shown that for the regions of the genome they studied with dense marker maps they were able to identify LD at the 5% level of significance in 96% of marker pairs within 0.5cM distance from each other, and 67% of pairs 0.5-1cM apart, 35% of pairs 1-2cM apart, 15% of pairs 2-4cM apart and 8% of pairs 5-10cM apart.

2.2.3.2 Variable mutation rates.

Mutation at a locus will also affect LD. If two alleles are associated at two loci, and a mutation event occurs at one locus, then it is possible the two alleles may no longer be associated. Some SNPs, particularly those in CpG islands have high mutation rates and are therefore not in LD with surrounding markers.

2.2.3.3 Genetic drift.

Genetic drift is the phenomenon of changes in gene or haplotype frequencies between generations due to random sampling of gametes. This effect is more pronounced in small populations. Increasing drift in these populations will increase the LD as some haplotypes

are lost. The influence of genetic drift has been demonstrated in the Irish population, based on geographical location, although it is not pronounced (Hill et al. 2000).

2.2.3.4 Admixture

LD between loci can be created by admixture or migration. When a population is affected by inward migration initially the LD is directly proportional to the allele frequencies in the individual populations (Chakraborty and Weiss, 1988). Subsequently, spurious LD between unlinked markers dissipates and LD between neighbouring markers is broken down by recombination.

2.2.3.5 Population bottlenecks, inbreeding and assortative mating

Population structure is important in how it determines patterns of LD in human populations. If a population goes through a bottleneck, where the population is reduced to a small number of founders by external events, haplotype diversity is reduced and LD extends over greater distances. Theoretically, LD between loci and disease causing alleles should be easier to detect in these populations.

Assortative mating occurs where individuals with similar phenotypes breed resulting in the increase in disease causing alleles in further generations, which in turn affects the LD relationships with surrounding loci.

2.2.3.6 Selection

Natural selection is the process whereby some of the inherited genetic variation will result in differences between individuals with respect to their ability to survive and reproduce successfully. Selection can be positive or negative based on the advantage or disadvantage conferred by the variant. A negative selection reduces fitness and in the absence of new mutations reduction in fitness thereby leads to reduction in the frequency of the mutant allele. In some cases, selection can act in the opposite direction by increasing fitness, particularly in autosomal recessive disorders where heterozygous status improves biological fitness, the 'heterozygote advantage'. The best example of this is the adaptation of the Thallosaemia group of illnesses. Heterozygotes show immunity to malaria, while

homozygotes develop thallosaemia with the consequent loss of fitness. The heterozygote advantage, however, led to high morbidity due to thallosaemias in Mediterranean and African populations.

Selection can affect LD in a number of ways:

(1) Hitchhiking: Where a region surrounding a favoured variant is elevated to higher frequency along with the variant.

(2) Selection against deleterious variants: Where deleterious haplotypes are eradicated from the population, this can lead to inflated LD.

(3) Epistatic selection for combinations of alleles on the same chromosome. This can cause association between different alleles at the same loci. (This has not been demonstrated in humans).

2.2.3.7 Gene conversion

This is where a short stretch of one copy of a chromosome is transferred to the other during meiosis. It is the equivalent of two very closely spaced recombination events and breaks down LD in a manner similar to recombination/ mutation. Rates of gene conversion are high in humans and are important in LD between tightly linked markers (Frisse et al. 2001), (Quintana et al. 2001) (Ardlie et al. 2002)

2.2.3.8 Measures of LD.

Lewontin's D is a measure of LD and is defined as the difference between the observed frequency and the expected frequency if two loci are segregating at random. For two loci Aa/ Bb, the frequency of a two-marker haplotype is P_{AB}. Assuming random assortment, $P_{AB} = P_A \times P_B$. Thus $D = P_{AB} - P_A \times P_B$. As D relies on allele frequencies its numerical value has little value in terms of measuring LD. Alternative measures of D have been developed. The most widely used measures for LD are D' and r^2. These measures are not identical, they have differing properties and measure different aspects of LD.

D' is calculated by dividing D by its maximum possible value, given allele frequencies at two loci. D'=1 if two loci are not separated by recombination (or recurrent mutation/ gene

conversion). When D'=1, the loci are said to be in complete LD. One of the problems with D' is that with relative values of D'< 1 there is no clear interpretation. D' is strongly inflated in small samples, particularly for rare alleles. Due to these sample size effects, LD in different samples can be difficult to compare. Intermediate values should not be used to compare the strength of LD between studies (Ardlie et al. 2002)

r^2 is the measure of choice for quantifying and comparing LD in association mapping (Pritchard and Przeworski 2001) (Weiss and Clark 2002). It is based on the correlation of alleles at two loci and is calculated by dividing D^2 by the sum of the frequencies of four alleles at two loci, i.e. for loci Aa and Bb:

$$r^2= \quad D^2/ P_{AB} \times P_{Ab} \times P_{aB} \times P_{ab}$$

$r^2=1$ if the two loci have not been separated by recombination and have the same allele frequency. This is known as 'Perfect LD'. An advantage of the measure is that individual values of r^2 are more easily interpreted. Consider two neighbouring loci where one is a disease causing allele and the other a nearby marker that are in LD = r^2. If a sample size x is required to detect the association between the sample and disease locus, a sample size of x/ r^2 is required detect association at the neighbouring locus in the sample (Kruglyak 1999; Pritchard and Przeworski 2001). Secondly, r^2 takes account of differences in allele frequencies between two markers. Finally, the calculation is less inflated in small samples.

2.2.4 Choice of population

The correct choice of population is important in genetic studies in complex disease. LD mapping studies use case-control and trio populations.

2.2.4.1 Case-control design

This study design is based on a population of individuals with the disorder and a comparative control group. A good control group should be matched evenly for age, sex, race, and ideally socio-economic status and anything else that might affect allele frequencies (e.g. hair/ eye colour, IQ, exposure to a given factor in the environment) . This

study design compares the frequency of putative susceptibility alleles in cases compared with controls. The null hypothesis is that if the alleles are not associated with disease then they should be present with equal frequency in both groups. A statistically significant departure from the null hypothesis in the cases is considered to indicate association between disease and allele.

Assuming adequate sample size the case-control design has better power to detect and association between a disease and putative disease-causing variant than other study designs such as trio-based samples (discussed below). The disadvantages of the study design include difficulties avoiding population stratification even with well-matched controls. Furthermore, studies involving children could be faced with the ethical dilemma of the appropriateness of obtaining blood samples from healthy children. Alternative control groups could be utilized, e.g other groups of sick children undergoing blood sampling, although this may also introduce bias.

2.2.4.2 The trio-based association study

This study design overcomes the problems of population stratification associated with case-control studies and eliminates the need for a well-matched control group. Here the transmissions of alleles from parents to affected offspring are compared. The null hypothesis is that parental alleles have an equal possibility of being transmitted, if a particular allele is transmitted more frequently with disease than would be expected by chance then this allele is a putative susceptibility allele. The statistics used to test for significant transmission are the Transmission Disequilibrium Test (TDT) (Spielman et al. 1993) and the Haplotype Based Haplotype Relative Risk (Falk and Rubinstein 1987).

The main disadvantages of the design are threefold. It necessitates one third extra genotyping, as DNA extraction and genotyping have to be performed in both parents. The TDT calculation can only be performed in heterozygotic parents and thus only a proportion of the sample can be utilised with a consequent loss of power. HHRR does not rely on heterozygotes, however is rarely more informative than the TDT. Finally, the presence of

both parents are normally required for this design which means that its application is somewhat limited to conditions presenting in early life.

2.2.5 Markers used in LD mapping.

Traditionally restriction fragment length polymorphisms (RFLPs) markers, have been used to perform association analysis. Such markers have the advantage of being widely available in the public databases with published amplification parameters and many with available frequency data (e.g. http://www.gdb.org). However, there still exist gaps in the genome where there is an absence polymorphic microsatellite markers and necessitates the use of SNPs to fill in these gaps for the purposes of dense association mapping.

There are two publicly available SNP databases, the SNP database (dbSNP) and The SNP Consortium (TSC). It is estimated that there are 2.84 million SNPs of which 1.65 million are considered non-redundant SNPs (i.e. minor allele frequency is sufficient for the purposes of genetic association testing). In addition, it is estimated that 80% of the publicly available SNPs are polymorphic and 50% are common SNPs (Marth et al. 2001).

In this study, microsatellite markers were used initially for exploratory studies across a region of interest, followed by a SNP-based association study to further narrow a region showing evidence of association.

2.2.6 Power

Calculation of power in complex disease is difficult and relies on several basic assumptions about the underlying genetic model. Power calculations are used to estimate the approximate size of sample required to detect linkage or association. The two main assumptions made are the frequency of the disease causing allele and the mode of inheritance. Kaplan et al. 1997 derived formulas for investigating sample sizes necessary to obtain a specified power and type I error rate. The calculations use dominant and recessive models for a variety of disease allele frequencies and penetrances. The estimates assume that only one locus contributes to the phenotype so the attributable risk for that locus is 1.0. In the presence of heterogeneity, required sample sizes increase. For a recessive model, the

required sample sizes are less than under the dominant model. This is because both parents can potentially contribute to the overall statistic (the allele must be received from both parents to be affected). Under the dominant model, the transmission need only occur from one parent. As the mode of inheritance is unknown in most complex disorders direct estimates of power cannot be performed. Similarly the size of the gene effect must be known in order to determine the size of the population required to detect association. For genes of minor effect, larger sample sizes will be required. The model frequently used for genetic inheritance in autism is one of either oligo- or polygenic inheritance associated with epistasis (Pickles et al. 1995; Risch et al. 1999). Power calculations are discussed in sections 5.2.4 and 6.4.11.

Chapter 3

Materials and methods

3.1 Introduction

Both clinical and laboratory methods are presented in this chapter. In addition, as management was an important component of this work and was of vital importance to the investigations conducted on chromosome 2q (Chapter 6) there follows a brief discussion of research collaborations that were established during the research.

3.2 Research collaborations

3.2.1 Characterisation and fine-mapping of a susceptibility region for autism on chromosome 2q

This collaboration is central to the work reported in Chapter 6. The collaboration involved Dr. Sean Ennis, Lecturer in Medical Genetics, UCD, and to a lesser extent, Prof Andrew Green, Professor of Medical Genetics, UCD, Dr. Raymond Stallings and Dr. David Barton all based at the National Centre for Medical Genetics, Our Lady's Hospital, Crumlin.

3.2.2 Functional and genetic aspects of the dopamine and serotonin systems in autism

This collaboration was established with Dr. George Anderson, Developmental Neurochemist at the Child Study Centre, Yale University. Dr. Anderson's group are involved in the biochemical analysis of dopamine and serotonin function in our clinical sample while a PhD student in the Irish Autism Genetics Group has been investigating and analyzing relevant genotypic data with respect to the biochemical data

3.2.3 National Alliance for Autism Research (NAAR)- Autism Genetic Programme (AGP):

The Irish autism genetics group established as a result of the research presented here has been invited to participate in a large international multi-centre study in autism genetics. This is a two-phase collaboration. The first phase includes nearly all groups that have published genomewide linkage studies in autism. During this phase all the families

included in these studies will be genotyped using a uniform set of genetic markers and linkage analysis will be performed. The second phase will follow up the findings from linkage analysis. Up to 2000 trios or cases and controls will be utilised for linkage disequilibrium fine mapping of the regions with the strongest evidence for linkage. The Irish sample will be included in the second phase. In addition cell-lines from all families involved in this major collaboration will be placed in a repository for use by autism researchers worldwide.

3.2.4 Irish-Portuguese autism genetics collaboration

This is a recently established collaboration with Dr. Astrid Vicente at the Gulbenkian Institute in Lisbon. Dr. Vicente and her group have a sample of 250 autistic trios. A proportion of these were recruited from an inbred population in the Azores. The collaboration will involve mutual replication of results commencing with the replication by Dr. Vicente's group of the findings on the chromosome 2q region presented here.

3.3 Clinical methodology

3.3.1 Ethical approval

Ethical approval was sought and obtained from the Eastern Regional Health Authority Child and Adolescent Psychiatry Ethics Committee (Formerly Eastern Health Board). Consent was obtained to participate in the study and to the blood sampling procedure. Additional consent was obtained for the audio-/ videotaping of interviews of assessments. Consent was obtained from both parents to participate. With the recommendation of the Ethics committee, consent was also obtained from subject where they were considered competent to give consent. The judgement of competency was given to the parents and caregivers of the individuals.

3.3.2 Ascertainment of a population sample of trios

Ascertainment was conducted through family support groups, special schools/ centres and special classes within the Autism Outreach Programme, part of Beech Park Autism Services, based in the South-Western Area Health Board of the Eastern Regional Health Authority. Cases were also ascertained through physician referral, mainly by Professor

Michael Fitzgerald, Henry Marsh Professor of Child and Adolescent Psychiatry, who obtained consent from parents of affected children to be contacted.

Information sheets regarding the study were distributed to parents through the relevant agencies. A copy of the information sheet is included in Appendix A. The information sheet included a detachable portion for parents to fill in their contact details and indicate consent to be contacted regarding the project. On receipt of this consent, parents were contacted by telephone and a screening interview was conducted to assess if the affected individual was likely to be suitable for inclusion in the study. A copy of this interview is included as an appendix (Appendix B). Where level of intellectual functioning was unknown, written consent was obtained from parents to obtain details of psychological assessment to confirm that the subject had an IQ greater than 35 or a mental age score greater than 18 months. At this point, suitable subjects were recruited to the study and an appointment was arranged to visit the family at home to conduct the ADI-R interview.

3.3.3 Recruitment procedures

3.3.3.1 Initial contact

At initial telephone contact, participants were offered the choice to come in to the clinical centre or to be interviewed at home. In two cases participants opted to come in to the centre for convenience, the remainder chose to be interviewed at home. At the first appointment the Autism Diagnostic Interview – Revised (ADI-R) assessment (see section 3.3.5.1) was conducted and blood was obtained from one or both parents. Signed consent was obtained from the parents to contact relevant professionals for further details regarding their child, e.g., psychologists, physicians, schools, fragile X/ cytogenetic test reports.

The ADI diagnostic algorithm was completed following the interview with parents. Individuals who did not meet the cut-off for autism on the ADI-R interview were excluded from further follow-up. Subjects who fell no more than one point under the cut-off for autism on the algorithm were followed up with the Autism Diagnostic Observation Schedule - Generic (ADOS-G) assessment. A study number was allocated to recruited

subjects at this point. The study number took the form of a three-digit number followed by M, F or C, to indicate mother, father or child.

3.3.3.2 Further contact.

Following initial contact, arrangements were made to conduct the ADOS-G assessment and to collect remaining data. In the majority of cases this was conducted (with parental consent) in the school/ centre individuals were attending. In a minority of cases parents were reluctant to have school-based assessments. These children were assessed in the research centre or in the family home. Normally on the same day, the clinical research team (L.G. and G.K) called to the family home again to complete the data collection. On this occasion blood/ buccal samples were obtained from subjects and remaining parents. Measurements were obtained for height and head circumference. A physical examination was conducted of the subject that included a brief neurological assessment and skin examination using a Wood's lamp to exclude Tuberous Sclerosis and Neurofibromatosis.

3.3.4 Collection of samples for DNA

Blood samples or buccal swabs were obtained for DNA. Three six millilitre samples of blood were obtained from each parent using 18 gauge needles in Vacuette® containers (Greiner BioOne), two EDTA and one serum tube. Samples were obtained from the affected offspring using 22 gauge butterfly needles and Vacuettes®. An additional two samples were obtained from children for Fragile X testing and chromosomal analysis and these samples were submitted to the National Centre for Medical Genetics. Blood samples for DNA were stored in Vacuettes® containing EDTA at -70°C until they were processed. Serum samples were spun at 6, 000rpm at room temperature and the supernatant was stored at -70°C.

Where blood samples could not be taken, buccal swabs were obtained using sterile cytology brushes (CytoSoft, Medical Packaging Corporation). The inside of the cheek and gums were gently scraped using the brushes following mouth rinsing with water. The swabs were reinserted into sterile, labelled packaging and delivered directly to the laboratory for extraction within one to two weeks. Where buccal swabs were obtained from

subjects it was not possible to test for Fragile X or chromosomal anomalies. If these tests had been previously conducted every effort was made to obtain the test results.

3.3.5 Clinical diagnostic instruments and procedures

3.3.5.1 The Autism Diagnostic Interview – Revised (ADI-R)

The ADI-R (Le Couteur et al. 1989) is a semi-structured investigator-based interview devised for use with an individual's primary caregiver. It provides a lifetime assessment of the range of behaviours relevant to the differential diagnosis of pervasive developmental disorders in individuals of any chronological age from 5 years to early adulthood and with any mental age from 2 years upwards. It was derived from an interview originally devised by Folstein and Rutter (1977) but was developed to offer greater specification and operationalisation. It has been shown to yield highly reliable ratings that effectively discriminate between autistic and non-autistic intellectually disabled individuals (Le Couteur et al. 1989).

It focuses on three main areas of abnormality:

(1) The qualities of reciprocal social interaction.

(2) Communication and language.

(3) Repetitive, restricted and stereotyped behaviour.

These represent key diagnostic features characterised in ICD-10 and DSM-IV. The main purpose is to obtain detailed descriptions of those behaviours that are essential for the differential diagnosis of pervasive developmental disorders. In addition the questionnaire also focuses on a variety of other behaviours that occur in association with pervasive developmental disorders and details regarding early developmental milestones are obtained.

3.3.5.1.1 Interview style and format

The questionnaire may be administered to one or both parents or primary caregivers. It normally takes three and a half to four hours to complete. It consists of 111 items with questions concerning aspects of behaviour as outlined above. The questions are semi-structured in that each question relies on specific probes that must be asked, followed by

suggested probes that may be asked so that the interviewer can code for that item. As the interview is an investigator-based instrument, the interviewer is required to be familiar with the codings required for each of the behavioural items and with the conceptual distinctions involved in each item. In addition the interviewer must ensure that sufficient information is obtained in order to code the item. Thus detailed descriptions of behaviour are required and it is necessary to obtain specific examples of behaviour by descriptions of specific incidents or events. The temporal occurrences of the behaviour are ascertained by asking the interviewee to relate them to specific events, for example commencement of nursery or primary school or around the time of the second or third birthdays. Because of the highly structured nature of the interview, training is essential in the administration and coding of the interview and reliability in coding must be achieved with the centre where training is received.

The interview is comprised of six sections. The first part provides general background information regarding the subject and their family. This information is intended for orientation purposes around the subsequent questioning. Section two is concerned with early developmental history and milestones. The subsequent three sections are concerned with the areas related to autism, i.e., communication and language, social development and play, and unusual interests and repetitive behaviours. The questioning in this section focuses on the behaviour currently and in the earlier years. The seventh section concerns non-specific behaviour difficulties, special abilities and some closing questions.

3.3.5.1.2 Coding

Coding of items depending on the presence or absence of the abnormalities in question are performed as follows:

0: Behaviour of the type specified in the coding is/ was not present

1: Behaviour of the type specified is/ was present in an abnormal form (or lack of behaviour was present), but not sufficiently severe, frequent or marked to meet the criteria for '2'.

2: Definite abnormality of the type specified that meets/ met the criteria given for that coding.

3: A more severe manifestation of '2'

7: Definite abnormality in the general area of the coding, but not of the type specified.

8: Not applicable (no opportunity to exhibit the behaviour because outside the relevant age range, does not have the required level of behaviour or because never in the circumstances that could elicit the behaviour).

9: Not known.

The interviewer codes each item during administration. Where behaviours show differing aspects under different items these can be coded separately. However the same behaviour should not be coded twice. For example, where a child that has a specific interest in trains and talks about this to the exclusion of other topics the behaviour can be coded under item 70 (circumscribed interests). However if the child also prefers to line up his toy trains rather than playing with them functionally, this behaviour can be coded under item 72 (repetitive use of objects).

Developmental items are coded as age in months. If an age-range is given, the value coded is the mid-point of the range rounded up to the nearest month. When no date can be obtained the following codings are used:

993: Regression – milestone achieved, but subject then relapsed.

994: Milestone never achieved.

995: Milestone still not reached.

996: Not known, apparently normal.

997: Not known, apparently delayed.

998: Not applicable.

999: Not known/ not asked.

All behavioural descriptions are recorded while the interviewer is administering the interview. This is done in sufficient detail so that a second person would be capable of coding the items based on what has been recorded.

3.3.5.1.3 ADI-R diagnostic algorithm

In order to produce an operationalised diagnosis, a diagnostic algorithm was derived from ICD-10 and DSM-IV criteria using the items scored in the interview. The algorithm is divided into four domains:

(1) Reciprocal Social Interaction.

(2) Language/ Communication.

(3) Restricted, repetitive and stereotyped behaviours.

(4) Abnormalities evident before the age of 36 months.

Item scores are summed for each domain and a cut-off is applied to each section for an autism diagnosis. Subjects must be above the cut-off on all four domains to meet the ADI-R criteria for autism. The cut-offs for each of the domains are 10, 8, 3 and 1 respectively for verbal individuals. The cut-offs are the same for non-verbal individuals with the exception of the communication domain where it is lowered to 7, as fewer items are applicable to this group.

3.3.5.2 Autism Diagnostic Observation Schedule –Generic (ADOS-G)

The ADOS-G (DiLavore et al. 1995) is a semi-structured assessment of communication, social interaction and play/ imaginative use of materials of individuals with a potential diagnosis of autism or pervasive developmental disorder (PDD). It was first introduced in the 1980s as a method of standardising direct observations of the abnormalities seen in autism/ PDD. It was developed at the University of Chicago Developmental Disabilities Clinic and is now published by Western Psychological Services. The instrument was derived from two earlier instruments, the ADOS (Lord et al. 1989), for children and adults with an expressive language level above 3 years, and the PL-ADOS (Pre-linguistic ADOS) (DiLavore et al. 1995) for non-verbal children. It was subsequently developed into the ADOS-Generic, consisting of four modules that are intended for children and adults with differing levels of expressive language and chronological age.

3.3.5.2.1 ADOS-G Format

It consists of standard activities that allow the examiner to assess the presence or absence of behaviour associated with autism / autism spectrum disorders. Planned social occasions ('presses') are created in which behaviour of a particular type is likely to occur. Structured activities and other less structured interactions using a variety of materials provide standard contexts in which social, communicative and other behaviours may occur.

The examiner selects the module most appropriate for the subject's expressive language level and age. As with the ADI, notes are kept throughout the assessment, recording the behaviours observed. Assessments take between 30-45 minutes to complete and are rated immediately after the administration. Assessments are normally videotaped to aid reliability exercises.

Module 1 is for children who don't use phrase speech (defined as non-echoed three word utterances that sometimes involve a verb and that are the child's meaningful and spontaneous utterances). Module 2 is for individuals with phrase speech who are not verbally fluent. Modules 1 and 2 are generally not appropriate for adolescents or adults as the materials presented would not be age-appropriate. Module 3 is intended for verbally fluent children for whom playing with toys is age-appropriate (i.e., less than 12-16 years old). Module 4 is intended for verbally fluent adolescents and adults. It contains socio-emotional questioning which is included as part of Module 3. However in Module 4 the information is obtained through an interviewing style rather than through play. (Verbal fluency is defined as the expressive language skills of a typical four-year-old: producing a range of sentence types and grammatical forms, using language to provide information about events out of context and producing some logical connections with sentences. Grammatical errors may occur).

Only one module of the ADOS-G is normally administered at any one time, however it is possible to shift from one module to another if necessary, for example if the language ability is different from that expected or if the tasks appear inappropriate.

3.3.5.2.2 ADOS-G coding and diagnostic algorithm.

Coding of the items on the ADOS-G is performed immediately after the assessment is administered. Codings range from '0' (no evidence of the abnormality) to '2' (definite evidence of the abnormality). Some items contain a '3' code to indicate severe abnormalities.

The codings applied to the items at the end of each module are used to derive a diagnostic algorithm based on ICD10 and DSM-IV criteria for autism. The algorithm is divided into 4 domains:

(1) Communication.
(2) Qualitative Impairments in Social Interaction.
(3) Play (Module 1) or Creativity (Modules 2-4).
(4) Restricted, Repetitive Behaviours and Interests.

Items were included in the algorithm based on reliability and validity studies when the instrument was developed. Scores for each domain are entered from the ratings for the individual item scores. For the purposes of the algorithm, scores of 3 are converted to 2. ADOS-G diagnoses are based on the scores for the Communication-Social domains, i.e. (1) and (2) in the above list. Diagnoses are three tiered. This means that to achieve an ADOS-G diagnosis for autism the subject must score above a cut-off for the Communication and Social domains individually and for the combined Communication and Social score. Cut-offs are determined for autism and autism spectrum disorders and vary depending on the module used.

?

3.3.5.3 Training and reliability in the ADI-R and ADOS-G instruments

As described, training is required in the ADI-R and ADOS-G instruments. For the purposes of the sample collection described here, two researchers were involved in the ascertainment and recruitment of the subjects to the study. These were L.G. (author) and G.K. (research nurse). L.G. spent a three and a half month period training and achieving reliability in the ADI-R and ADOS-G instruments at the MRC Child Psychiatry Unit, Institute of Psychiatry, London, under the supervision of Dr. Tony Bailey. G.K. was trained using

ADI-R training tapes under the supervision of L.G. by arrangement with Dr. Tony Bailey. G.K. attended a training course in the ADOS-G instrument at the Institute of Psychiatry and achieved reliability with L.G.

3.3.5.4 Limitations of the ADI-R and ADOS-G assessments

Both tools require a good clinical understanding of the behaviours being measured and therefore are best suited to those with clinical experience of childhood disorders, e.g., psychology, psychiatry, speech and language therapy. Training is required to establish reliability and the instruments should not be used until this has been achieved. Thus there is often a time lag when recruiting new clinical personnel to ensure that they are trained in the use of the instrument. Training courses run infrequently and this can serve as a further delay. In addition, reliability exercises for both instruments must be undertaken periodically. The authors recommend every 2-3 months within a research group. While this ensures the reliability of the instruments, it can also be time-consuming.

3.3.5.4.1 ADI-R limitations

(1) The ADI-R is a lengthy tool, taking approximately 3.5 to 4 hours to apply in the research format. As the interviewer must get specific examples of behaviour in order to code items, the length of administration can extend far beyond this, particularly in the case of a garrulous informant. This has obvious limitations for research purposes, particularly with respect to slowing down the rate of recruitment of individuals.

(2) The specificity of the tool in the profoundly intellectually disabled is poor and therefore it is not suitable for use in those with an IQ score < 30. The definition of autism in this population is particularly difficult as it is not unusual to find autistic symptoms in these individuals despite the absence of classic autistic syndrome (Wing and Gould 1979).

(3) The diagnostic algorithm was based on the criteria for ICD-10. However several items that were not included were shown to be discriminating for autism. None of these showed sufficient reliability to be included in the algorithm, however it begs the question as to whether the ICD-10 criteria are too narrow.

(4) The ADI-R was developed for use with children in the fifth year and is not suitable for use in pre-school children.

(5) As it is solely informant based the ADOS-G must also be performed to establish a research autism diagnosis. This lengthens recruitment process.

3.3.5.4.2 ADOS-G limitations.

(1) The ADOS-G must be supplemented by a diagnostic interview for valid diagnosis.

(2) No module is suitable for the assessment of low functioning adults as the materials provided in module 1 are not age-appropriate.

(3) The ADOS-G provides diagnoses for autism and autism spectrum disorders only. It does not differentiate between individuals on the autism spectrum with Asperger syndrome or PDD-NOS for example.

3.3.5.5 The Vineland Adaptive Behaviour Scales – Interview Edition (VABS)

The VABS (Sparrow S.S 1984) is a revised edition of the Vineland Social Maturity Scale, designed to measure personal and social sufficiency in individuals from birth to adulthood. It is a semi-structured interview to a parent or caregiver familiar with the individual's typical day-to-day behaviour. It contains 297 items and provides a general assessment of adaptive behaviour, which is useful for determining areas of strength and weakness. It can be used with individuals from birth to 18 years, or with low-functioning adults. The domains, sub-domains, and item content covered by the scales are:

(1) Communication: Receptive, Expressive, Written

(2) Daily Living Skills: Personal, Domestic, Community

(3) Socialisation: Interpersonal Relationships, Play and Leisure Time, Coping Skills

(4) Motor Skills: Gross, Fine

Administration takes between 20-60 minutes.

3.3.6 Inclusion/ Exclusion Criteria

Inclusion to the study depended on confirmation of the diagnosis of autism using the ADI-R and the presence of autism/ autism spectrum disorder with the ADOS-G. Reasons for exclusion from the study were known medical causes of autism (as outlined in Chapter 1), the presence of chromosomal abnormalities and the Fragile X anomaly and an IQ less than 35 or a mental age score less than 18 months. To this end subjects were screened for cytogenetic abnormalities and the Fragile X anomaly. IQ scores were determined from medical/ psychology case notes where available or if not Mental Age scores < 18 months were determined by the Vineland Adaptive Behaviour Scales. A physical examination of subjects was undertaken that involved a brief physical examination and skin examination using an ultra-violet Wood's lamp to exclude Tuberous Sclerosis or Neurofibromatosis.

3.3.7 Consensus diagnosis:

A consensus diagnosis for each individual was determined between L.G. and G.K. based on all the clinical data collected and a final decision was made regarding inclusion of each subject in the study.

3.4 Laboratory solutions and procedures

Consumables were obtained from Biolabs, Sigma Chemicals, Gibco BRL Life Technologies, Pharmacia Ltd., MWG, and Applied Biosciences.

3.4.1 Buffers and Solutions

(1) TE Buffer:

2M Tris-Hcl (pH 7.5)

0.5M EDTA (pH 8.0)

(2) Extraction Buffers/ Solutions:

Lysis Buffer (x10)

50mM Tris-HCl, pH 7.5

25mM $MgCl_2$

0.6M sucrose

5% (v/v) Triton x100

Distilled H_20

Suspension Buffer (x10)

2M Tris-Hcl (pH 7.5)

0.5M EDTA (pH 5.0)

5M NaCl

(3) Electrophoresis Buffers:

10x TAE:

0.4M Trisma Base

0.5M EDTA

The pH of the buffer was brought to 8.2 with glacial acetic acid.

TBE (x10) (Invitrogen):

0.9M Boric Acid

0.01M EDTA

(4) PCR Buffers:

10X PCR Buffer (Invitrogen Platinum Taq):

200mM Tris-HCl (pH 8.4)

500mM Kcl

(5) Loading Buffers:

0.25% (w/v) bromophenol blue

15% (w/v) ficoll

Hi-Di Formamide Solution

3.4.2 Laboratory methods

3.4.2.1 DNA extraction from blood samples

Extraction of DNA from blood samples was conducted using a two-day Phenol/ Chloroform method. On day one, 6ml blood samples were removed from the freezer and allowed to thaw slowly on ice for a two- four hour period.

Day one involved the following steps:

(1) Blood samples were transferred to a labeled 50ml Falcon tube. Blood tubes were rinsed with small amount of sterile distilled water which was added to the Falcon tube. Sterile distilled water was added to 12.5ml.

(2) Lysis buffer (x1) was added to 25 ml. The tubes were inverted and placed on ice on a rotary shaker for 30 minutes.

(Lysis buffer is a salt solution that is slightly osmolar to the contents of the red blood cells (RBC). The RBCs absorb absorb water from the solution by osmosis, swell and rupture while the white cells remain intact.)

(3) The tubes were centrifuged at 3, 500 rpm for 15 minutes.

(4) The supernatant was poured off to approximately 4ml.

(5) Lysis buffer was then added again to 25mls. The pellet was dislodged into the solution by banging end of tube.

(6) The tubes were on ice on a rotary shaker for 10 minutes and subsequently centrifuged at 3, 500rpm for 15 minutes.

(7) The supernatant was poured off until the white cell layer on the surface of the pellets just began to move.

(8) The pellets were suspended in 1ml of suspension buffer (x1) and the contents were transferred to a labeled 15ml Falcon Tube. Each 50ml tube was rinsed with a further 1ml of suspension buffer and this was added to the corresponding 15ml tube.

(9) One hundred and fifty microlitres of 10% Sodium Dodecyl Sulphate (SDS) and 60µl of Proteinase K was added to each.

(SDS is negatively charged with a high affinity to protein. Thus it encourages proteolysis. Proteinase K is a powerful proteolytis enzyme that ensures the degradation of nucleoproteins)

(10) The tubes were placed on a rotary shaker overnight at room temperature.

Day 2 of the extraction process involves the use of Phenol and Chloroform/ Isoamylalcohol (24/1, v/v). As Phenol is neurotoxic, the Phenol/ Chloroform steps were conducted in a fume hood.

Day two involved the following steps:

Phenol/ Chloroform steps:

(1) The tubes were removed from the rotary shaker and 2ml of Phenol was added to each and thoroughly mixed.

(2) The tubes were placed on a rotary shaker for 10 minutes and then centrifuged at 6, 000rpm for 10 minutes.

(3) The aqueous white layer was removed into a clean labelled 15ml tube and 1ml of Phenol and 1ml of Chloroform was added to each. The tubes were again placed on the rocker for 10 minutes and centrifuged at 6, 000rpm for 10 minutes.

(4) The top layer was removed into a clean 15ml tube and 2ml of Choloroform/ IAA was added to each tube. The tubes were rocked for 10 minutes and centrifuged at 6, 000rpm and the supernatant was removed into a clean 15ml tube.

(5) Fifty micro litres of 3M Sodium Acetate was added followed by Absolute Alcohol at a temperature of -20°C to a volume of 6mls.

(6) The tubes were gently inverted until the formation of a white fluffy DNA pellet was observed.

(7) The pellet was removed from the tube using a sterile 1ml pipette tip and transferred to a labelled 1.5ml sterile Eppendorf.

(8) One millilitre of 70% ethanol (-20°C) was added to remove excess salt and the samples were microfuged at 13, 000rpm for 5 minutes.

(9) Ethanol was carefully removed with a pipette tip.

(10) A further 1ml of 70% ethanol (-20°C) was added and the tubes were microfuged at 13, 000rpm for 5 minutes.

(11) Ethanol was removed again.

(12) The DNA was dried overnight at room temperature.

3.4.2.2 DNA suspension.

The dried DNA pellet was suspended in 250µl of TE Buffer and stored in the fridge in 1ml Eppendorf tubes at 4°C for four to five days until the pellet had resuspended. The tubes were mixed daily by flicking the tubes to ensure the DNA was fully dissolved. This stock DNA was stored at -20°C.

3.4.2.3 Buccal cell extraction.

DNA from buccal cells collected with cytology brushes were extracted within one week of their receipt in the lab. The method used to extract DNA used BIORAD Instagene Matrix according to the following protocol.

(1) Two brushes from the same individual were cut into a labeled sterile 1.5ml Eppendorf. (Four to six brushes were obtained per subject, thus each subject had 2-3 eppendorfs).

(2) Five hundred microlitres of 50mM NaOH was added to each eppendorf to dentaure the proteins.

(3) The tubes were capped and vortexed by hand for 60 seconds.

(4) The tubes were then placed in a heatblock at 95°C for 10 minutes.

(5) Fifty microlitres of 1M Tris pH 8.0 was added to each tube.

(6) The tubes were capped and vortexed for 30 seconds.

(7) The BIORAD Instagene matrix was mixed well by hand prior to use and 200µl was added to each tube.

(8) The tubes were placed in a water bath at 56°C for 30 minutes.

(9) They were then vortexed for 10 seconds and placed in the heatblock again at 100°C for 8 minutes.

(10) They were again vortexed for 10 seconds and the brushes were removed using a tweezers.

(11) Finally the tubes were microfuged at 12, 000 rpm for 3 minutes. The supernatant was transferred to new tubes and used for PCR.

The samples were stored in a fridge at 4°C.

3.4.2.4 DNA quantification

Rough estimates of DNA concentration were performed with the extracted DNA samples to confirm that the extraction procedure was successful. DNA samples (5µl) were electrophoresed on 1% agarose gels. The gel was subsequently stained with ethidium bromide and visualized under an ultra-violet (UV) lamp. The bands were compared with fluorescence of known concentrations of molecular weight standards such as λHindIII.

DNA quantification of all samples was later performed using spectroscopy. For each sample a solution of 5ml of DNA and 1ml of TE buffer was made. Each sample was quantified in a Spectrometer by comparing the optical density (OD) of the DNA/ TE solution to the OD of 1ml TE Buffer. Readings for the OD of the solutions were taken at 260nm and 280nm. The 260nm reading indicated the amount of DNA in the solution and the 280nm reading indicated the quantity of RNA in the sample. Quantification of DNA in nanograms per microliters was calculated. For 1ml of DNA per 1ml of TE Buffer, an OD of 1 is equal to 50ng of DNA per microliter. The OD was calculated using the following formula:

If the OD260nm = 1 the DNA Concentration = 50ng/ ml

Then for OD260nm= X the DNA Concentration = {(50 x X)/ 5} ng/ml

The ratio of OD260nm:OD280nm was calculated as an indication of the impurities in the DNA solution.

3.4.2.5 Working solutions of DNA

DNA concentrations varied between 0.4μg/μl and 0.8μg/μl. Working solutions for each sample was made containing 15μl of DNA stock solution and 235μl of TE Buffer. Stock solutions were frozen at -20°C and thawed intermittently as required for the purposes of preparing more working solutions. Working solutions were stored in a refrigerator at 4°C.

3.4.2.6 DNA amplification by polymerase chain reaction (PCR)

Standard primer extension reactions were conducted for the amplification of genetic markers in the HOXA1/B1 genes and the Microsatellite markers used in a genetic association study on chromosome 2q. As the volumes and conditions for these markers were variable they are considered separately below. SNP genotyping was performed by the Human Genome Mapping Project, MRC geneservice using Taqman Assays on a commercial basis.

3.4.2.6.1 Genotyping HOXA1/B1.

PCR reactions for HOXA1 and HOXB1 were carried out in reaction volumes of 25μl containing 60ng of genomic DNA, 20pmol of each primer, 200μm of each dNTP, 50 mM of KCl, 10mM of Tris HCl (pH 9), 1.5mM of MgCl2 (HOXA1)/ 2.0mM of MgCl2 (HOXB1), 0.01% gelatine and 1U of Taq polymerase. PCR reactions were conducted in a PTC-200 Peltier Thermocycler (MJ Research) programmed for a three-stage reaction.

The primer sequences for HOXA1/B1were as follows:

HOXA1

(Forward) 5'-GCA AGA ATG AAC TCC TTC CTG – 3'

(Reverse) 5'-ACC AAC CAG CAG GAC TGA CCT – 3'

HOXB1

(Forward) 5' – GCA TGG ACT ATA ATA GGA TG – 3'

(Reverse) 5' – TCT TGG GTG GGT TTC TCT TTA – 3'

PCR parameters for HOXA1 were:

Denaturing Step:	94°C x 3.5mins
Amplification Step:	94°C X 35s
	64°C X 45s
	72°C X 35s
Final Extension Step:	72°C X 5 mins

This produced a fragment size of 661bp

PCR parameters for HOXB1 were:

Denaturing Step:	94°C x 3.5mins
Amplification Step:	94°C X 45s
	64°C X 45s
	72°C X 45s
Final Extension Step:	72°C X 5 mins

This produced a fragment size of 576bp.

PCR products for HOXA1 and HOXB1 were digested overnight with HphI and MspI restriction enzymes respectively, as recommended by the manufacturers.

3.4.2.6.2 DNA detection with polyacrylamide / agarose gels and ethidium bromide staining (HOXA1/HOXB1)

Twenty microlitres of digestion product and 3µl were electrophoresed on non-denaturing 15% polyacrylamide gels. Electrophoresis was conducted at room temperature for six hours. DNA fragments were detected by staining the polyacrylamide gels with 80µl of ethidium bromide at a final concentration of 1µg/ml for 20 minutes. The gel was then washed with distilled water for 10 minutes to remove excess ethidium. The fragments were visualized under a UV transilluminator. Allele size was estimated by comparison of the DNA fragments with λHINDIII. Agarose gels were also stained with ethidium bromide to the same concentration and fragments were visualized in the same way.

Digestion fragments for the HOXA1 and HOXB1 genes were as follows:

HOXA1:	A allele –	210/ 199/ 198bp
	G allele –	409bp
HOXB1:	no insertion -	134/ 200bp
	9bp insertion –	143/ 200bp

3.4.2.6.3 Chromosome 2 microsatellite markers

Microsatellite markers selected from publicly available databases were used for association mapping of a region on chromosome 2q (http://genome.ucsc.edu/cgi-bin/hgGateway, http://gdbwww.dkfz-heidelberg.de/, http://www.ensembl.org). The construction of this association map is discussed in more detail in Chapter 6 (section 6.3.1). Details of the markers used are given in Table 3.1.

Standard PCR reactions in volumes of 10µl were conducted for each marker on PTC-200 Peltier Thermocyclers (MJ Research) in 96-Well PCR Plates. PCR mixes consisted of 24ng of genomic DNA, 8pmol of each primer, 80µmol of each dNTP, 1.5-2.5mM $MgCl_2$, 1 U Platinum Taq Polymerase® and 1x PCR Buffer supplied by the manufaturer. Primers were 5'- labelled with fluorescent labels (FAM-6, HEX, TET) for the purposes of Genotyping on the Applied Biosystems 3100 Genetic Analyzer. PCR conditions are outlined in Table 3.1. Conditions were adjusted for some markers particularly for amplification of DNA obtained from buccal swabs. The samples were run in PCR tubes rather than plates. The $MgCl_2$ concentration was increased to 2.0mM for markers D2S426 and D2S270. Marker D2S1787 did not amplify for buccal DNA despite multiple attempts at optimization. A sample of PCR products for individual markers were electrophoresed on 2% Agarose Gels for 45 minutes to check PCR quality. Gels were stained with ethidium bromide and examined under the UV transilluminator to examine the quality of the PCR product. PCR reactions yielding no/ poor products of amplification were further optimized and tested again in the same manner. One marker D2S1391 failed to amplify despite multiple attempts at optimization and was excluded. As it was located 72kb from the flanking marker D2S1383 this did not make a marked difference to the association map.

3.4.2.6.4 Genotyping with Applied Biosystems 3100 Genetic Analyzer (ABI 3100)

Genotyping of these markers was performed on an Applied Biosystems 3100 Genetic Analyzer. This is a multi-colour fluorescence-based DNA analysis system using capillary electrophoresis. It is fully automated from sample loading to data analysis. PCR product was diluted prior to running on the ABI 3100. Optimal dilutions were first determined by running samples of varying dilutions for each individual marker. Differently labelled markers were then pooled. Some markers with the same label were also pooled where PCR fragment sizes varied.

Table 3.1 Microsatellite Markers Used in Association Mapping of Chromosome 2q Region

GDB ID	Alternative ID	Primer sequences	No. of Reported alleles	Hetrozygosity	Size Range bp	*PCR conditions
D2S2310	AFMb355xd5	-CGACTTGAGTAGACGCACTATTC- -GCATCTAAACTGTGAAATGAGC-	Not reported	0.79	244-260	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S2077	UniSTS:34364	-CATGGAGCCGACGTTCAG- -CATTTACTTATTTTCTGAATCACCC-	8	Unknown	149	94°C x 45s 58°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S364	AFM303ya9	-GCTAATAATCTCTATGGGAATGCAG- -AGGATTCTGACAGCAGCATAAC-	9	0.80	Min: 232	94°C x 45s 56°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
**SHGC 1659		-GCTAATAATCTCTATGGGAATGCAG- -GTTATGCTGCTGTCAGAATCCT-	Not reported	Unknown	189	94°C x 45s 55°C x 45s 72°C x 45s
D2S350	AFM292wd1 UniSTS:75558	-TGTGTCACAGCGAAATTAC- -TCCAAGCAACTCCTCATAC-	4	0.42	Min: 152	94°C x 45s 54°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S2273	AFMb297xc1 UniSTS:15570	-AGGTCACTGGACTTCCC- -AGATAATACTTGGTTCATGGC-	9	0.80	140-164	94°C x 45s 54°C x 45s 72°C x 45s $MgCl_2$ = 1.0mM

Table 3.1 contd.

GDB ID	Alternative ID	Primer sequences	No. of Reported alleles	Hetrozygosity	Size Range bp	*PCR conditions
D2S2281	AFMb310xf5 UniSTS:77352	-TTATTAGCAGTATAAAGGCAGCA- -GATTAGGAGGCACTCTGGG-	6	0.7721	215 – 229	94°C x 45s 58°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S2366	AFMa057vg9 UniSTS:3483	-AACTTCAGCATTCTAAGAGACCTTT- -CGTGCCCAGCAGTGAT-	7	0.7503	Min: 172	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S1383	CHLC.GATA45A07 UniSTS:11697	-AGACAGTTGTAGGTGTGCAGC- -GCTGGAGGCATCACATTATT-	Not reported	Unknown	Min: 202	94°C x 45s 60°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S1391	CHLC.GATA65C03 UniSTS:36563	-CTCACTGTCTGGATTTCTTGG- -TAAGGACAAGTTAAAAAAGCTGG-	Not reported	Unknown	Min: 124	Failed to amplify and excluded
D2S1787	CHLC.GATA85E09 UniSTS:6142	-AATGCTAAATTTCCTGGCCT- -GATCTCGTGATCCGCCCT-	Not reported	Unknown	Min: 276	94°C x 45s 58°C x 45s 72°C x 45s $MgCl_2$ = 1.0mM
D2S1361	CHLC.GATA14E05 UniSTS:60630	-TAAGAAGCCGTTCTTGGATG- -TCAAATTCAAGTTAACATTCATCA-	8	Unknown	Min: 177	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S2069	WI-5296 UniSTS:38497	-CTGCGTTGAGACATGAAGGA- -ACTTTAATTTTTTTGCAAATTTTGG-	Not reported	Unknown	Min: 103	94°C x 45s 54°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM

Table 3.1 contd.

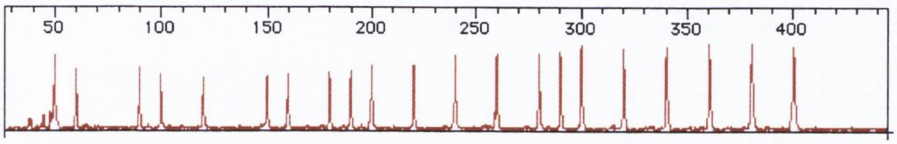
D2S270	UT866 UniSTS:50550	-GCTGTTGGGTGGAATATGTG- -TCACAAAGAGACAAAGAAGGA-	Not reported	0.75	Min: 369	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S152	AFM207xg1	-AGCTGCTGGTATATATTATCTTCCA- -TAATGAAATTATTTAATGCATCTCC- -GACCCAAACAATATCATTATGG- -GTGGACTATCTAGTTTGACCACA-	9 Not reported	0.7712 Unknown	269 – 285 Min: 191	94°C x 45s 57°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S103	UniSTS:5344	-TATTATTTCGAGGTCCACCT- -AATAAATATACTGTTGGAGCT-	9	0.82	109 – 125	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S426	CHLC.GATA12B05	-CACCACTCACTCACTGATGC- -TGAGACCCTGCCTCAAAATA-	4	Unknown	155	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM
D2S2262	AFMb082ye1	-AGCAATGATAAACCACATCT- -AAGCCCCTCGTTTTCT-	6	0.6526	224 – 244	94°C x 45s 55°C x 45s 72°C x 45s $MgCl_2$ = 1.5mM

* All PCR reactions had an initial denaturation step of 96°C x 5 mins/ final extension step of 72°C x 10 mins

**This Marker was optimized using HotStar® Taq (Qiagen)

For each sample, 1µl of the pooled dilutions was mixed with 0.5µl of ROX400 Size Standard and 10µl of Hi-Di® Formamide Solution. ROX 400 has peaks at 50bp, 60bp, 90bp, 100bp, 120bp, 150bp, 160bp, 180bp, 190bp, 200bp, 220bp, 240bp, 260bp, 280bp 290bp 300bp, 320bp, 340bp, 360bp, 380bp, and 400 bp. This is illustrated in Figure 3.1.

Figure 3. 1 ROX 400 Size Standard Peaks



Genotyping was performed using the ABI 3100 data collection software and Genemapper v.2.7. Genotypes were scored by two individuals. Individual and family ID numbers were not available when genotyping was conducted to ensure that affectation status was unknown. At this point samples that had failed to amplify were identified and were re-amplified under optimized PCR conditions and genotyped again in the same manner.

3.1.13.6.5 Chromosome 2 single nucleotide polymorphisms

A smaller region of approximately 868kb was identified for follow up association mapping based on the results of the microsatellite testing. The construction of this map is discussed in more detail in Chapter 6 (section 6.3.1) Twenty-six SNPs were selected from the publicly available databases (http://www.ncbi.nlm.nih.gov/SNP/, http://snp.cshl.org/). These SNPs were genotyped commercially by the MRC Geneservice, a division of the Medical Research Council funded Human Genome Mapping Project. SNP references and surrounding sequence information was provided to the MRC Geneservice for the purposes of assay design and genotyping. Genotyping was conducted using TaqMan Assays®. The raw data was returned to the author for the purposes of analysis.

3.4.2.7 Statistical methods

3.4.2.7.1 Allele frequencies and Hardy-Weinberg equilibrium

Allele frequencies for all markers were calculated as the proportion of occurrences of each allele in the sample. The Hardy Weinberg principle assumes that allele frequencies should conform to the following mathematical equation:

$$p^2 + 2pq + q^2 = 1$$

Where p^2 = the expected fraction of the population homozygous for p

q^2 = the expected fraction homozygous for q

2pq = the expected fraction of heterozygotes

Each marker was assessed for deviation from Hardy-Weinberg equilibrium by calculating the above formula based on the observed and expected allele frequencies for each marker.

3.4.2.7.2 Haplotype-based haplotype relative risk (HHRR)

This tests the transmission of transmitted compared to non-transmitted alleles from parents to affected offspring (Falk and Rubinstein 1987). The format is a simple 2x2 contingency table that generates a χ^2 statistic and p significance value. Calculations of HHRR were obtained from the Epistat program. This program generates an estimate of the relative risk.

3.4.2.7.3 Transmission disequilibrium test (TDT)

The TDT is a test of association in the presence of linkage (Spielman et al. 1993). The test examines a single allele of a marker to determine whether it is transmitted from heterozygous parents to affected offspring more often than would be expected by chance, i.e. on more than 50% of occasions. The statistic relies on the transmissions from heterozygotic individuals only and is calculated according to the following formula:

$$\chi^2 = \frac{(a-b)^2}{(a+b)}$$

a = number of transmitted alleles

b = number of non-transmitted alleles

3.4.2.7.4 Relative risk (RR)

The Relative Risk (Woolf 1955) is an epidemiological term that is used to describe the proportion of people in the population with a disease/ disorder who have been exposed to a given risk factor divided by the proportion with the condition who have not been exposed to the risk factor. In terms of genetic association the term is used to measures the strength of association between a disease and genetic marker and was calculated according to the following formula:

$$RR = \frac{a/(a+c)}{b/(b+d)}$$

a = number of transmissions of 'associated' allele.

b = number not transmitted of 'associated' allele

c = number of transmissions of 'other' allele(s)

d = number not transmitted of 'other' allele(s)

The relative risk can take a value from 0-α. Confidence intervals for the relative risk estimates are quoted.

3.4.2.7.5 Attributable fraction (AF)

The AF (Levin 1953) for an'associated marker' is the fraction of the cases which would not have occurred had the risk factor been removed or reduced to population base-line levels. Where a positive association was identified the AF for the marker was calculated according to the following formula:

$$AF = \frac{f(R-1)}{1+f(R-1)}$$

f= Frequency of the 'disease-associated' allele in the population. In TDT the frequency of the non-transmitted parental alleles are used as a conservative estimate of the frequency of the disease associated allele in the population.

R = Relative Risk.

3.4.2.7.6 Odds ratio (OR):

The OR is also a measure of the strength of association between a marker and disease. It measures the probability that disease is present in the presence of a disease allele compared with the probability that it is absent. It was measured according to the following formula

$$OR = a/b$$

a = No. of transmissions of particular allele by heterozygotic parents

b = No. of non-transmissions of the allele by heterozygotic parents.

3.4.2.7.7 Extended- transmission disequilibrium test (e-TDT)

The eTDT is a program for testing multi-allelic association. It is an extension of the TDT and uses likelihood ratio tests based on logistic regression (Sham and Curtis 1995).

The *genotype-wise* test is a test of the probability of success for each of the observed heterozygous parental genotypes, or the probability in transmitting the first allele of the genotype and test whether the probabilities exceed 50%. The test is based on the comparisons of the likelihood maximized over separate transmission probabilities for each genotype with the likelihood assuming all these probabilities are 50%.

The *allele-wise* test is allele specific. Maximum-likelihood estimation of the allele-specific parameters are calculated using logistic regression analysis. The likelihood maximised over

these allele-specific parameters is compared to the likelihood that all parameters are equal using a chi-squared statistic with m-1 degrees of freedom (where m is the number of alleles).

The *goodness of fit* of the allele-wise and genotype-wise data are estimated by comparing the likelihoods for the genotype-wise test and the allele-wise test. A monte-carlo simulation is conducted as a test of significance. This performs 1000 simulations using the input data to generate genotypes and compares these simulations with the input data. The program runs in MS-DOS mode and requires an input file in LINKAGE file format.

3.4.2.7.8 Graphical overview of linkage disequilibrium (GOLD)

This is a software package that gives a graphical representation of linkage disequilibrium in genetic data (Abecasis and Cookson 2000). D' and delta^2 values are generated with relevant χ^2 statistics and p value significance. The program runs in MS-DOS mode and utilizes an input file with population genotype data for a series of markers (both microsatellite and SNP data). This program was used to calculate D' values across the region that underwent association mapping on chromosome 2q.

3.4.2.7.9 Haplotype analysis using TRANSMIT (v2.5)

TRANSMIT tests for association between genetic marker and disease by examining the transmission of alleles or genotypes from parents to affected offspring (Clayton 1999). It also examines transmissions of multi-locus haplotypes, phase known or unknown. The program produces asymptotic chi-squared tests for each haplotype or allele, a test on 1-df for excess transmission of that haplotype and a global test for association on H-1 df, where H is the number of haplotypes for which transmission data are available. The program runs under remote access via the HGMP website (http://www.hgmp.mrc.ac.uk/). The input file for transmit is a SPLINK file generated by the SPLINK program on HGMP from a pedigree data file. Haplotype analysis for the dense SNP mapping on chromosome 2q was conducted with this program. Features of TRANSMIT used in this analysis were the –agg and –c flags which allowed rare alleles to be aggregated or rare haplotypes to be ignored. This reduces the computational burden and prevents skewing of the χ^2-statistic. In addition

for haplotypes showing evidence of increased transmission, bootstrap significance testing was utilized. This performs haplotype simulations with the input data and generates a p value, which represents the proportion of bootstrap samples that give an equal or larger value of the test statistic to that observed. These statistics are at 1-df for individual haplotype and H-1df for the Global χ^2 statistic (where H is the number of haplotypes generated for a given number of loci).

3.4.2.7.10 Haplotype analysis/ tagging of SNPs using Haploview

Haploview is a software package still in development at the Whitehead Institute (Massuchessets Institute of Technology) http://www-genome.wi.mit.edu/personal/jcbarret/haplo/. It is designed to simplify and expedite the process of haplotype analysis by providing a common interface to several tasks relating to such analyses. HaploView currently allows users to examine block structures, generate haploypes in these blocks and save the data in a number of formats. It also calculates D' and r^2 measures of linkage disequilibrium, tests alleles at loci for association, constructs haplotypes of alleles at a series of loci and tests these for association and tags haplotypes with single nucleotide polymorphisms (SNPs) that are most informative about a given haplotype. The program only takes input data for SNPs or biallelic microsatellites. It runs in a java environment and the input file is formatted as a tab delimited file that contains information about the family and subject identity and the genotypes for a given set of loci.

Chapter 4

Description of clinical data

4.1 Introduction

In this chapter the clinical data collected for individuals is presented. This includes general demographic data, data derived from the clinical diagnostic instruments, the Autism Diagnostic Interview Revised (ADI-R) and the Autism Diagnostic Observation Schedule – Generic (ADOS-G), and data describing the available psychological measures. Following this is a more detailed description of four cases that were identified as possessing chromosomal abnormalities.

4.2 Description of sample

One hundred and fourteen children and their parents were recruited. The mean age of subjects was 10.33 years (range 4-34, sd = 6.57). 77.8% of the sample were male and 23.2% were female, a male to female ratio of 3.35:1. The sample was Caucasian and largely ethnically Irish. In four families, there was one Irish parent and one non-Irish (two were North American, one was British and one was Dutch). Mean maternal age at the time of birth of subjects was 32.38 (range = 20.5 – 42.75, sd = 4.53). Mean paternal ages at birth of subjects were 34.18 years (range = 20.58 - 40.58y, sd = 5.22).

Two families consisted of only mother and affected child, where the fathers were absent. The remaining families were trios of affected child and both parents. Blood samples were collected from parents with the exception of two cases with needle phobia where buccal swabs were obtained. Blood samples were collected from children with the exception of 12 cases. Buccal swabs were obtained from these children. No Fragile X testing or chromosomal analysis were conducted in these cases.

4.2.1 Exclusions and withdrawals

Three families withdrew from the study following the initial assessment. Reasons for withdrawal were not given in one case and for the other two were based on disagreement about participation between parents. Eleven subjects were excluded. Four did not meet the

required diagnostic criteria. One was found to have a cytogenetic abnormality of Ch2q (see section 4.5). One subject was found to have a 47XYY karyotype (see section 4.3). One was excluded on the basis of extreme prematurity (26 weeks gestation with associated neonatal problems). A further individual was excluded on the basis of low IQ (i.e. mental age score = 15 months using Vineland Adaptive Behaviours Scale). One was excluded as one parent was deceased a fact not disclosed at the time of initial contact, this family may have been included as a duo but the surviving parent was not considered a sufficiently good historian.

4.2.2 ADI data

The ADI-R data for study subjects are provided in Appendix B. The cut-offs for an ADI-R diagnosis of autism have been outlined above. Of the 100 subjects included in the final analysis, 31 were non-verbal and 69 were verbal. Average age when the ADI-R was conducted was 124 months (Range 48 – 408, median age-group 60-72 months).

In one case (AS001) the subject did not meet the criteria for autism in the repetitive behaviour domain. This subject demonstrated repetitive behaviours on assessment with the ADOS-G, including behaviours not reported by the ADI-R informant. A clinical judgement was therefore made to include this subject as the combined information from the ADI-R and ADOS-G supported the presence of abnormalities across all three domains.

4.2.3 ADOS data.

The ADOS-G assessment was performed in all cases to support the ADI-R diagnosis and the data are provided in Appendix C. In two cases it was not possible to complete the ADOS-G. These cases were adults who were non-verbal. Non-verbal adults are not suited to the ADOS-G assessment owing to the nature of the materials used in Module 1. An age appropriate version for non-verbal adults has not as yet been developed. Since these individuals had a clear history of autism and met ADI-R criteria they were not excluded from the study.

The ADOS-G generates two diagnostic categories, Autism and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). Sixteen of the children received

diagnoses of PDD-NOS using the ADOS-G. All of these children met the ADI-R criteria for autism. Six of the 15 were one point under the cut-off for Autism on the ADOS-G algorithm. The remaining 7 were two or more points under the cut-off for autism but all were above the cut-off for PDD-NOS. Consensus diagnoses were derived by comparing the ADI-R and ADOS-G data and the clinical history. Individuals scoring only one point under the cut-off for autism on the ADOS-G received a consensus diagnosis for autism. The rest were diagnosed autism spectrum disorder (ASD). The molecular data presented in the work here is for the whole sample.

4.2.4 Screening of subjects for Fragile X and chromosomal analysis.

Blood samples, where obtained, were sent to the National Centre for Medical Genetics, Our Lady's Hospital, Crumlin for Fragile X testing and chromosomal analysis. The incidence of Fragile X and chromosomal anomalies has been discussed in sections 1.3.5 and 1.3.6 respectively. During the course of study three cases were identified with chromosomal anomalies. These were a case of an apparent deletion on Ch2q, a case of 47XYY and an inversion on chromosome 9 that is a common variant in the population. Two further cases of identical twins with trisomy 15q11-q13 were referred to the study for phenotyping. The cases of 47XYY and the twins with trisomy 15q11-q13 will be discussed first followed by the chromosome 2q case, as this case subsequently determined the direction of the study.

4.3 A case of autism associated with a sex chromosomal aneuploidy

C. was 8 years 1 month when he was recruited to the genetic study. He was the second oldest of three children and lived with both parents and his siblings in a suburban setting. He was placed in an autism class in a local primary school. His parents had been concerned about his development since he was a young child. His father reported that he never cried and as a two-year old he appeared to dwell in a world of his own. His mother became concerned when he was 4 years old because he was still not toilet trained, did not mix well and his speech and language development appeared delayed for his age. He had an unusual preoccupation with washing machines, demonstrating an in depth knowledge of the details of the makes and features of the various models. During his early school years his teachers were not concerned with his behaviour and reassured his parents. Thus he first presented to

his local child and adolescent mental health services for assessment when he was 7 years old.

By the age of 7 years his parents were also concerned that he had become overactive and talkative. However, in the classroom environment he was generally quiet and withdrawn although his teachers found him distractible. His school performance was rated at the bottom of the class despite an apparently average intellectual ability. He continued to have problems mixing with his peers and he had no friends.

4.3.1 Obstetric history and early development:

He was the second child of a 28-year-old (para 1) mother. The antenatal history was complicated by the presence of antenatal bleeding for the duration of the first trimester that eventually ceased at 16 weeks gestation. This was attributed to a cervical polyp. In the second and third trimester there was reduced fetal movement and his mother commenced a fetal movement chart. He was born at term by uncomplicated delivery. His birth weight and his birth records were not available. He was well in the neo-natal period and went home with his mother within 24 hours of delivery.

At his six-week check he was found to have small testicles. No other abnormalities were noted. He was a very quiet baby. His parents recalled that when he was put down to sleep he would lie contentedly, appearing not to mind if he was alone. He sat at eight months and walked at fifteen months. He began using single words at 18 months and phrases at 30 months, thus he did not meet the criteria for speech and language delay as defined by the ADI-R, i.e. delay of single words later than 18 months and a delay of phrase speech until or later than 36 months. Toilet training was not achieved until he was 52 months. He had ongoing persistent nocturnal enuresis.

His father and a paternal uncle were reported to have reading disability. The same paternal uncle had meningitis as a child and was described as a poor mixer. This uncle had not been formally diagnosed with a social and communication disorder. Both the maternal grandmother and a second paternal uncle were diagnosed with bipolar disorder.

4.3.2 Assessment with ADI-R and ADOS-G

He was aged 8 years and nine months at the time of recruitment. Formal assessment for autism was conducted with the ADI-R and ADOS-G assessments. These highlighted abnormalities in the three domains of communication, social interaction and behaviour.

Communciation abnormalities were present and included the use of a stereotyped style of speech, particularly repetitive use of phrases from television. He had limited conversational skills, he generally only talked on subjects that were of interest to him. He rarely expressed an interest in others and never asked questions unless the answer concerned a special interest. He often asked inappropriate questions of others and appeared unconcerned if his questions were distressing, for example he repeatedly asked his separated aunt about her ex-husband. His speech was poorly modulated, loud and fast. In addition, he rarely pointed to express interest and rarely used gesture. This was more markedly abnormal when he was aged between 4-5 years. Similarly at this age, he seldom nodded or shook his head to indicate yes or no.

His social interactions were also abnormal. His eye contact was reduced when he was between 4-5 years, although at the time of assessment his mother rated it as good. He did not smile to be sociable and at 4-5 years he rarely greeted others when they came to visit. At this age also his social interactions were noted to be abnormal in that he rarely showed his parents things of interest to him, never shared, never showed sympathy to others if they were sick or unwell, nor did he seek out comfort if he was sick. His use of facial expression was limited and at times inappropriate. He was unaffectionate and actively disliked to be cuddled. He was unaware of social cues or rules and thus at times his behaviour was socially embarrassing, for example he shouted and was cheeky and rude even to strangers. He frequently used profanities in social situations often shouting them out loud without reservation to tell others to go away. As a young child he often wandered off and did not appear to experience normal separation anxiety. He did not engage in imaginative play on his own or with his peers and rarely joined in with normal children's social games, such as

ring-a-rosie. He was generally uninterested in other children and tended to avoid them. He had been subject to bullying in the past. He never played with a group of his peers.
His interests were restricted to activities he enjoyed and he engaged in these to the exclusion of anything else. He often spent hours drawing pictures (up to 15 at a time) and spent a long time ensuring that his drawings were perfect. As mentioned previously, he was preoccupied with washing machines and often examined the washing machine when he went visiting other people's houses. His use of objects was mildly repetitive, for example he occasionally engaged in spinning wheels of toy cars. He had olfactory sensory interests that had been more pronounced as a younger child. He engaged in hand mannerisms on a daily basis and demonstrated complex whole body mannerisms, e.g. bouncing and spinning occasionally. His gross-motor co-ordination was reported to be clumsy. He engaged in self-injury occasionally and was 'always on the go' at home.

His scores on the ADI-R algorithm were:

ADI-R:

Qualitative Impairments in Reciprocal Social Interaction	22
Communication	17
Repetitive Behaviours and Stereotyped Patterns	6
Evidence of Abnormality at/before 36 months	2

These scores met the cut-offs for a diagnosis of autism.

Direct assessment was performed with the ADOS-G assessment using a module 3, designed for children with fluent speech. Similarly abnormalities of communication, social interaction, play and behaviour were noted. His voice was monotonous. During the assessment he used stereotyped speech using phrases repetitively. His conversational skills were poor and conversation dried up quickly after one or two interchanges with the examiner. His use of gesture and eye contact was poor throughout the session. He used little facial expression and showed abnormalities in the nature of his social overtures and responses. He had very little understanding of friendship when asked, and appeared to have no friends. His use of the available play materials was limited, for example he did not build

any narrative into his play and he flicked the action figures repetitively rather than using them as independent agents. He had sensory interests that were evident in the form of scratching the surface of objects. Hand and finger mannerisms were observed during the course of the assessment. His general behaviour was notable for mild agitation/ overactivity.

His scores on the ADOS-G algorithm were as follows:

ADOS-G:

Communication:	4
Qualitative Impairments in Reciprocal Social Interaction:	9
Communication and Social Total:	13
Play:	1
Stereotyped Behaviours and Restricted Interests:	3

The ADOS-G assessment confirmed a diagnosis of autism.

A clinical psychology assessment had been performed at the time of his referral to the child and adolescent psychiatry services. A shortened version of the Stanford-Binet test battery was applied and he was determined to be in the average range of intelligence.

Physical examination did not reveal the presence of any obvious dysmorphology. He was noted to be of above average height (132.5cm, 75th percentile) and his head circumference was 54cm. Gross neurological examination was normal.

Blood samples were obtained for fragile X testing and chromosomal analysis. Cytogenetic analysis of G-banded metaphases revealed that he had a 47, XYY karyotype.

4.3.3 Discussion of 47, XYY and autism

The 47, XYY occurs in approximately 1 per 1000 male births (Nielsen J 1991). The condition is mild, affected individuals are usually not detected at birth. Intellectual function is within normal limits and there is usually no dysmorphology detected. Associated features

include above average height, mild learning difficulties and speech and language delay (Ratcliffe 1999). One study reported poor social relatedness in addition to speech and language difficulties in 11 out of 21 cases studied (Nielsen et al. 1973). Conduct and impulsivity problems occur more often and an association has been reported with increased criminality (Gotz et al. 1999). Approximately 10% of cases have a 46, XY, 47, XYY mosaicism and a further 10% have 48, XXYY. There have been several case reports in the literature of autism or pervasive developmental disorders associated with a supernumerary Y chromosome (Gillberg and Wahlstrom 1985; Gillberg et al. 1987; Nicolson et al. 1998).

This case presents with features described in association with the 47, XYY condition. He was of above average height, was considered overactive, and had speech and language delay and conduct and aggression problems. In addition he met the criteria for autism on both the ADI-R and the ADOS-G. In his family history there was a history of dyslexia, possible social relatedness problems in an uncle and a history of adult onset bipolar disorder in both maternal and paternal families.

There are two questions that arise in relation to this finding. The first is whether the karyotype observed in the individual had any effect on the phenotype. The second is whether an additional Y chromosome tells us anything about autism and its potential underlying pathophysiological processes. The answers to both must unfortunately be speculative. In the first instance, it appears from the history that there was a possible family history of autism and other developmental difficulties were observed (dyslexia). From population studies of individuals with 47XYY it is also known that deficits such as developmental delay and speech and language and reading disorders occur (Walzer et al. 1990; Abramsky and Chapple 1997). This suggests some similarity with the features of autism. However features of conduct disorder, behavioural difficulties and subsequent elevated rates of criminality are also reported in association with the 47, XYY karyotype (Rutter et al. 1975; Gotz et al. 1999). Furthermore there have been reports of an increased incidence of the 47, XYY karyotype in childhood-onset psychosis (Kumra et al. 1998). With the evidence that an additional Y chromosome has such a broad range of effects it seems more likely that it is potentially acting as a modifier with respect to the clinical

phenotype. The question then arises as to the potential mechanisms through which the Y chromosome could possibly influence neurodevelopment, specifically the expression of neurodevelopmental genes.

4.3.3.1 Potential role of genes on the Y chromosome in neurodevelopment

The structure of the Y chromosome has been well characterized and it has very few known functioning genes, probably fewer than 50 (Thompson 2001). Sex determination and spermatogenesis appear to represent the main function of genes present on the Y chromosome with the exception of those contained within the pseudoautosomal region. Of the 20 annotated genes on the Y chromosome contained within NCBI database, 5 appear to have testis-specific function, one is a transcription factor, one is a translation initiation factor, one is involved in extracellular-matrix protein production. The remainder have either unknown or inferred function or no known protein products.

Other sex chromosome aneuploidies, e.g., 47, XXY and 47, XXX karyotypes also present with abnormalities in development, such as speech and language delay, lowered IQ, reading difficulties. XXY boys have been reported to present with peer and sibling relationship difficulties but were less anti-social than those with the 47, XYY karyotype (Robinson et al. 1986). The IQ levels in the 47, XXX females are more significantly lowered compared with controls than either of the other karyotypes (Robinson et al. 1990; Stewart et al. 1990). Psychiatric difficulties have been reported in association with this karyotype in the form of depression, drug abuse and obsessive-compulsive disorder. The XO phenotype presents with lowered IQ well-characterised dysmorphology. More recently face and emotion recognition deficits have been reported and attributed to anomalies in amygdala development (Lawrence et al. 2003). Subsequently imaging studies have demonstrated the presence of enlarged amygdala volume in XO females compared with normal male and female controls. The authors attempted to map genes on the X chromosome that may predispose towards the neurocognitive deficits using females with X-chromosome deletions who also showed the face and emotion recognition deficits (Good et al. 2003). A 4.5Mb region on the X-chromosome was identified that contained several genes including those encoding genes for MAO-A and MAO-B. The authors postulate that the observed

phenotype may be attributable to haploinsufficiency for genes contained within the region. If susceptibility genes within the region were found to be subject to gene dosage effects it may also be postulated that the associated phenotype observed with the supernumerary X sex chromosome abnormalities may be attributable to an excess gene dosage effect. This does not explain the phenotype observed in the 47, XYY karyotype unless susceptibility genes subject to dosage effects were identified within the pseudoautosomal region. The role of the Y chromosome in autism has been discounted in one study that found no difference between Y chromosome haplogroups and in different populations of individuals with autism (Jamain et al. 2002b).

4.4 Identical twins with maternally derived trisomy 15q

A. (Twin 1) and R. (Twin 2) were 5-year-old twin girls referred for assessment with delayed language development and behaviour problems. They were the first live-born children of a healthy unrelated couple.

4.4.1 Obstetric history and early development

The girls were born by Caesarean Section at 33 weeks gestation following a pregnancy that had been complicated by minor antenatal bleeding. Both twins weighed approximately 2,500g. Apgar scores were normal. Both were jaundiced after birth and Twin 1 required phototherapy. The neonatal period was complicated by feeding difficulties associated with failure to thrive and sleep disturbance. In the first year of life both girls had ear and throat infections and recurrent bouts of diarrhoea. Early difficulties were largely attributed to prematurity.

Twin 1 achieved her motor milestones normally. Toilet training was not achieved by the time of the fifth birthday. She was non-verbal. At 12 months she used approximately four words, which were predominantly echoed. She never acquired phrase speech. Twin 2 also reached motor milestones normally. Toilet training was also not successfully completed by the fifth birthday. Some speech had been acquired; this was mainly comprised of echoed speech and single words. All of her verbal interactions were restricted to the communication of her needs.

Their parents consulted the family GP when they were aged 15 months, as they were concerned about tantrums, developmental delay and absence of normal play. Both twins were assessed in a paediatric neurology clinic and referred to the study for assessment of the clinical phenotype.

4.4.2 Family history

The twins are the eldest of three children. They have a younger brother who is phenotypically normal and has a normal karyotype. The maternal grandmother had five miscarriages of unknown cause. She had one other live female birth besides the twin's mother. This maternal aunt was reported to be socially odd and very rigid and orderly, but had no formal psychiatric diagnosis. She was single and had no offspring.

4.4.3 Assessment of twin 1 with ADI-R and ADOS-G

4.4.3.1 ADI-R assessment

This girl appeared to dwell in a world of her own. She made few non-verbal attempts to communicate. She ignored others unless vigorous attempts were made to attract her attention. She rarely used communicative gestures and pointing was restricted to indicating her needs, not to express interest. She rarely imitated family members. Her understanding of language was restricted.

Her eye contact was fleeting during social interactions and she had little social smiling. Any social overtures were normally highly motivated by her own needs and such overtures showed poor integration of gaze, gesture and verbal attempts to communicate.. Her use of facial expression was limited and inappropriate at times. She was disinhibited in social situations appearing to be unaware of the social expectations of the situation. She was generally unresponsive to the social approaches of others especially those of strangers and did not interact with her peers. She did not show things or share objects or enjoyment with others. She did not offer comfort to others, although she would seek out her parents if she was sick or hurt.

She did not engage in normal play at all. Her use of objects was very repetitive including spinning wheels, collecting together stones and tapping objects. She preferred routine and became distressed at unexpected changes. There were food rituals that involved the consumption of food according to colour and texture. She developed unusual attachments to objects such as a part of a doll or piece of lego or jigsaw and became distressed if these were misplaced. She demonstrated a number of sensory interests including examination of objects by peering at them, feeling their texture and the noise they produced when tapped. She showed sensitivity to the noise of household appliances such as the vacuum cleaner, washing machine and lawnmower limiting their use at home. She engaged in finger-flicking mannerisms regularly and occasional complex mannerisms including bouncing and flapping. She demonstrated occasional toe walking.

Her scores on the ADI-R were as follows:

ADI-R:

Qualitative Impairments in Reciprocal Social Interaction	28
Communication	14
Repetitive Behaviours and Stereotyped Patterns	6
Evidence of Abnormality at/before 36 months	4

4.4.3.2 ADOS-G assessment

A module I ADOS-G assessment for non-verbal subjects was conducted. During the assessment she made few attempt to communicate, appearing to be unaware of the examiner. Her eye contact was poor and her facial expression was unvarying throughout the assessment. She did not respond to her name or efforts to initiate joint attention. She showed no interest or enjoyment in any of the tasks. However, she had a high-pitched scream that she used to indicate her distress or displeasure at something, which was mainly directed at the examiner's attempts to engage with her. On one highly motivated approach (offering her a snack) she indicated her preference for food items by reaching out her hand to the snack containers. Her play was exceptionally limited. She did not play functionally with any of the toys preferring instead to tap them repetitively. She peered at the light and the corners of the room. She became distressed towards the end of the assessment and lay

on the floor screaming. Complex mannerisms in the form of bouncing were observed during the assessment.

She met the criteria for autism on both assessments showing deficits in all three areas of communication, social interaction and behaviour.

Her scores on the ADOS-G assessments were as follows:

ADOS-G:

Communication	6
Qualitative Impairments in Reciprocal Social Interaction	14
Play	4
Stereotyped Behaviours and Restricted Interests	6

4.4.4 Physical examination of twin 1

She was on the third centile for weight and height. She was not grossly dysmorphic but had a large port-wine haemangioma on her left leg extending from her foot to buttock. She also had a right divergent strabismus. There was no evidence of kyphoscoliosis.

4.4.5 Assessment of twin 2 with the ADI-R and ADOS-G

4.4.5.1 ADI assessment

This twin had some single words that were mainly used to communicate her needs. She had similar deficits to her sister in that she did not imitate, point or make use of gesture. She showed limited response to others without deliberate efforts to catch her attention. She had shown sensitivity to noise from household appliances in the past, although this was not a problem at the time of assessment.

Unlike her sister her direct gaze was reported to be better while undertaking tasks with others and she was more likely to share objects and food. She also shared her enjoyment of social games such as tickling. Her social overtures had better integration of gaze and gesture than those of her sisters but these were also limited. Unlike her sister she demonstrated spontaneous affection. She often attended to new toys and was more likely to

engage in the to-and fro aspects of tickling games. She was more receptive to the approaches of other children although would never attempt to initiate any interactions.

Her understanding of language was also judged to be limited. Like her sister, she did not show objects out of interest, was unaware of the enjoyment of others, showed social disinhibition and had a restricted range of facial expressions. She rarely responded to social interactions by strangers and had little interaction with her peers. In addition, her use of objects was mainly repetitive. She was upset by changes in routine and had the same eating ritual as her sister. She had similar sensory interests. She demonstrated complex mannerisms in the form of bouncing and flapping.

Her scores on the ADI-R were as follows:

ADI-R:

Qualitative Impairments in Reciprocal Social Interaction	25
Communication	14
Repetitive Behaviours and Stereotyped Patterns	5
Developmental Abnormality at/ before 36 months	4

4.4.5.2 ADOS-G assessment:

She was more compliant on direct assessment than her sister. However her social interaction was also poor. She did not seek to engage with the examiner during the assessment and showed little enjoyment of the tasks performed with her. Her eye contact was generally poor. She showed some play with one of the cause and effect toys but had no functional or pretend play with miniatures. She also demonstrated some sensory interests, she peered at objects for a length of time. She tapped objects repeatedly in a similar manner to her sister. There was also evidence of complex mannerisms (flapping and bouncing). She was quite overactive during the assessment.

Her scores on the ADOS-G assessments were as follows:

ADOS-G:

Communication	5
Qualitative Impairments in Reciprocal Social Interaction	14
Play	4
Stereotyped Behaviours and Restricted Interests	6

The ADI-R and ADOS-G assessments for both of the girls were consistent with diagnoses of autism. The scores were remarkably similar despite the qualitative observation that one child (Twin 2) was more interactive than her sibling.

4.4.6 Physical examination of twin 2

General physical examination of this girl was normal and there were no dysmorphic features seen.

Psychological assessments were performed on both girls for the purposes of school placement but were incomplete owing to lack of co-operation with the assessment.

4.4.7 Cytogenetic analysis

The twins were referred for cytogenetic testing following their presentation to the paediatric neurology services. An additional small satellite chromosome was identified in both girls. G banded and Fluorescent in Situ Hybridisation (FISH) analysis identified the additional chromosome to be a fusion chromosome made up of the heterochromatic short arm of Chromosome 15 and the proximal long arm of chromosome 15 to band q13.1 with a small telomeric piece of the long arm of chromosome 21 fused to the long arm of chromosome 15.

There was trisomy for these parts of 15 and 21. FISH analysis indicated that the derived Chromosome 15 painted with whole Chromosome 15 paint and was positive for the SNRPN probe, which maps to the Prader-Willi/ Angelman's critical region (PWACR) on Chromosome 15. The additional chromosome also hybridised with the Chromosome 21 telomere probe D21S1219 confirming the presence of Chromosome 21 material at the

distal end of the derived chromosome. Both twins were therefore trisomic for a maternally derived additional chromosome consisting of the heterochromatic short arm of chromosome 15, the euchromatic proximal long arm of chromosome 15, and the telomeric end of the long arm of chromosome 21.

Their mother was found to carry a balanced reciprocal translocation 15;21 with breakpoints in bands q13.1 and q22.3 respectively. The additional chromosome found in the twins was judged to have arisen from an unbalanced 3:1 segregation of the translocation in the maternal meiosis. The maternal grandfather was also found to carry the balanced translocation seen in the mother. Both the grandfather and the mother were phenotypically normally. The maternal aunt declined cytogenetic testing.

Molecular genetic analysis of the twins using the Jefferies multilocus fingerprinting probe showed an identical fingerprinting pattern, indicating a very high likelihood that the twins are monozygotic.

4.4.8 Discussion of maternally derived Ch15q

The most commonly reported clinical findings found in association with Trisomy 15q11-q13 are hypotonia, epicanthal folds, high arched palate, large ears, low stature, and the presence of seizures. Developmental abnormalities observed include delayed motor milestones and lack of speech acquisition/ speech delay and moderate to severe intellectual disability. Autistic behaviours occurred frequently, including lack of social reciprocity, lack of eye contact, no interest in peers, failure to seek comfort, stereotyped behaviours and repetitive play, insistence on routines and hyperactivity (Wolpert et al. 2000).

The cases presented here are similar to those in the literature with respect to the clinical features of trisomy 15q and the autism presentation. Both girls were on the 3rd centile for height, one was completely non-verbal and the other had very little speech and no established phrase speech. They both appeared to be significantly delayed intellectually although no formal measure of this was available. They both met the criteria for autism on assessment with the ADI-R and the ADOS-G. These assessments differ substantially in the

scores that each of the girls received. However from the reports of schoolteachers and parents it was apparent that there was an observable difference in the considered level of functioning. R. was considered more socially communicative (albeit in the context of significantly abnormal social interactions) and used verbal attempts to communicate while her sister was completely non-verbal. It is not possible from the case history to determine any obvious environmental factors that may have influenced the differing outcomes in these girls. They had similar obstetric histories and neonatal histories. Just as in the last case, the question also arises as to whether the phenotype is solely attributable to the cytogenetic abnormality or if in such cases there is a prior susceptibility to autism. There was a suggestion from the family history that others in the family had developmental difficulties and social abnormalities. It seems however likely that the cytogenetic findings play a substantial role in the influence of the clinical presentation in these girls. Interestingly there is converging evidence for susceptibility genes for autism within this region on chromosome 15q.

Maternally derived trisomy for Ch15q11-q13 occur at a rate of about 1% in autism and are the most commonly occurring cytogenetic abnormality found in the disorder (Cook et al. 1997a; Schroer et al. 1998; Weidmer-Mikhail et al. 1998). Autistic features occur in cases of trisomy 15q11-q13 that include the Prader-Willi Angelman Critical Region (PWACR). In a report of two individuals with a chromosome 15q duplication, Baker et al., 1994, confirmed the presence of the PWACR and postulated that this region may be of aetiological significance. Two studies reported autism in maternally derived interstitial duplications of the proximal portion of the long arm of Ch15, which contained the PWACR (Browne et al. 1997; Cook et al. 1997a) and in a review of the literature on case reports Wolpert found nine out of the 17 cases where the duplications were maternally derived (Wolpert et al. 2000). As discussed in the introduction, linkage disequilibrium (LD) studies have been conducted across the region in autistic samples and evidence of LD has been shown around a microsatellite marker in the γ-aminobutyric $acid_A$ receptor subunit gene (GABRB3-155CA-2) (Cook et al. 1998) and at a marker 60kb beyond the 3' end of the gene, GABRB3 (UniSTS:156408) (Martin et al. 2000). Further evidence for a potential candidate gene within the region has been obtained from genomewide linkage studies. A

principal components analysis of the clinical data (ADI-R scores) collected by the Collaborative Linkage Study in Autism showed that there was increased evidence of linkage in families where affected individuals showed savant skills (Nurmi et al. 2003). As discussed in the introduction, several candidate genes have been identified within the region and studies have shown some evidence for LD between variants within the genes and autism. The genes identified include GABRB3, UBE3A and ATP10C (Cook et al. 1998; Martin et al. 2000; Nurmi et al. 2001).

An important aspect to this finding is the length of the supernumerary marker chromosome, which includes the short arm and proximal long arm of chromosome 15 to band q13.1. This is a larger portion of chromosome 15 than those previously reported and is very rare. In addition a telomeric part of the long arm of chromosome 21 is present at the distal end. The potential role of this additional portion of chromosome 21 for which the girls are also trisomic is uncertain without mapping the breakpoints at the molecular level.

4.5 A case of autism associated with chromosome 2q anomaly

K. M. was a fourteen-year-old boy who was ascertained and initially recruited to the trio-based association study in autism.

4.5.1 Obstetric history and early development

He was born by spontaneous vaginal delivery at term to healthy parents (mother 20y, father 27y) following a normal pregnancy. Early milestones were normal. He did not have any congenital malformations, and his physical development and growth were normal.

Abnormalities in his development were first noted by his parents at 30 months when he stopped using previously acquired language. This was associated with the onset of self-injury, tantrums and a dislike of crowds. At the time of his third birthday he was referred to a paediatrician who diagnosed developmental delay. Toilet training was not fully achieved until 72 months and there were episodes of nocturnal enuresis occurring until 96 months. He had difficulties in mainstream school, which he commenced at 51 months and these continued despite a change of school. Following psychological assessment he was

diagnosed as being on the autism spectrum with a mild learning disability and was referred to a local special needs school. At the time of recruitment his academic progress was in keeping with his level of learning disability, however he appeared to display ability for reading and language comprehension that exceeded the level expected in the normal population for this age group. Parents and teachers reported this although formal psychological testing had not substantiated it.

4.5.2 Family history

K was the younger of two children. He had an older phenotypically normal sister. Two paternal uncles were reported to be "odd" but had no psychiatric diagnoses. There was no other family history of note.

4.5.3 Assessment with the ADI-R and ADOS-G

4.5.3.1 ADI-R assessment

The ADI-R indicated that he met the criteria for classic autism at age 4-5 years. He had deficits in all three areas of communication, reciprocal social interaction and behaviour. His language was limited to simple phrase speech at 4-5 years and primarily for the purposes of addressing his needs. He did not make use of gaze and gesture at this age. At the time of assessment his language was quite advanced for his chronological age and general level of ability but he still had limited integration of gesture with language for the purposes of communication. At 4-5 years he was described as living in a world of his own and did not mix with his peers. Play was limited and lacking imagination. These difficulties forming normal social relationships had persisted and he had been the subject of bullying. He had a restricted range of interests in books and locomotive engines, which dominated his conversation. He showed evidence of repetitive use of objects in the past. He was somewhat routine and ritualistic. He had stereotyped complex mannerisms in which he engaged when bored or frustrated. He engaged in self-injury in the past but this was no longer in evidence.

His scores on the ADI-R algorithm were as follows:

ADI-R:

Qualitative impairments in reciprocal social interaction	20
Communication	12
Repetitive behaviours and stereotyped patterns	5
Abnormality of development evident at/ before 36 months	6

4.5.3.2 ADOS-G assessment

Direct assessment with the ADOS-G derived a diagnosis of pervasive developmental disorder. He scored mildly under the cut-off for classic autism on reciprocal social interaction. Notably his eye contact was considered quite good.

His scores on the ADOS-G algorithm were:

ADOS-G:

Communication	4
Qualitative impairments in reciprocal social interaction	4
Combined communication and social score	8
Imagination/ creativity	1
Stereotyped behaviours/ restricted interests	1

4.5.4 Physical findings

On physical examination he was noted to be tall and slim with long fingers and toes. He displayed hyperextensibility around his wrist and had flat feet, a high-arched palate and pectus carinatum. A community paediatrician had raised the possibility of Marfan syndrome. There was no family history of Marfan's. An echocardiogram was normal. An eye examination for lens dislocation was normal. His appearance was therefore of a Marfanoid habitus, rather than actual Marfan's syndrome.

4.5.5 Cytogenetic studies

Blood samples were sent to the National Centre for Medical Genetics in Our Lady's Hospital for Sick Children for cytogenetic analysis. Chromosomal analysis of G-banded metaphases showed a male karyotype with what appeared to be a small interstitial deletion

on part of the long arm of chromosome 2. (del(2)(q32.1q32.2) or (q32.2q32.3)) (Fig. 1). Exact breakpoints were difficult to define but the deletion appeared to involve either band 2q32.1 or 2q32.3. As there are reports in the literature of an association between autism and a deletion of 2q37 {Ghaziuddin, 1999 #35}, FISH analysis using the 2qtel probe was performed for this region. This showed the presence of the distal portions of both chromosomes 2.

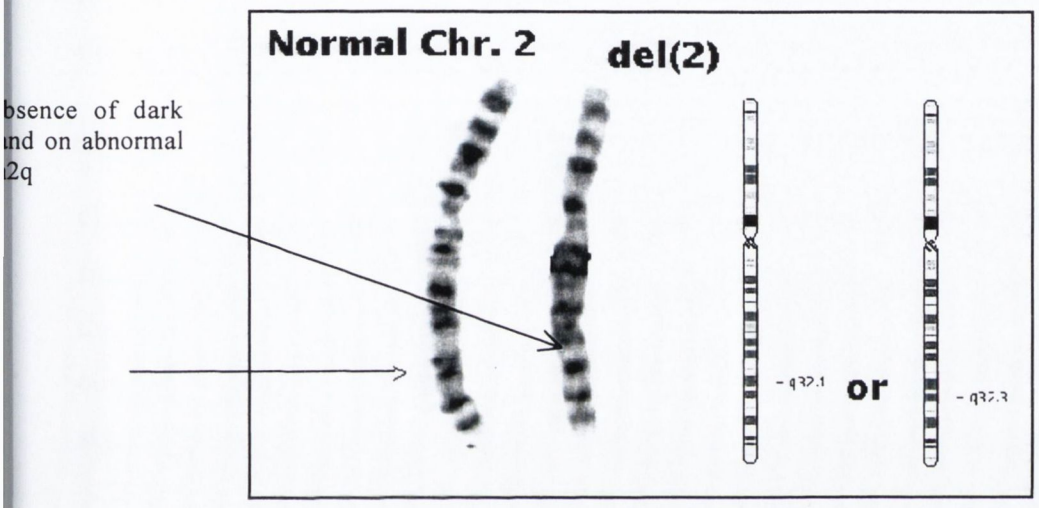


Figure 4. 1 Two copies of chromosome 2 from K.M. showing the absence of a dark cytogenetic band on one chromosome (del(2))

Following the identification of the abnormality, parental blood samples were also sent for analysis. Chromosomal analyses of the parents revealed normal karyotypes in both cases. Thus the identified deletion seems to have arisen *de novo.*

4.1.21 Discussion of case of autism with apparent deletion on chromosome 2q

This is the first published case reporting an association between autism and an interstitial deletion of chromosome 2q. There have been reports of an association between the disorder and deletions involving 2q37 (Ghaziuddin and Burmeister 1999; Stein *et al,* 1992) but the presence of this region in this individual was confirmed. There have been a number of other reports of deletions of 2q31q33 (Taysi et al. 1981; Al-Awadi et al. 1983; Buchanan et al.

1983; Franceschini et al. 1983; Young et al. 1983), Bensen *et al,* 1986) and at least one report of a duplication of this region (Ramer et al. 1990). None of these cases report an association with autism although as discussed in chapter 1, section 1.3.6.3, some of these reported cases were too young at the time for a diagnosis of autism to have been made. The most consistent clinical manifestations of these reports include low birth weight, cleft palate, developmental delay/ learning disability generally in the moderate range, microcephaly and abnormalities of the digits including syndactyly and camptodactyly. Pai reported siblings with a deletion of 2q32 (Pai et al. 1983). Both cases had developmental delay, microcephaly, micrognathia and abnormal ears.

This case differs from all of the above in that the deletion reported is much smaller and the clinical manifestations are mild. Apart from above average height, which appears to be familial, and high-arched palate no dysmorphic features were noted. The level of learning disability was in the mild range with an above average ability for reading and language comprehension out of keeping with the degree of disability. His clinical presentation is significant mainly for the presence of an autistic spectrum disorder.

A comparison of the cytogenetic bands that were reported to be deleted were compared with a region of putative linkage on chromosome 2q as discussed in section 1.3.8.1. Approximate relationships between the physical positions of microsatellite markers showing evidence of putative linkage as reported by IMGSAC, Buxbaum and Shao and the apparent deletion were estimated from the publicly available databases (http://www.ncbi.nlm.nih.gov/, http://www.ensembl.org/Homo_sapiens/, http://genome.ucsc.edu/cgi-bin/hgGateway). This revealed that the markers showing evidence of linkage in the IMGSAC study and Buxbaum's study were centromeric to the suspected deletion. The markers in Shao's study were located telomeric to the 'deletion'. The locations of these regions relative to the deletion are illustrated in Figure 6.1. These converging findings are evidence for the potential presence of a candidate gene for autism within this region. Based on this finding it was determined to follow up this region further. A collaboration was established with the National Centre for Medical Genetics. The plan for follow-up was:

(1) To map the extent of the deletion.

(2) Determine the breakpoints particularly with respect to the possible disruption of a known or predicted gene,

(3) Define the candidate genes within the region based on the published databases.

(4) Conduct a fine-scale linkage disequilibrium mapping experiment across the region defined by the observed deletion in a sample of individuals with autism

(5) Conduct a linkage disequilibrium fine-mapping experiment across the region of putative linkage as identified by (Buxbaum et al. 2001), (IMGSAC 2001) and (Shao et al. 2002a).

The results of these investigations are described in chapter 6.

4.6 General discussion

4.6.1 Clinical data

The data reported here represents the first phase of clinical sample collection to this autism genetics study. This phase of the study took place over a two-year period. The collection of this kind of clinical data are necessarily slow because of the lengthy research tools that must be utililised in order to derive the best possible research diagnoses for autism. The advantage of this detailed recruitment process is that cases are well characterized. While the molecular studies that are presented here must be considered exploratory in view of the small sample size, the advantage is that recruitment has continued and future studies will be possible in a larger well-characterized sample. It would also have been preferable to determine the psychological test scores using a standardized tool rather than relying on data available from schools and case notes. However, no funding was available at the time. To ensure that cases fell within the criteria for recruitment on their mental age scores, the Vineland Adaptive Behaviour Scales was utilized in the absence of neuropsychological testing.

As discussed above (section 4.2), the sample includes 7 children who met the criteria for autism spectrum disorder (ASD) and not the narrower diagnosis of autism. As this

represented a small group of children, it was decided to include them in the molecular investigations. To remove them from the sample would further decrease the power of the study for the purposes of genetic investigations. In addition, these children met ADI-R criteria for autism so it was decided that it would not substantially increase the heterogeneity of the sample by including them in the analysis. Following the second phase of clinical sample collection it is anticipated that molecular investigations will be reported for a subdivided sample, the core autism group and the ASD group.

4.6.2 Chromosomal abnormalities and autism

In a review of epidemiological data for autism Fombonne reported that chromosomal anomalies occur at rates of around 5% (Fombonne 1999). The majority of anomalies observed consist of the Fragile X anomaly and Trisomy 15q11-q12. The rate of detection of chromosomal abnormalities in the study population was 1.8% (2 in 114 subjects). The twins with the maternally derived trisomy 15 were referred to the study for the purposes of phenotyping following the identification of the abnormality and are therefore not included in this calculation. This rate of occurrence is thus lower than the population data reported by Fombonne. Of course the study sample does not represent a population sample and therefore the rate of occurrence of cytogenetic abnormalities in the sample is not directly comparable with the detection rate in a population sample. Additionally, many of the children referred to this study had already undergone cytogenetic testing and Fragile X analysis as part of a general medical assessment. Thus in addition to the twins reported in section 4.4, at least two other cases with cytogenetic abnormalities and one family with the fragile X anomaly were excluded from the sample reported here. Clinical assessments of these children and molecular investigations of the cytogenetic abnormalities are planned. The current recommendations regarding chromosomal analysis in children with autism or pervasive developmental disorders are that children with evidence of dysmorphology should be routinely screened (Filipek et al. 2000).

Chapter 5

An association study of variants in the HOXA1/ HOXB1 gene and autism

5.1 Introduction

The HOXA1 and HOXB1 genes are paralogous genes occurring on chromosomes 7p and 17q respectively. They have a single ancestral source analogous to the labial genes of Drosophila (Ruddle et al. 1994). They belong to the homeobox family of genes, a superfamily with the general function of encoding for homeodomain proteins. The homeodomain is a DNA binding domain effecting the transcription of other genes. The homeobox family of genes is well conserved in evolutionary terms and widespread in terms of phylogenetic distribution. The expression of the HOXA1 and HOXB1 genes has been shown to occur in the hindbrain during neural tube formation (Murphy and Hill 1991)

5.1.1 Identification of human HOXA1/ HOXB1.

(Hong et al. 1995) determined the structure of HOXA1 by cloning the cDNA which predicted a 335 amino acid protein. The protein was seen on in-vitro translation and an alternative splice copy was also observed. HOXA1 is the first gene occurring in the HOXA cluster and was localized by Fluorescent In-situ Hybridisation (FISH) to 7p15.3 (Apiou et al. 1996). The authors also localized the HOXB1 gene to 17q21-q22

5.1.2 Variants of the HOXA1/ HOXB1 genes in humans.

A common polymorphism was identified in the HOXA1 gene (Ingram et al. 2000a). It consisted of an A-to-G substitution at base 218 that changed the codon for one histidine in a series of histidine repeats to an arginine at position 73. The G allele had a frequency of 20 to 60% among the Coriell Human Diversity Panel. It was not identified in Asian individuals including Indians, Japanese, and Chinese. A second variant was detected in the first exon of HOXA1 in an affected cousin of a proband and her father. The polymorphism included the same base substitution as the A218G allele and three codons for histidine were absent from a series of histidine repeats.

A 9bp insertion was identified after base 88 in the HOXB1 gene that encodes histidine-serine-alanine. Two additional sequence changes were noted in association with the insertion. This included a thymidine substitution for adenine at base 315 changing a codon for glutamine to histidine. The second observed change was an adenine substitution for guanine at base 456, which appeared to have no effect. These associated changes were not detected in any of the cases where the insertion did not occur. This polymorphism was also determined as a high frequency allele in European or African populations. Only one individual of Asian origin was a heterozygote.

5.1.3 Mouse HOXA1/ HOXB1 mutants.

Mouse HOXA1 occurs on chromosome 6p and HOXB1 on chromosome 11p. Mouse mutants for the HOXA1 and HOXB1 genes were found to have abnormally developing rhombomeres (a transient subdivision of the hindbrain early in development) HOXA1 mutant homozygotes demonstrate the absence in development of the 5th rhombomere (Carpenter et al. 1993; Mark et al. 1993). The 4th rhombomere is also disrupted and the neural crest structures emanating from it are hypoplastic (Lufkin et al. 1991; Chisaka et al. 1992; Carpenter et al. 1993; Mark et al. 1993). Specifically they lack the superior olive, the Abducens nerve and all but the anterior portion of the facial nerve.

HOXB1 mouse mutants show a milder phenotype. The organization of the rhombomeres is not disrupted and the neural crest structures are maintained. The identity of cells in the 4th rhombomere appears to be affected. They appear to have greater deficits of the facial nucleus neurons and neurologic dysfunction in the muscles of facial expression (Goddard et al. 1996; Studer et al. 1996).

Double mutants for HOXA1 and HOXB1 show a more marked phenotype with loss of both the 4th and 5th rhombomeres and the associated neural crest tissues are missing. These observations in mice have led (Rossel and Capecchi 1999) to argue in favour of commonality of purpose for these genes and/ or overlap in function.

5.1.4 Autism and abnormalities of development similar to those observed in mouse null mutants for HOX genes.

Autism is found to occur at greater frequency as a result of *in-utero* exposure to several teratogens including thalidomide, ethanol and retinoic acid. Stromland reported 86 cases of thalidomide exposure where 15 of these individuals also had autism (Stromland et al. 1994). These individuals had the characteristic ear abnormalities associated with thalidomide exposure and one individual had both ear and limb abnormalities. However none had limb abnormalities alone. Since limb abnormalities alone were associated with thalidomide later exposure than ear abnormalities or the combined type, the finding appears to suggest that the neurodevelopmental insult resulting in autism occurred around 20-24 days post-conception. At this time the neural tube is closing, the CNS is divided into rhombomeres and simultaneously, the motor neurons are developing (specifically those of the eyes, face, tongue, jaw, throat and larynx).

Valproate exposure during the in-utero period has also been associated with ear malformations (Ardinger et al. 1988; Moore et al. 2000). A decrease in the number of neurons in cranial nerve nuclei has also been observed (Rodier et al. 1996). Ingram et al, 2000, reported reduced cerebellar Purkinje cell number associated with pre-natal valproate exposure in rats (Ingram et al. 2000b). This feature has been reported in post-mortem findings in autism (Ritvo et al. 1986; Bailey et al. 1998; Kemper and Bauman 1998).

In addition, an individual with autism was found at post-mortem to have similar abnormalities to the knockout mice previously described (Rodier et al. 1996). The reported abnormalities included impaired motor function associated with loss of the facial nucleus and superior olive.

5.1.5 Autism and the HOXA1/ HOXB1 variants.

Following the identification of the HOXA1 and HOXB1 variants described above, an investigation of their frequency in autism was undertaken (Ingram et al. 2000a). Fifty-seven probands with autism spectrum disorders were included (40 with autism, 10 with Asperger disorder, 6 with Pervasive Developmental Disorder and one with Disintegrative Disorder).

The HOXA1 G allele was found to occur at a frequency of 0.202 in "affecteds", 0.203 in "affected relatives", 0.164 in "unaffected reltives" and 0.109 in the convenience population (χ^2=7.92, p=0.005). Among affected offspring, a significant deviation from Mendelian expectation in gene transmission was observed. An increase in the number of transmissions of the G allele from matings between A/G mothers and A/A fathers was observed. In addition they noted that all nine affected females were heterozygotes (χ^2= 9.0. p=0.003). Four non-affected female family members were also heterozygotes. In contrast the male ratio of heterozygote to homozygote from the same matings was effectively equal (19:18).

There was no significant deviation from Hardy-Weinberg equilibrium for the HOXB1 variant. There was no significant deviation in transmission from Mendelian expectation although a non-significant trend towards transmission of the heterozygote status (+/INS) to affecters was noted.

In a replication study, Li et al, 2002, sequenced the variants again in 24 individuals with autism (Li et al. 2002). The same variants were identified in both genes and further variants in the HOXB1 gene were identified; A315T and G456A which were determined to be in strong LD with the insertion polymorphism. Association testing of the variants was conducted in 110 multiplex families (affected sibling-pairs) Statistical analysis was performed using the sib-TDT test in ASPEX and the McNemar Chi-squared. There was no significant deviation from Hardy-Weinberg Equilibrium in their population for variants in either the HOXA1 and HOXB1 genes. Parental transmissions were investigated as Ingram had reported an increase in transmissions from heterozygote mothers. This was not replicated in this investigation.

A second investigation of the transmission of the HOXA1 variant in a sample of multiplex and singleton families was undertaken (Devlin et al. 2002). Two hundred and thirty-one families were genotyped in total, 131 trios and the remainder were affected sib-pairs. Analyses were conducted in two diagnostic categories. A broad category consisted of individuals with a diagnosis of Autism, Asperger disorder and Pervasive Developmental Disorder resembling the diagnoses of the individuals within Ingram's sample. The narrow

diagnostic category consisted solely of those individuals with a diagnosis of autism (confirmed with the ADI-R). Analyses were performed in the sample under a number of differing models of inheritance using the FBAT programme (Horvath et al. 2001). Under the broad diagnostic category there were 95 informative transmissions. No significant transmission of either allele was observed (χ^2=0.637, p=0.524). Similarly in the narrow diagnostic no increased transmission was observed (χ^2=1.16, p=0.244). No increased transmission was observed under the recessive model. Parental transmissions did not show a significantly increased transmission of either allele from either parent. There was a non-significant trend towards transmission of the A allele to affected daughters. The analysis was also restricted to multiplex families, as Ingram's sample was comprised solely of multiplex cases. This did not reveal any association with the G allele. Restricting the analysis to those of European origin alone (also mimicking Ingram's sample) did not show any positive associations either. No sex-specific bias in transmission of the alleles was observed in the population.

An association study in the Irish sample of trios described in this work was conducted. The advantages of this sample compared to Ingram's sample were that the clinical heterogeneity was minimized through the careful phenotyping as described previously and secondly that the population was more genetically homogeneous. Detection of association over greater distances has been shown in the Irish population previously (Kendler et al. 1999). The laboratory methods for this study have been described in section 3.4.2.6.1 and 3.4.2.6.1 and statistical methods were described in section 3.4.2.7.

5.2 Results

5.2.1 Association testing of the HOXA1/ HOXB1 variants

The HOXA1 variant was an A/G substitution which was tested for association with autism in the sample. A 9bp insertion in the HOXB1 gene was tested for association with autism.

5.2.1.1 Allele frequencies and Hardy Weinberg Equilibrium (HWE)

Genotype information was obtained for 83 trios and 4 duos for the HOXA1 gene. The G allele had a frequency of 13.6%. No significant deviation from HWE was observed.

For the HOXB1 gene genotype information was available for 78 trios and 3 duos. No significant deviation from HWE was observed and the frequency of the 9bp insertion in the sample was 12.5%. The difference in numbers of trios tested for the two variants was attributable to genotyping failures.

5.2.1.2 HHRR association testing:

For the HOXA1 gene there was a small increase in transmission of the G allele. HHRR testing was non-significant for this allele (χ^2=1.36, p=0.24, RR=1.2, 0.9<RR<1.59). Testing of the HOXB1 gene showed no increase in transmission of the 9bp insertion,), (χ^2=037, p=0.54, RR=0.9, 0.64 <RR<1.27). These data are tabulated in Table 5.1

5.2.1.3 TDT association testing:

The TDT analysis showed a non-significant excess transmission of the G allele of the HOXA1 gene, (χ^2=2.08, p = 0.1996, OR = 1.6, CI 1.0 - 2.56). No increased transmission was observed for the 9bp insertion of the HOXB1 gene (χ^2=0.257, p=0.316, OR=0.84 (CI 0.53<OR<1.35) (Table 5.1).

Table 5. 1 Association testing of the HOXA1 and HOXB1 variants with TDT/ HHRR analysis.

GENE	TDT			HHRR		
	OR	χ^2	p- value	RR	χ^2	p- value
HOXA1 (A/G)	1.6 (1.0<OR<2.56)	2.08	0.1996	1.2	1.36	0.24
HOXB1 (9bp INS)	0.84 (0.53<OR<1.35)	0.257	0.316	0.9	0.37	0.54

Table 5. 2 TDT testing based on parent of origin

HETEROZYGOTE MATINGS	NO. OF MATINGS	NO. OF TRANSMISSIONS	% OF TRANSMISSIONS χ^2
HOXA1: Mother(AG)/ Father (AA)	12	8	1.3
Mother (AA)/ Father (AG)	15	11	3.3
Both parents AG	6 (12 Heterozygotes)	5	0.3
HOXB1: Mother (+/INS)/ Father (+/+)	10	7	1.6
Mother (+/+)/ Father (+/INS)	9	4	0.1
(+/INS)/ (+/ INS)	7	3	0.1

5.2.2 Transmission of variants from heterozygote parents.

As Ingram reported an increase in transmission of the HOXA1 variant from heterozygote mothers, transmission of the variants based on parent of origin in this sample was examined (Table 5.2). There were 18 heterozygote mothers where the genotypic data was complete for the family. Twelve offspring were affected heterozygotes and eight of these arose from a mating with a homozygotic (A/A) father. Of the 21 fathers where the genotypic data was complete, 15 were mated with A/A mothers and 11 of these matings resulted in offspring who were heterozygotes for the variant. In six cases mothers and fathers were both heterozygous for the variant and in five out of the six cases the offspring was also heterozygotic for the variant. This data did not support the findings of Ingram that there was increased transmission of the variant from heterozygote mothers. A similar analysis of

the data was conducted for the HOXB1 variant (Table 5.2). Once again there was no evidence of increased transmission of the variant from either parent.

5.2.3 Distribution of the variants based on gender.

The distribution of the allelic variants based on the sex of the affected individuals was examined as Ingram reported greater frequency of the HOXA1 variant in affected females. TDT analysis was conducted for male and female offspring (Table 5.3). The results of TDT for male offspring showed evidence of excess transmission of the G allele (χ^2=5.3, p=0.03). This is in direct contrast to the finding reported by Ingram, i.e., an increase in transmission to female offspring. Li et al, 2001, did not describe any increase transmission in either sex. A similar analysis with similar outcome is presented for the HOXB1 variant (Table 5.3). The evidence of greater association in males for the HOXB1 9bp insertion was not noted previously.

Table 5. 3 TDT based on sex of offspring

	OR	χ^2	P
HOXA1:	2.83	5.3	0.03
Males	(1.37<OR<5.88)		
		1.6	0.34
Females	2.33		
	(0.83<OR<6.54)		
HOXB1:	2.6	4.5	0.096
Males	(1.17<OR<5.78)		
	2	1	0.5
Females	(0.71<RR<5.62)		

5.2.4 Study design and power calculations.

An estimate of the power of this study to detect an association with the HOXA1 and HOXB1 variants was performed using the Genetic Power Calculator (Purcell et al. 2003) (http://statgen.iop.kcl.ac.uk/gpc/). Based on the observed frequencies of the variants seen in

the present study, it was estimated that a study comprised of roughly 100 trios would have 75% power to detect increased transmission at a non-significant level of 0.1.

5.3 Discussion

There is interesting evidence suggesting a role for the HOXA1 and HOXB1 variants as susceptibility genes in autism. Firstly, the observation that individuals exposed to teratogens in utero are likely to develop autism and that the occurrence of autism is associated with particular congenital anomalies suggests that a neurodevelopmental insult may occur between 20-24 days post conception, a critical time for HOXA1/B1 expression. The second line of evidence is the evidence from animal studies showing that mouse null mutants for the genes had disruption of neural crest structures, in particular disruption of cranial nerve nuclei (i.e. abducens and facial nerves) and the superior olive. Functional abnormalities of the cranial nerves are observed with higher frequency in autism (Stromland et al. 1994). Following identification of variants in the HOXA1 and HOXB1 genes, Ingram reported an association between autism and the G allele of the HOXA1 gene in a sample of related individuals with autism and autism spectrum disorders.

Two attempts to replicate these findings have been unsuccessful. As outlined above, the findings from the studies by Li and Devlin were in direct contrast to that identified by Ingram. The present study found a small non-significant trend towards transmission of the G allele of HOXA1 in individuals with autism and autism spectrum disorders. This finding, although non-significant is more supportive of the findings reported by Ingram.

As Ingram had reported an increase in transmission of the G allele from mothers to affected offspring, we examined our data for this phenomenon. In contrast to Ingram's finding and reflecting the findings of Li and Devlin, no increased transmission of the G allele was observed from mothers to affected offspring. In fact when the transmissions of alleles were examined based on the gender of the affected individual it was found that a significant increase in the transmission of both the HOXA1 G allele and the HOXB1 9bp insertion were observed in males. This is the opposite of the finding by Ingram and was not identified in the studies by Li and Devlin.

The findings in relation to the 9bp insertion of the HOXB1 gene in the present study appear to concur in all the studies reported. There appears to be no evidence suggestive of an association between this variant and autism.

The findings in the Irish sample are similar to the two attempted replication studies in that no significant trend towards increased transmission of the G allele was identified in association with autism. Unlike any of the studies, the present study found an increased transmission of the variant to male offspring. The interpretation of the data must be cautious, as there are weaknesses to this study. The first is that the sample size is relatively small. While this has the effect of reducing the power to detect an association, it also means that any association in this sample requires replication in a larger sample, particularly when the numbers of negative associations that have been reported are taken into account. The second weakness is that none of the data here has been corrected for multiple testing and therefore it is quite likely that the observed association between the G allele and autism in male offspring may have occurred by chance.

There are several differences between this study and the others in the literature that should be taken into account. Ingram's sample was comprised entirely of multiplex families, while this was a trio-based population study. Devlin reported findings in a mixed group of multiplex and singleton families. In Ingram's entirely multiplex sample, the subjects potentially represent a more genetically inherited form of autism than a population-based sample such as the one described here. Devlin's sample was comprised of 231 families, 131 trios and the remainder multiplex families. Based on the power calculations used in that study there was sufficient power to detect a susceptibility of moderate effects. Li's sample is comprised of 110 families with 230 affected individuals. Again power calculations here estimated that the sample was sufficient to detect an allele with a moderate effect. The power calculations in the present study indicated that there was insufficient power to detect an association with either variant. A meta-analysis of the data are not really possible as the samples differ in terms of their design, i.e affected sib-pairs, trios, and mixed groups, and in terms of the diagnostic criteria used. Therefore to completely exclude the HOXA1 gene as

a candidate gene for autism, a large-scale association study using a well-characterised sample would be preferable.

A population-based sample is potentially less 'enriched' for susceptibility genes compared to a sample of multiply affected families. However one of the advantages of the Irish sample is the minimization of clinical heterogeneity by careful phenotyping and the exclusion of any potential phenocopies or individuals with chromosomal anomalies. Ingram's sample included individuals with Autism, Asperger disorder and PDD-NOS. Li's sample included only those individuals meeting criteria for autism on the ADI-R. Devlin performed analyses on two diagnostic categories. The entire sample was included in the broad diagnostic category, autism, Asperger disorder and PDD-NOS. The narrow category included only individuals with an autism diagnosis.

A further advantage of the Irish sample may be relative ancestral homogeneity and simple population structure and history. It has been estimated that LD is detectable between a genetic variant and a disorder over greater distances in the Irish populaiton (Kendler et al. 1999). This is of relevance particularly if the two observed alleles are not the actual disease causing loci, but are in linkage disequilibrium with a nearby susceptibility allele. It may be the case that the variants reported here are in linkage disequilibrium with other potentially disease-causing variants. There are no reports to date of more extensive screening of either gene for new variants (e.g. the promotor region or non-coding sequences). It is also possible that for example a promotor variant might influence the expression of the gene in an as yet undetermined way.

Alternative hypotheses have not been considered by any of the authors. Ingram's hypothesis that autism is a form of neural tube deficit may be inaccurate. Exposure of rats to thalidomide and valproate in-utero has been demonstrated to result in abnormal levels of monoamines in the CNS, with elevated hippocampal serotonin and pre-frontal dopamine (Narita et al. 2002). Given serotonergic abnormalities observed in autism (discussed in chapter 1), it may well be that the clinical presentation is secondary to a differing pathophysiological mechanism entirely.

In conclusion, given that there have now been two reports of an association between autism and the G allele of the HOXA1 gene and two further reports showing no association, it seems prudent that further studies will be required to clarify the findings. Since the present study had insufficient power, further investigation of genetic association in a larger sample of Irish autistic trios is planned. Future directions should also include further characterization of the gene particularly the promotor and association testing around other variants within the gene.

Chapter 6

Characterization and fine mapping of a susceptibility region for autism on chromosome 2q

6.1 Introduction

Chapter 4 detailed the clinical description of an individual, K.M., with a chromosomal abnormality on chromosome 2q32.1-32.2. As discussed in chapter 1, this region is of particular interest in autism owing to reports from genomewide linkage studies that have highlighted putative evidence of linkage to autism in this region. Buxbaum reported evidence for linkage to autism between markers D2S335 and D2S364, located at 175.91cM and 186.21cM on chromosome 2q respectively (Buxbaum et al. 2001). The findings were more significant when the analysis was restricted to a group of individuals with phrase speech delay. The International Molecular Genetics Consortium in Autism (IMGSAC) found a Maximum Lod score (MLS) of 4.8 at marker D2S2188, located at 180.79cM (IMGSAC 2001). However, unlike the analysis performed in the Buxbaum study, the significance of these findings was not found to increase when the analyses were restricted to a narrow diagnostic category in the sample. A third genomewide linkage study reported evidence of linkage at marker D2S116, located at 198.65cM (Shao et al. 2002a). Like Buxbaum's study the significance of the findings increased when the analyses were restricted to a group with phrase-speech delay. The relative positions of these genetic markers are illustrated in Figure 6.1. Localization of genes from linkage peaks can be difficult as the regions within the peaks can be broad. Based on the data from the three studies the genetic region showing evidence for linkage represents 22.74cM. Identification of candidate genes within such a region might involve a number of approaches. Fine-mapping linkage disequilibrium (LD) experiments within a region such as this are likely to be useful to attempt to narrow the region and the number of candidate genes to be screened further. Then candidate genes lying close to linkage peaks and/ or close to markers showing evidence for association with a disorder may be selected for mutational screening.

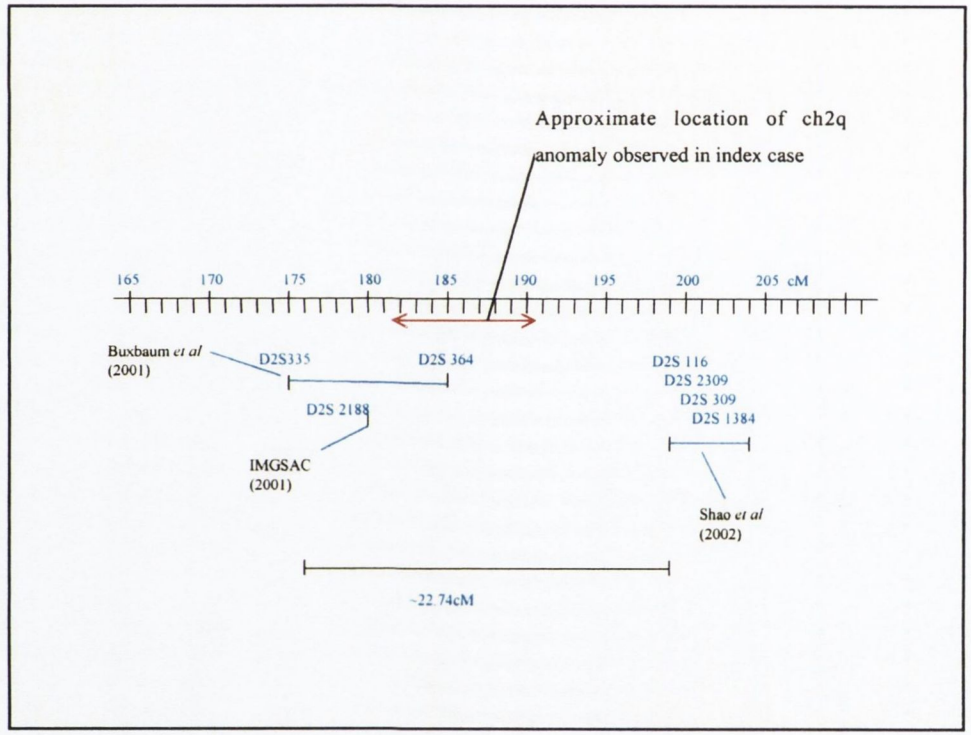


Figure 6. 1 Results of genomewide linkage studies showing evidence for linkage on chromosome 2q and relationship to chromosomal anomaly observed in index case.

Linkage and association approaches are not the only strategies that may be utilized for the purposes of gene identification. The genes for many Mendelian disorders have been identified based on the observation of cytogenetic abnormalities, which include, among others the Retinoblastoma gene (RB1) (Lee et al. 1987) the Wilms' tumor gene (WT1) (Huff et al. 1991) and the dystrophin gene in Duchenne Muscular Dystrophy (Zneimer et al. 1993). Thus the discovery of a case of autism with a chromosomal abnormality within a region showing some converging evidence for linkage was interesting and worthy of further investigation. Some of the possibilities that might arise from such a finding include the disruption of a gene or gene controlling sequence around the breakpoints of the abnormality, a haploinsufficiency of a gene within the region (particularly if the abnormality is a deletion) or an alteration in controlling factors of gene expression (particularly where the chromosomal segment is inverted or translocated).

Following the identification of this case, a plan was established to:

(1) Characterize the cytogenetic abnormality observed and define the breakpoints.
(2) Conduct a fine-mapping experiment across the region in a sample of individuals with autism in order to identify the presence of linkage disequilibrium (LD) within the region with autism.

There will follow a brief discussion of the characterization of the chromosomal abnormality followed by the results of an *in silico* candidate gene search. The results of two fine-mapping experiments will then be discussed. The first involved fine mapping of a 10Mb region on chromosome 2q using microsatellite markers from D2S2310 (at 182.4Mb) to D2S2246 (at 192.5Mb). The second investigation involved fine mapping of a smaller region of interest (approximately 850kb) using single nucleotide polymorphisms (SNPs) at an average distance of 40kb between microsatellite markers D2S2310 (at 182.4Mb) and D2S364 (at 183.2Mb).

6.2 Description of the mapped region

A collaboration was established with the National Centre for Medical Genetics where the cytogenetic testing had been undertaken to determine the extent of the apparent deletion. The first investigation was to determine the extent of the deletion using gene dosage methods. This investigation indicated that the apparent deletion was 8.5Mb however the breakpoints could not be determined. The significance of this first investigation was that the results were used to define a region for fine-mapping association experiments as described in section 6.3 and the investigations outlined here were conducted in tandem with the association experiments. The outcome of this was that the centromeric end of the breakpoint on chromosome 2q was not included in the association experiments (see Figure 6. and Section 6.4.1 for the discussion)

6.2.1 Defining breakpoints using haploid cell lines

As the microsatellite method had been unsatisfactory in the attempt to map the exact extent of the breakpoints an alternative method was sought to complete this. The initial strategy was to utilize a newly developed technique available commercially from GMP Genetics that involves synthesis of human haploid cell lines (Yan et al. 2000a). The technique involves the production of human haploid chromosomes through the fusion of human mononuclear blood cells to a novel recipient E2 cell line (mouse). The fused cells are monoallelic for human chromosomes. The company claims that the human chromosome complement is stable, with identical patterns of retention after growth for 90 generations following the initial genotyping. Using this technology it was planned to map the deletion by genotyping haploid chromosome 2 DNA. Samples from the index case, K.M. were sent to the company but the procedure failed on two occasions due to technical difficulties.

6.2.2 Characterisation of the region using comparative genome hybridization

Since the previous attempts had not yielded adequate information about the chromosomal abnormality, comparative genome hybridization (CGH) was undertaken. Fluorescently labelled case (red) and control (green) DNA were hybridized to a 600 band cytogenetic spread. The chromosome spread was counter-stained with 4',6-Diamidino-2-phenylindole dihydrochlorate (DAP-1) and evaluated. An observed decrease in

fluorescence over the region on chromosome 2q would confirm the presence of a deletion and aid the identification of the breakpoints. However no change in fluorescence was observed. This was puzzling, however detection of a deletion of approximately 8Mb approaches the limits of CGH and it was assumed that the deletion was smaller than had been previously estimated. The results of this investigation are illustrated in Figure 6.2. An alternative strategy had to be utilized to characterize the region involving Bacterial Artificial Clone (BAC)-arrayed CGH.

6.2.3 Attempt to characterize region of interest using BAC arrayed CGH

Collaboration was established with a group at the Sanger Centre who had developed a technique for BAC-arrayed CGH. This is a higher resolution procedure (1Mb) that involved the hybridization of DNA from the case onto arrayed genome-wide BAC-end PCR products. Again the procedure did not identify the presence of any deleted material. The conclusion from these CGH investigations was that the abnormality had not been a chromosomal deletion and that the observed anomaly on the cytogenetic spread may have been attributable to either an inversion or a translocation.

6.2.4 Identification of 2q translocation using flourescent-in-situ hybridisation (FISH)

A chromosome 2q fluorescent probe was applied to the cytogenetic spread. This revealed the presence of chromosome 2q on Chromosome 9. The probe was reapplied with a chromosome 9 probe and this confirmed that the abnormality was a cytogenetically visible *de novo* translocation of bands 32.3 on chromosome 2q inserted into bands q31.1-q31.2 on chromosome 9 (46, XY, ins (9;2)(q31.1;q31.2q31.3). The chromosome 2q (red) and 9q (green) probes are illustrated in Figure 6.3. Figure 6.4a shows the original cytogenetic spread that had been judged to be a deletion and Figure 6.4b is a comparison between chromosomes 2 and 9 showing the presence of an inserted band on chromosome 9q.

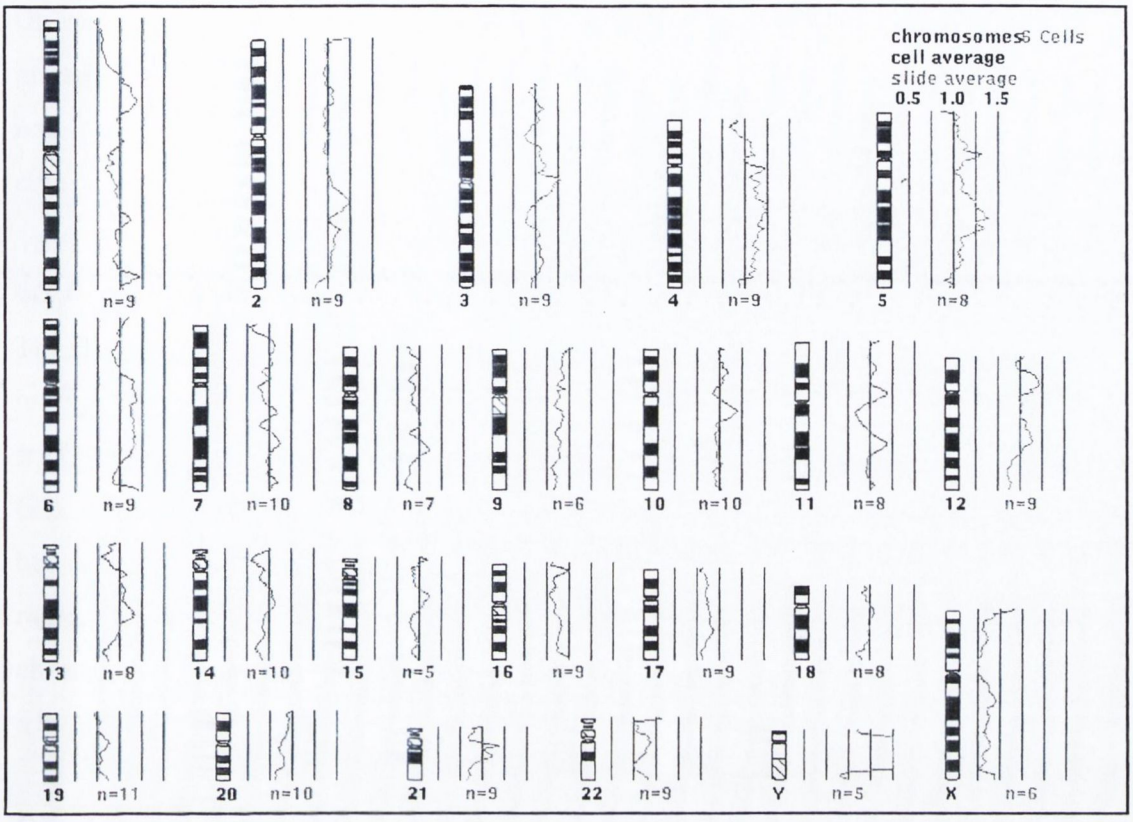


Figure 6.2 Comparative genome hybridisation showing no evidence of deletion on chromosome 2q31

(Graphs to side of chromosomes show the presence of all chromosomal material. A significant deviation in a graph would indicate a deletion (if negative) and an insertion (if positive))

6.2.5 Defininition of the breakpoints using BAC clones using flourescent-in-situ-hybridisation (FISH) walking.

Having identified the translocated portion of chromosome 2q on chromosome 9q, the deletion breakpoints on chromosome 2q and the insertion breakpoints on chromosome 9 had to be defined. This involved the use of FISH walking with labeled BAC clones. Labeled clones were hybridized to the cytogenetic spread commencing at a point on chromosome 2q that was known to lie outside the deleted region. BAC clones overlapped in a tile-like fashion. This enabled the identification of the breakpoint to within 10-20kb.

On chromosome 9q the method was used in reverse, commencing at the centre of the inserted portion of chromosome 2q and walking out to the insertion breakpoints. The importance of this investigation is to discover if a gene on either chromosome 2q or chromosome 9q was disrupted by the breakpoints.

6.2.6 Chromosomal breakpoints on chromosome 2q:

The BAC clones utilized to map the extent of the deletion on chromosome 2q are illustrated in Figures 6.5a and 6.5b. At the 5' end BAC RP11-33 I19 is not deleted, while the adjacent BAC RP11-122 M5 is deleted. At the 3' end of the region on chromosome 2q RP11-24 G21 is deleted and RP11-270 G18 is not deleted. There is approximately an 11kb overlap between these BACs which is likely to contain the breakpoint. The size of the deleted region on chromosome 2q is 8.75Mb. No gene at the 5' end of the deleted region on chromosome 2q was observed to be disrupted by the breakpoint. At the 3' end of the region a hypothetical gene, AK022607 was disrupted.

On chromosome 9q the inserted chromosomal material was found to be orientated inversely. The translocation breakpoint has been mapped to between RP11-52 E12 and RP11-80 N14 (Figure 6.6). This currently represents a region of 3.2Mb and the exact breakpoint needs localization within this region.

6.2.7 *In silico* characterization of candidate genes within the chromosome 2q region

Based on the results of these investigations, a review of the publicly available databases was undertaken to identify candidate genes within the region. The databases utilized included the University of Santa Cruz (UCSC) Genome Browser Gateway (http://genome.ucsc.edu/), Ensembl (http://www.ensembl.org/Homo_sapiens/), NCBI Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM) and Genecards (http://bioinformatics.weizmann.ac.il/cards/). Known genes within the region are detailed in Table 6.1. While no known gene appeared to be disrupted, a human mRNA transcript with a hypothetical protein product was disrupted at the breakpoint at the 3' end. This was

AK022607, 'Weakly Similar to Vegetable Incompatibility Protein' (HET-E1). It spans 34.9kb and has 22 exons.

As yet no known gene has been identified at the breakpoints on chromosone 9q31.1q31.2 as the region has been mapped to within 3.5Mb. The known genes contained within this region are indicated in Table 6.2.

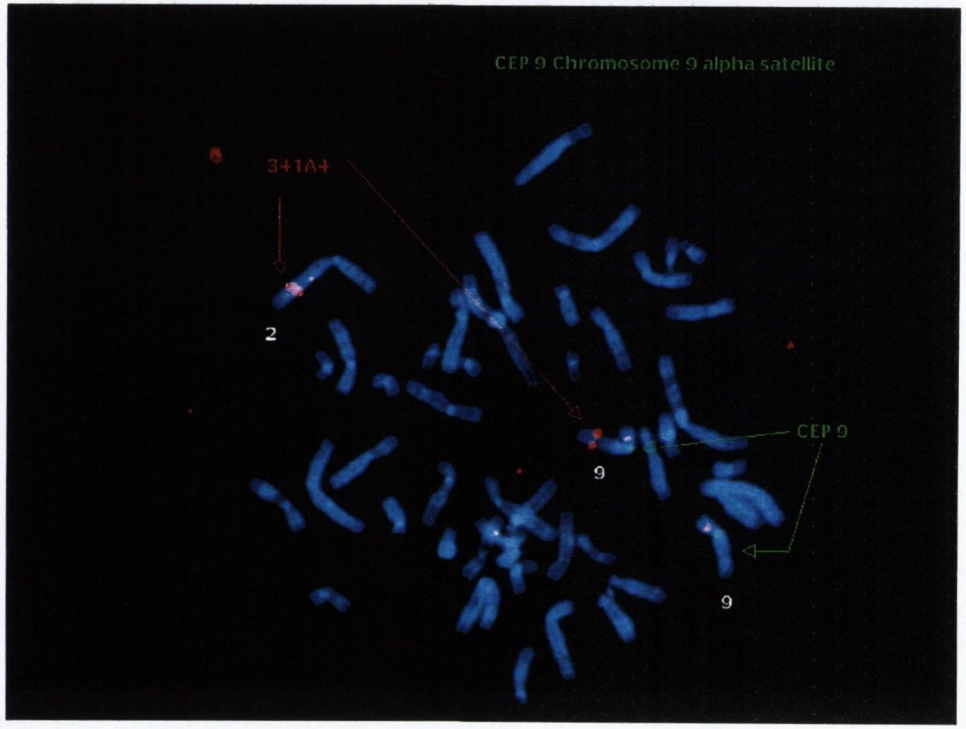


Figure 6.3 Flourescent in situ hybridization (FISH) showing the presence of a chromosome 2q probe (341A4) translocated to chromosome 9q.

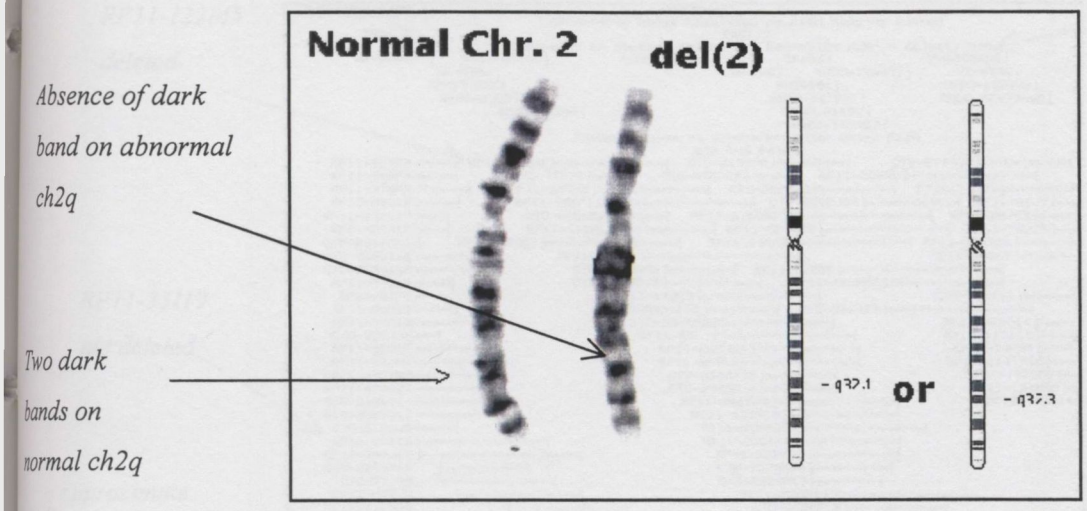


Figure 6.4 a Cytogenetic Spread from index case showing evidence for deletion of Ch2q31.1-31.2

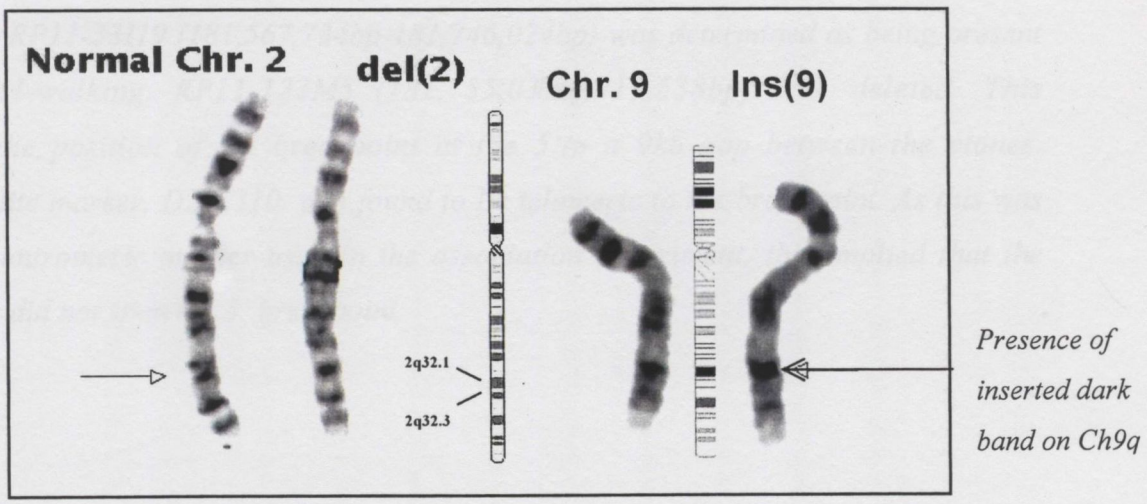


Figure 6.4 b Cytogenetic spread showing the presence of translocated chromosome 2q32.1-32.2 on Chromosome 9q31.1-31.3

Figure 6.5 a Position of BAC clones at 5' end of chromosome 2q region indicating the 5' breakpoint.

BAC clone RP11-33I19 (181,567,734bp-181,746,024bp) was determined as being present using FISH-walking. RP11-122M5 (181,755,037bp-181,911,538bp) was deleted. This localised the position of the breakpoint at the 5'to a 9kb gap between the clones. Microsatellite marker, D2S2310, was found to lie telomeric to the breakpoint. As this was the most centromeric marker used in the association experiment, this implied that the experiment did not span the 5' breakpoint.

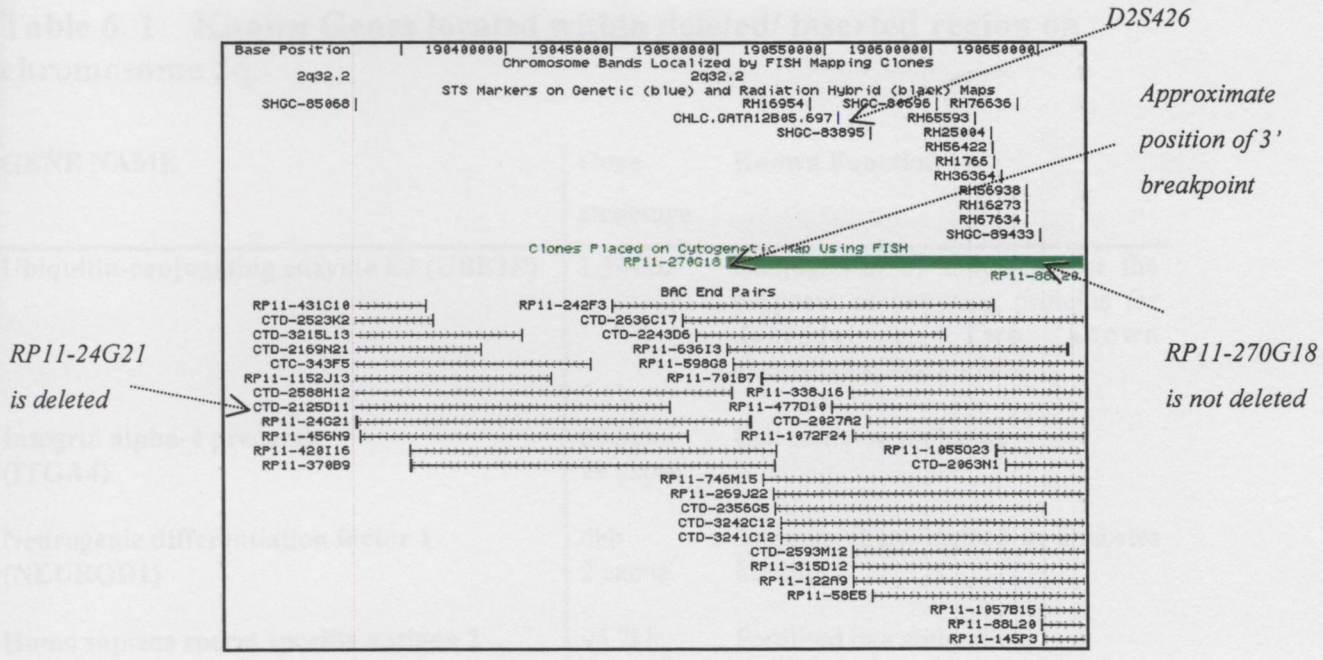


Figure 6.5 b Positions of BAC clones at 3' end of the chromosome 2q region indicating the 3' breakpoint.

The RP11-24G21 clone, at 190350000bp-190502000, indicated by the arrow on the left of the diagram was found to be deleted. The RP11-270G18 clone, which ends at 190501000bp (approx), shown on the right of the diagram in green was not deleted. This indicates that the position of the breakpoint is located in the region of approximately 11kb where these clones overlap (190501000bp – 190502000bp). The microsatellite marker, D2S426, indicated by the arrow on the left of the diagram is physically located on the undeleted clone, RP11-270G18. This marker was genotyped in the microsatellite-based association experiment, which thus spanned the breakpoint at the telomeric end of the 2q region.

Table 6. 1 Known Genes located within deleted/ inserted region on chromosome 2q.

GENE NAME	Gene structure	Known Function
Ubiquitin-conjugating enzyme E2 (UBE2E)	1.394kb	Conjugation of ubiquitin for the purposes of targeting proteins for degradation. Two known alternative splice variants
Integrin alpha-4 precursor (ITGA4)	80.5kb 29 exons	Cell adhesion molecule
Neurogenic differentiation factor 1 (NEUROD1)	4kb 2 exons	Variants characterized in Diabetes Mellitus
Homo sapiens sperm specific antigen 2 (SSFA2)	95.2kb 5 exons	Fertilised ova antigenicity
Homo sapiens protein phosphatase 1 regulatory subunit 1A	1.18kb	Protein phosphatase inhibitory function. Expressed in skeletal muscle
Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1A (PDE1A)	96kb 19 exons	Hippocampal derived
Endoplasmic Recticulum-resident protein (Erdj5)	64kb 25 exons	Expressed at high levels in pancreas and testis. Possible role in protein transport within the cell
Frizzled-related protein precursor	34kb 7 exons	Possible role in patterning of body axis in vertebrates. Interacts with WNT family. Postulated role as tumor suppressor.
Nck-associated protein 1	114.7kb 32 exons	Role in apoptosis/ Alzheimer disorder
Dual specificity protein phosphatase 19	22.9kb 5 exons	Protein Product, function inferred. Stress-activated protein kinase pathway-regulating phosphatase 1
Nucleoporin (Nup35)	62.6kb 9 exons	Protein product, function inferred. Mitotic Phosphoprotein

(continued)

Table 6.1 (contd.)

GENE NAME	Gene structure	Known Function
Uncharacterized hypothalamus protein (HT010)	23.9kb 10 exons	Erythropoietin 4 immediate early response. Ten alternative splice variants. Widely expressed.
PRO2710.	1.5kb	Hypothetical gene, unknown function.
Integrin alpha-V precursor (ITGAV)	91.6kb 31 exons	Vitronectin receptor. Vascular development.
Hypothetical protein (FLJ34104)	23kb 10 exons	Protein product, function inferred. Erythropoietin 4 immediate early response.
Calcitonin gene-related peptide type 1 receptor precursor	105.9kb 16 exons	Expressed in cerebellum/ CNS.
Tissue factor pathway inhibitor precursor	36.8kb 6 exons	Role in coagulation.
CED-6 protein	26.5kb 5 exons	Apoptosis.
Homo sapiens disrupted in renal cancer protein (DIRC1)	57.2kb 2 exons	Disrupted in renal cancer. Protein function unknown.
Alpha 1 Type III collagen (COL5A2)	38.4kb 49 exons	Ehler-Danlos Syndrome.
Collagen alpha 2 (COL2A)	148kb 54 exons	Ehler-Danlos Syndrome.
Hypothetical protein FLJ12519	34.1kb 22 exons	Protein product, function unknown.
Iron-regulated transporter IREG1 (SLC40A1)	22.9kb 9 exons	Role in haemachromatosis Alternative splice variants.

Table 6. 2 Known genes contained within 3.2Mb region on chromosome 9q31

GENE NAME (ID)	Gene structure	Known Function
Inversin (INVS)	202.6kb 19 exons	Renal tubular development and function. Left-right axis determination. Mutations associated with nephronophthisis type II.
Hypothetical protein (FLJ20287)	51.6kb 15 exons	Protein product, function unknown.
Transmembrane protein with EGF-like and two follistatin-like domains 1 (TMEFF1)	105kb 10 exons	Possible tumor suppressor gene in brain cancer.
Hypothetical protein (FLJ20300)	140.9kb 9 exons	Member of the phosphatidic acid (PA) phosphatase-related phosphoesterase family. Protein product, function unknown.
Bile acid Coenzyme A: amino acid N-acyltransferase (BAAT)	24.3kb 4 exons	Catalyzes the transfer of the bile acid moiety from the acyl-CoA thioester to either glycine or taurine.
Mitochondrial ribosomal protein L50 (MRPL50)	9.4kb 2 exons	Protein synthesis within the mitochondrion.
Zinc Finger Protein (ZNF189)	12.6kb 4 exons	Protein product, function unknown.
Aldolase B, fructose-bisphosphate (ALDOB)	15.2kb 9 exons	Catalyzes the reversible conversion of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. Associated with hereditary fructose intolerance.
Hypothetical protein (MGC12992)	58.9kb 5 exons	Protein product, function unknown.
Ring finger protein (RNF20)	29kb 20 exons	Ubiquitination of histones.
Glutamate receptor, ionotropic, N-methyl-D-aspartate 3A. (GRIN3A)	169kb 15 exons	Glutamate regulated ion channel, NMDA receptor.
Protein phosphatase 3, regulatory subunit B (PP3R2)	34kb 1 exon	Calcium ion binding, expressed in cytoplasm.
Cylicin, basic protein of sperm head cytoskeleton 2. (CYLC2)	23.2kb 11 exons	Expressed in testis, part of cytoskeletal calyx of mammalian spermheads.

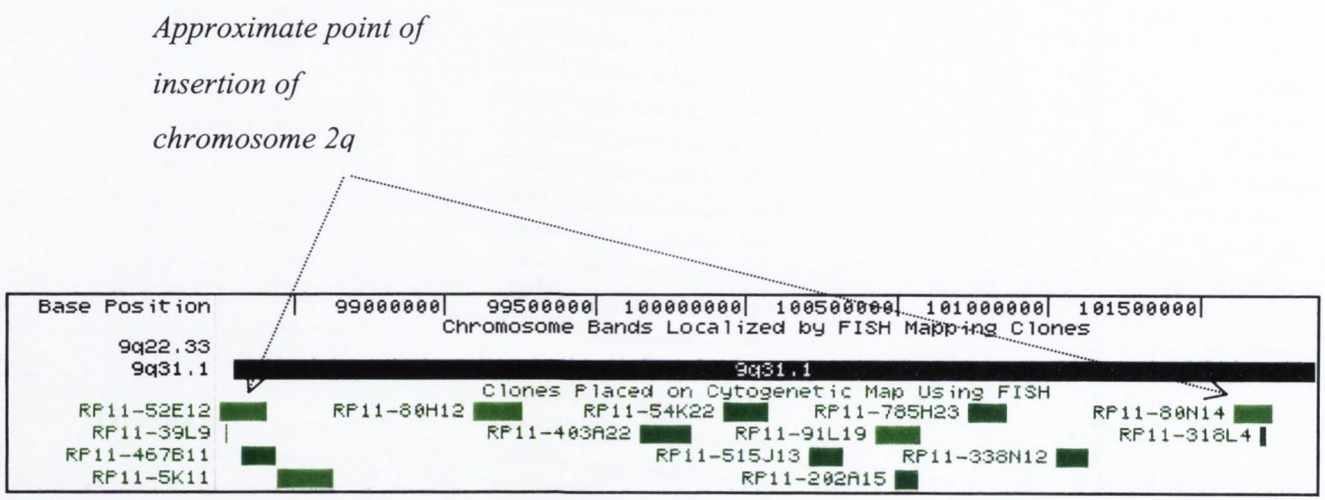


Figure 6.6 The approximate point of insertion of the chromosome 2q translocation on chromosome 9q based on estimations from FISH clones.

The point of insertion of the chromosome 2q on chromosome 9q has been localised to between clones RP11-52E12 (98,250,207bp-98, 410,010bp) and RP11-80N14 (101,609,007bp-101,736,685bp).

6.3 Fine mapping of the translocated region using microsatellite markers

6.3.1 Selection of markers for LD mapping

Due to the difficulties in determining that the chromosomal anomaly on chromosome 2q was a translocation, a fine-mapping experiment was undertaken prior to completion of this work. Thus the region that was defined for fine mapping with microsatellite markers was based on the extent of the perceived 'deletion'. Figure 6.7 shows the microsatellite markers selected for the linkage disequilibrium fine mapping experiment. The position of the 5' translocation breakpoint has been illustrated in Figure 6.6a and the position of marker D2S2310 is indicated in this illustration. From this it can be observed that D2S2310, which represented the first marker at the 5' end of the microsatellite-based linkage disequilibrium experiment, lies within the translocated region. As a result no marker at the 5' end of the deletion has been genotyped that lies outside of the breakpoints on chromosome 2q.

Twenty-two markers were selected from the publicly available genetic maps, including the Marshfield Chromosome 2 Sex Averaged Map, The Whitehead Chromosome 2 Radiation Hybrid Map and the Genethon Chromosome 2 map. The maps were all accessed through the University of California, Santa Cruz Human Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway). Markers with known heterozygosity were preferentially selected for genotyping across the region. However due to low numbers of markers in the region with published heterozygosity scores, several were selected where the heterozygosity was unknown. In order to determine heterozygosity, these markers were genotyped in a sample of controls prior to genotyping of the autism sample. Figure 6.7 also illustrates the physical distance between microsatellite markers. While the average distance is approximately 500-600kb it can be seen that parts of the region have a sparse distribution of markers. This applies in particular to the region around D2S270. The closest marker to it in the region was D2S152, however this marker was homozygous in all samples despite the use of two sets of primers.

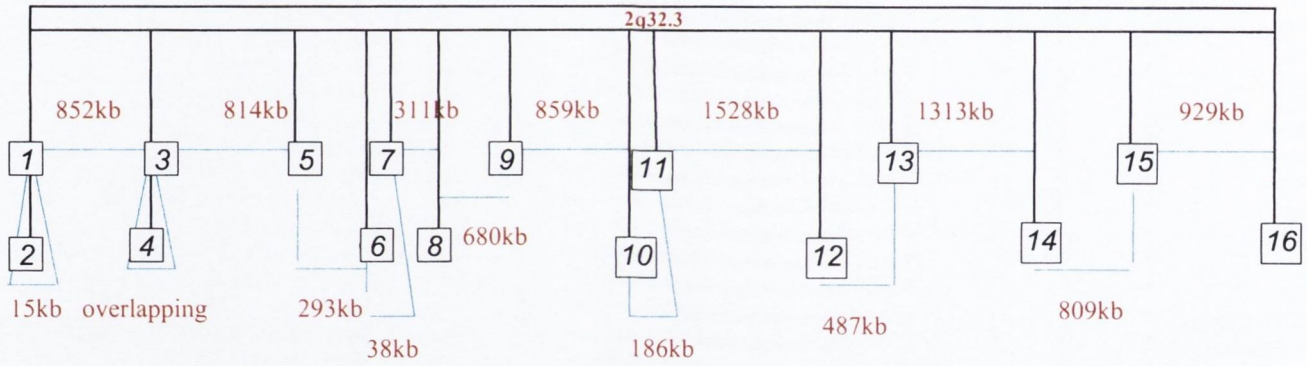


Figure 6.7 Positions and physical distances of microsatellite markers used in LD fine mapping experiment on chromosome 2q.

Markers used in the microsatellite association experiment are numbered 1-16.

1=D2S2310	*5= D2S350*	*9= D2S1383*	*13= D2S103*
2=D2S2077	*6= D2S2273*	*10=D2S1787*	*14= D2S152*
3=D2S364	*7=D2S2281*	*11=D2S1361*	*15=D2S426*
4=SHGC1659	*8=D2S2366*	*12=D2S270*	*16=D2S2262*

Thus D2S270 is located 1528kb from the next marker upstream and 1800kb from the next marker downstream. Of the original twenty-two markers, 18 were genotyped in the autism sample and fifteen of these provided informative genotypic information. Allele frequencies were calculated in the parents and no deviation from Hardy-Weinberg Equilibrium was observed. All markers included in the analysis for fine mapping were genotyped in 96 autism trios.

6.3.2 Microsatellilte marker association analysis in autism trios

The fifteen markers genotyped across the region of interest were analysed for association using the ETDT program (Sham and Curtis 1995). The results are presented in Table 6.3.

Table 6. 3 Results of allelewise ETDT association testing for all microsatellite markers used in fine mapping experiment on chromosome 2q

ETDT Association Testing for Microsatellite Markers

	χ^2	df	p value	Empirical p
D2S2310	10.04	9	0.35	0.49
D2S2077	5.73	2	0.017	0.013
D2S364	8.02	9	0.53	0.66
SHGC1659	13.04	19	0.11	0.17
D2S350	4.41	7	0.73	0.832
D2S2273	9.75	9	0.37	0.47
D2S2281	6.71	6	0.35	0.466
D2S2366	8.58	5	0.12	0.12
D2S1383	5.16	6	0.52	0.66
D2S1787	25.35	19	0.15	0.326
D2S1361	5.68	7	0.58	0.664
D2S270	11.55	8	0.17	0.221
D2S103	5.37	6	0.50	0.496
D2S426	3.93	3	0.27	0.28
D2S2262	16.40	11	0.13	0.21

Shaded cells indicate markers showing evidence of association with autism

Analysis of marker D2S2077 revealed evidence of significant association with autism (Table 6.4). This was a bi-allelic marker located close to the 5' end of the mapped region. It had two alleles of 144bp and 148bp, which are identified as alleles 1 and 2 respectively. Allele 2 was observed to be transmitted with significantly greater frequency (χ^2=5.8, p=0.013, OR = 1.75). The frequency of this allele in parents was 62.3%.

Table 6. 4 TDT testing of microsatellite marker D2S2077 in autism sample

	D2S2077	
Alleles:	1	2
Transmitted	28	49
Not transmitted	49	28
χ^2	5.7	
P-value	0.017	
Odds Ratio	1.75	
Parental Transmissions:		
Mothers (χ^2, p-value)	2.68, p=0.1	
Fathers (χ^2, p-value)	3.13, p=0.07	

Some association was seen around marker D2S270, which was a nine-allele marker with allele sizes ranging from 359bp-375bp (numbered 1-9 in Table 6.6). The allele-wise TDT was not significant for this marker although a small trend towards association was observed (χ^2= 11.55, 8 df, p = 0.17). However a genotype-wise association was detected (χ^2=43.48, 19 df, p=0.001). This marker is located closer to the 3' end of the region and for the reasons discussed above, there are larger intermarker distances between this marker and surrounding microsatellites. The transmissions for individual alleles for both markers are shown in Table 6.5. Analysis of individual allele transmissions for the D2S270 marker show that alleles 2, 3 and 9 appear to be transmitted more frequently than might be expected by chance (although not significantly). These were low frequency markers, with frequencies in parents of 1.2%, 7% and 6% respectively.

Table 6. 5 ETDT testing of microsatellite marker D2S270

D2S270

ETDT Analysis

	χ^2 (df)	P value
Allele-wise	11.55 (8)	0.17
Genotype-wise	43.48 (19)	0.001
Goodness of Fit	31.93 (11)	0.0008

Table 6. 6 ETDT association testing of individual alleles of microsatellite marker D2S270

D2S270

Alleles

	1	2	3	4	5	6	7	8	9
Transmitted	1	4	14	30	27	14	29	1	9
Not transmitted	0	0	7	34	35	17	28	2	6
χ^2			2.333	0.25	1.032	.290	0.018	0	0.6
Uncorrected P-value			0.126	0.6171	0.3097	0.5900	0.8946		0.4386

Shaded alleles indicate those alleles that appear to be over transmitted

To minimize loss of information as a result of missing parental genotypes, markers were also tested for association using the haplotype-based haplotype relative risk (HHRR). For multi-allelic markers, only those markers were tested where there was evidence that one of the alleles was overtransmitted. HHRR was conducted by comparing the transmission of the overtransmitted allele with the combined transmissions of the remaining allele. Marker D2S2077 showed significant transmission of the 2 allele based on the HHRR (χ^2=5.6, p=0.018, RR=1.3, 1.04<RR<1.66, AF=0.16). No significant transmission was detected for the remaining markers. These results are shown in Table 6.7.

Table 6. 7 Association testing of microsatellite markers using HHRR

Marker	χ^2	p value	Relative Risk	Confidence intervals
D2S2310	3.09	0.08	1.36	1.01<RR<1.81
D2S2077	5.6	0.018	1.3	1.04<RR<1.66
D2S350	0.78	0.38	1.12	0.86<RR<1.46
D2S2273	1.31	0.25	1.17	0.91<RR<1.5
D2S2366	3.26	0.07	1.23	0.99<RR<1.52
D2S1383	0.54	0.46	1.11	0.85<RR<1.45
D2S270	1.49	0.22	1.28	0.89<RR<1.83
D2S103	2.11	0.15	1.27	0.95<RR<1.7
D2S426	3.11	0.07	1.22	0.99<RR<1.51
D2S2262	0.4	0.53	1.07	0.87<RR<1.32

Shaded cells indicate marker that was found to be significantly over-transmiitted using HHRR

6.3.3 Measure of LD (D') between markers

An indication of Linkage Disequilibrium (LD) between markers was obtained by calculating D' values between all markers using the Graphical Overview of Linkage Disequilibrium (GOLD) program (Abecasis and Cookson 2000). A graphical representation of the LD in the region is illustrated in Figure 6.8. As expected from the physical distances separating markers, the D' values for intermarker LD are small.

6.3.4 Definition of region for single nucleotide polymorphism (SNP) mapping

Based on the association findings around microsatellite markers in the investigations described above, two regions were highlighted for further follow up. Follow up of the region around the D2S2077 is described in the following section. The second was the region around marker D2S270, a work currently in progress.

	D2S2310	D2S2077	D2S364	SHGC1659	D2S350	D2S2273	D2S2281	D2S2366	D2S1383	D2S1787	D2S1361	D2S270	D2S103	D2S426
D2S2077	0.678													
D2S364	0.145	0.092												
SHGC1659	0.205	0.114	0.715											
D2S350	0.132	0.224	0.176	0.189										
D2S2273	0.149	0.112	0.192	0.177	0.275									
D2S2281	0.113	0.122	0.221	0.189	0.173	0.343								
D2S2366	0.16	0.148	0.2	0.161	0.109	0.169	0.168							
D2S1383	0.143	0.089	0.195	0.131	0.078	0.145	0.111	0.144						
D2S1787	0.222	0.095	0.24	0.249	0.173	0.188	0.217	0.192	0.199					
D2S1361	0.199	0.111	0.15	0.245	0.303	0.223	0.173	0.159	0.233	0.269				
D2S270	0.231	0.197	0.223	0.229	0.152	0.189	0.212	0.209	0.167	0.196	0.235			
D2S103	0.162	0.09	0.192	0.204	0.223	0.157	0.153	0.111	0.207	0.266	0.215	0.23		
D2S426	0.15	0.233	0.148	0.171	0.329	0.142	0.115	0.124	0.131	0.165	0.145	0.181	0.238	
D2S2262	0.216	0.155	0.125	0.149	0.122	0.148	0.158	0.143	0.132	0.236	0.2	0.17	0.177	0.247

D' = 0.6 – 0.79

D' = 0.4 – 0.59

D' = 0.2 – 0.39

D' < 0.2

Figure 6.8 LD relationships between microsatellite markers on chromosome 2q

6.3.5 Selection of SNPs for linkage disequilibrium fine mapping.

Twenty-six SNPs were selected at average intervals of 30-40kb across a 850kb region that spanned from D2S2310 to D2S364. Markers and surrounding sequence were obtained from the publicly available databases of SNPs the SNP Database (dbSNP) (http://www.ncbi.nlm.nih.gov/SNP/) and the SNP Consortium (TSC) (http://snp.cshl.org/). Details of the SNPs and surrounding sequences were sent to the British Medical Research Council (MRC) Geneservice where Taqmann assays were optimized for genotyping. Of the original twenty-six SNPs only 19 were suitable for optimization, owing to poor sequence data surrounding the SNP or the presence of a further SNP within the surrounding sequence. The SNPs used in the fine mapping experiment are shown in Table 6.8. This table also includes the flanking microsatellite markers (D2S2310, D2S364 and D2S2077) and the positions of SNPs and microsatellite markers in the current maps (University of Santa Cruz Genome Browser – July 2003) and the physical distance separating markers.

6.3.6 SNP marker association testing using the transmission disequilibrium test and haplotype-base haplotype relative risk.

Genotype distribution in parents showed no significant difference from those expected according to Hardy-Weinberg equilibrium. The results of TDT analysis are shown in Table 6.9 and include the results for the flanking microsatellite markers and marker D2S2077 that show evidence for association in the autism sample. There was no significant association detected with any of the SNPs in the autism sample. However, one SNPs (SNP 4) showed a non-significant trend towards increased transmission of allele 2 (the c allele) (χ^2=3.3, p=0.07, OR=1.68). Two further SNPs, SNP 26 and SNP 3 showed a non-significant trend towards increased transmission in the autism sample (SNP 26: χ^2=2.7, p=0.1, OR=1.47, SNP 3: χ^2=2.9, p=0.09, OR=1.63). Association testing was also conducted using HHRR and this data are shown in Table 6.10. Analysis of the SNP 4 marker showed significant evidence of association to autism (χ^2=5.16, p=0.023, RR=1.48, 1.01<RR<2.16). The association testing for SNP 26 and SNP 3 remained non-significant however SNP 26 continued to show a trend towards significance (χ^2=2.98, p=0.08, RR=1.21, 0.98<RR<1.5).

Table 6. 8 Physical distances between microsatellite and SNP markers in region subjected to linkage disequilibrium mapping on chromosome 2q.

MARKER	SNP Consortium Reference	dbSNP Reference	Base Change	Allele 1	Allele 2	Position (bp) (UCSC July 2003)	Distance separating markers (kb)
D2S2310	-	-	-	-	-	182130476	
D2S2077	-	-	-	-	-	182146268	15.8
SNP 26	TSC0289130	rs976384	a/t	a	t	182156774	10.5
SNP 1	TSC0261664	rs959382	a/c	a	c	182170825	14.1
SNP 3	TSC1049286	rs2035207	c/t	c	t	182253484	82.7
SNP 4	TSC0675369	rs155149	a/c	a	c	182293670	40.2
SNP 6	TSC0235369	rs921257	c/t	c	t	182358590	64.9
SNP 8	TSC0065947	rs763860	a/g	a	g	182386947	28.4
SNP 9	TSC0531811	rs1823335	a/c	a	c	182430564	43.6
SNP 10	n/a	rs732483	a/g	a	g	182475363	44.8
SNP 11	TSC0033824	rs1000764	c/t	c	t	182518536	43.2
SNP 14	n/a	rs1528027	c/t	c	t	182637361	38.4
SNP 15	TSC0512025	rs1365806	a/g	a	g	182669175	31.8
SNP 16	TSC0634796	rs1427332	a/g	a	g	182679295	10.1
SNP 17	TSC0326375	rs1157595	a/g	a	g	182713591	34.3
SNP 19	n/a	rs2288330	g/t	g	t	182748017	34.4
SNP 20	TSC0188915	rs908391	a/g	a	g	182783803	35.8
SNP 21	TSC1089774	rs1196290	a/g	a	g	182817229	33.4
SNP 22	TSC0976902	rs1921143	a/g	a	g	182878498	61.3
SNP 25	TSC0644900	rs1196065	g/t	g	t	182996603	118.1
D2S364	-	-	-	-	-	182998556	2.0

UCSC- University of Santa Cruz Genome Browser Gateway

Table 6.9 TDT and ETDT association testing of SNPs and MS markers in region subjected to SNP based linkage disequilibrium mapping.

TDT/ ETDT association analysis SNPs and surrounding MS markers

	χ^2	p value	Odds Ratio
D2S2310	10.04 (9 df)	0.35	
SNP 19	1.286	0.257	1.54
D2S2077	5.73 (2)	0.017	1.75
SNP 26	2.649	0.1037	1.47
SNP 1	0.13	0.718	1.09
SNP 3	2.88	0.0897	1.63
SNP 4	3.314	0.069	1.68
SNP 6	1.067	0.302	1.3
SNP 8	0.023	0.879	1.03
SNP 9	0.057	0.812	1.08
SNP 10	0.077	0.782	1.16
SNP 11	n/a	n/a	n/a
SNP 14	n/a	n/a	n/a
SNP 15	0.474	0.491	1.375
SNP 16	0.818	0.366	1.44
SNP 17	1.991	0.275	1.24
SNP 20	0.013	0.908	1.027
SNP 21	n/a	n/a	n/a
SNP 22	1.052	0.305	1.26
SNP 25	n/a	n/a	n/a
D2S364	8.02 (9 df)	0.53	

Shaded cells indicate markers that showed evidence of over-transmission in the autism sample.

n/a – Markers that were not informative for association testing due to a low frequency of the minor allele.

Table 6.10 Haplotype-based Haplotype Relative Risk (HHRR) association testing of SNPs and MS markers in region subjected to SNP based linkage disequilibrium mapping.

HHRR analysis of SNPs and surrounding MS markers in the autism sample

	χ^2	p value	Relative Risk	Confidence Intervals
D2S2310	3.09	0.08	1.36	1.01<RR<1.81
SNP 19	1.07	0.301	1.18	0.88<RR<1.59
D2S2077	5.6	0.018	1.3	1.04<RR<1.66
SNP 26	2.98	0.08	1.21	0.98<RR<1.5
SNP 1	0.26	0.609	1.09	0.85<RR<1.34
SNP 3	2.01	0.156	1.23	0.91<RR<1.68
SNP 4	5.16	0.023	1.48	1.01<RR<2.16
SNP 6	0.41	0.524	1.08	0.85<RR<1.37
SNP 8	0.03	0.86	1.02	0.8<RR<1.3
SNP 9	0	1	1	0.68<RR<1.47
SNP 10	0.08	0.77	1.08	0.65<RR<1.81
SNP 11	2.05	0.15	1.52	1.0<RR<2.3
SNP 13	n/a	n/a	n/a	n/a
SNP 14	n/a	n/a	n/a	n/a
SNP 15	0.5	0.48	1.2	0.7<RR<2.05
SNP 16	0.85	0.36	1.39	0.63<RR<3.06
SNP 17	0.46	0.49	1.08	0.87<RR<1.35
SNP 20	0.02	0.89	0.99	0.79<RR<1.23
SNP 21	1	0.316	2.01	1.81<RR<2.22
SNP 22	1.25	0.26	1.14	0.91<RR<1.42
SNP 25	0.02	0.896	1.02	0.79<RR<1.32
D2S364	-	-	-	-

Shaded cells indicate those markers that were transmitted significantly in the autism sample

The relationship of the SNPs and surrounding microsatellite markers to genes and predicted genes in the region is shown in figure 6.10.

6.3.7 LD relationships between SNPs and MS markers and haplotype definition.

6.3.7.1 LD relationships using Graphical Overview of Linkage Disequilibrium (GOLD)

A measure of LD between SNPs and the surrounding microsatellite markers was conducted using GOLD (Abecasis and Cookson 2000). This program generates D' values of LD based on the genotypes observed in parental data and uses information from both SNPs and multi-allelic microsatellite markers. The LD relationships between SNPs and microsatellite markers is graphically illustrated in figure 6.10.

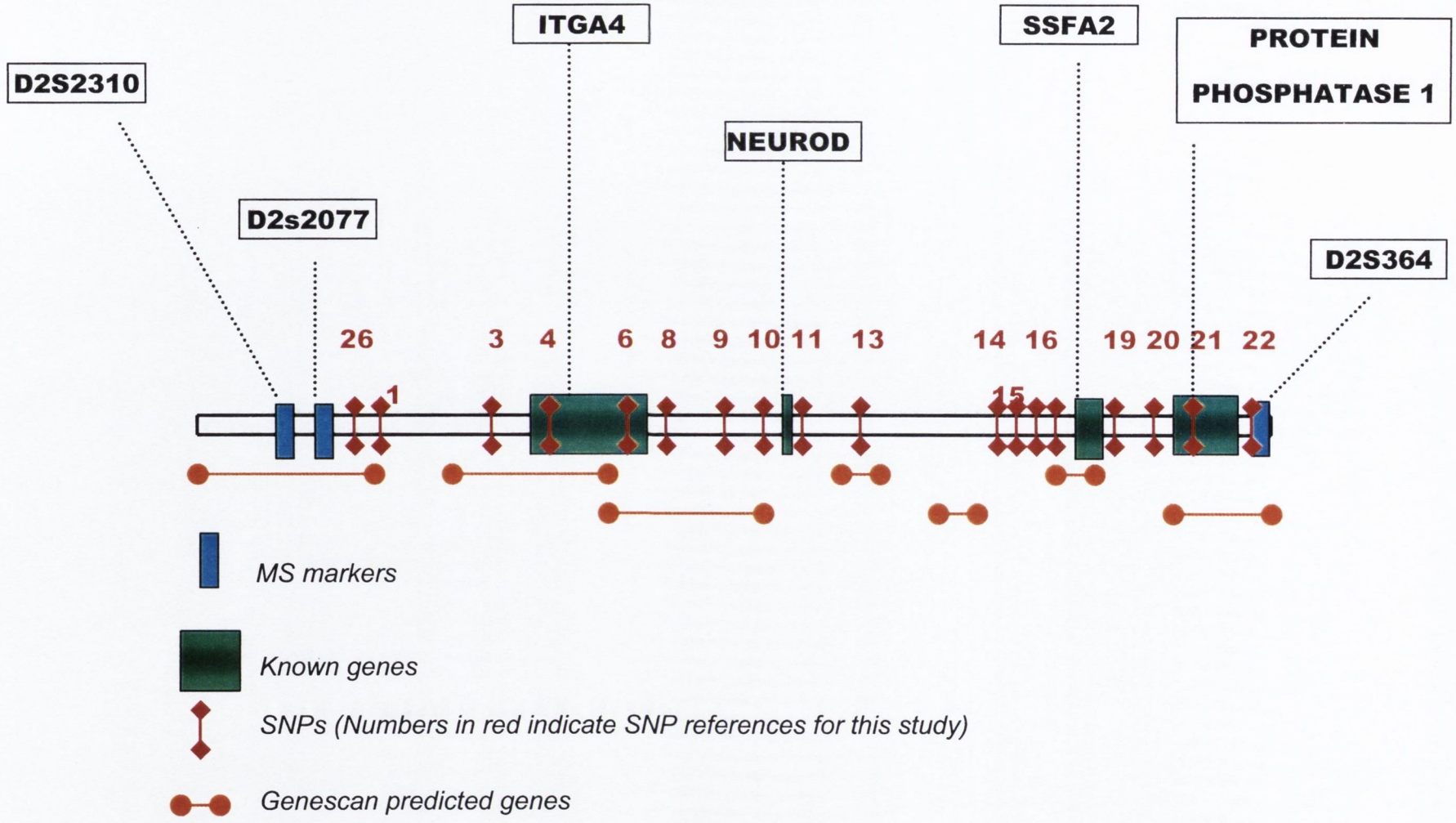
6.3.7.2 LD and haplotypic structure using Haploview

The Haploview program provides information about underlying LD and haplotype structures based on D' values between markers. Information from parental genotypes for SNPs and microsatellite marker (D2S2077, as it was a biallelic marker) was analysed using the Haploview program to analyse the underlying LD block structure in the region. A graphical representation of the LD in the region is shown in figure 6.11a. This figure also indicates the haplotypic relationships of SNPs in the region and provides a schematic representation of the physical relationships of the markers. From the figure, it can be seen that there are three discrete blocks (in red) within the region comprised of D2S2077/ SNP26, SNP4/6/8/9/10 and SNP14/15/16/17.

The haplotypes that were generated by Haploview are shown in Figure 6.11b. This shows the SNP alleles within each block. Haplotype blocks linked by dark lines indicate those that were observed to be in LD with a frequency greater than 10%, indicating the presence of a 'two-block haplotype'. The SNPs that are marked in red indicate 'tagging' SNPs, i.e. the SNPs that are most informative markers for that block. The haplotypes generated showed the relationship between a haplotype comprised of D2S2077/ SNP26/ SNP1 and SNP3 with a haplotype comprised of SNP4/ SNP6 which are all in strong LD. The three commonest haplotypes for the D2S2077/ SNP26/ SNP 1 combination are all in strong LD with two of the observed haplotypes for SNP4/ SNP6/ SNP8.

A second two-block haplotype relationship generated is the relationship between D2S2077 alone with the SNP4/ SNP6/ SNP8 haplotype and a haplotype block comprised of SNP26/ SNP1/ SNP3/ SNP4/ SNP6/ SNP8. This shows that D2S2077 is in LD with both of these haplotype blocks. It is probably a reasonable assumption that the entire region spanning from D2S2077 to SNP9 is in LD. A better picture of the underlying haplotypic structure would probably be gained be genotyping SNPs in the region at higher density (e.g. every 5-10kb).

Figure 6.9 Relationship between MS Markers, known genes, predicted genes and SNPs in 868kb region that underwent SNP based LD mapping.



	D2S2310	D2S2077	SNP26	SNP1	SNP3	SNP4	SNP6	SNP8	SNP9	SNP10	SNP11	SNP14	SNP15	SNP16	SNP17	SNP19	SNP20	SNP21	SNP22	SNP25
D2S2310	0.059																			
D2S2077	0.092	0.908																		
SNP26	0.067	0.873	0.95																	
SNP1	0.179	0.017	0.27	0.281																
SNP3	0.179	0.261	0.133	0.225	0.181															
SNP4	0.165	0.101	0.17	0.208	0.167	0.825														
SNP6	0.199	0.148	0.084	0.068	0.101	0.812	0.91													
SNP8	0.53	0.128	0.189	0.023	0.384	0.518	1	0.722												
SNP9	0.226	0.018	0.06	0.369	0.077	0.133	1	1	0.26											
SNP10	0.398	0.999	0.138	0.342	0.077	0.104	1	0.801	0.297	0.787										
SNP11	0.309	0.408	1	1	0.063	0.023	1	0.48	1	1	1									
SNP14	0.124	0.608	0.104	0.05	0.269	0.992	0.138	0.03	1	0.959	0.434	1								
SNP15	0.285	0.711	0.561	0.153	0.273	0.462	0.03	0.2	0.998	0.003	0.027	1	0.853							
SNP16	0.076	0.053	0.068	0.029	0.073	0.445	0.025	0.268	0.37	0.177	0.221	0.999	0.437	0.31						
SNP17	0.136	0.399	0.394	0.169	1	0.111	0.643	0.085	0.511	1	0.236	1	1	1	1					
SNP19	0.086	0.118	0.023	0.04	0.092	0.111	0.198	0.064	0.208	0.249	0.027	1	0.343	0.029	0.163	0.138				
SNP20	0.986	1	0.998	0	1	1	0	1	1	1	1	1	1	1	1	1	0.994			
SNP21	0.074	0.002	0.023	0.014	0.007	0.202	0.175	0.155	0.025	0.084	0.239	0.226	0.46	0.998	0.232	0.164	0.512	0.999		
SNP22	0.113	0.308	0.001	0.164	0.005	0.091	0.26	0.036	0.052	1	1	0.911	0.297	1	0.158	0.067	0.707	1	0.706	
SNP25	0.133	0.084	0.051	0.075	0.164	0.269	0.208	0.16	0.16	0.572	0.671	0.687	0.268	0.267	0.206	0.434	0.27	0.998	0.429	0.676

D'>0.8 D'>0.4 D'<0.2

D'>0.6 D'>0.2

Figure 6.10 LD relationships between SNPs based on D' values

The graph indicates the strength of the LD relationships between the SNPs used in the SNP-based association experiment. D' values greater than 0.4 are considered to be in relative LD.

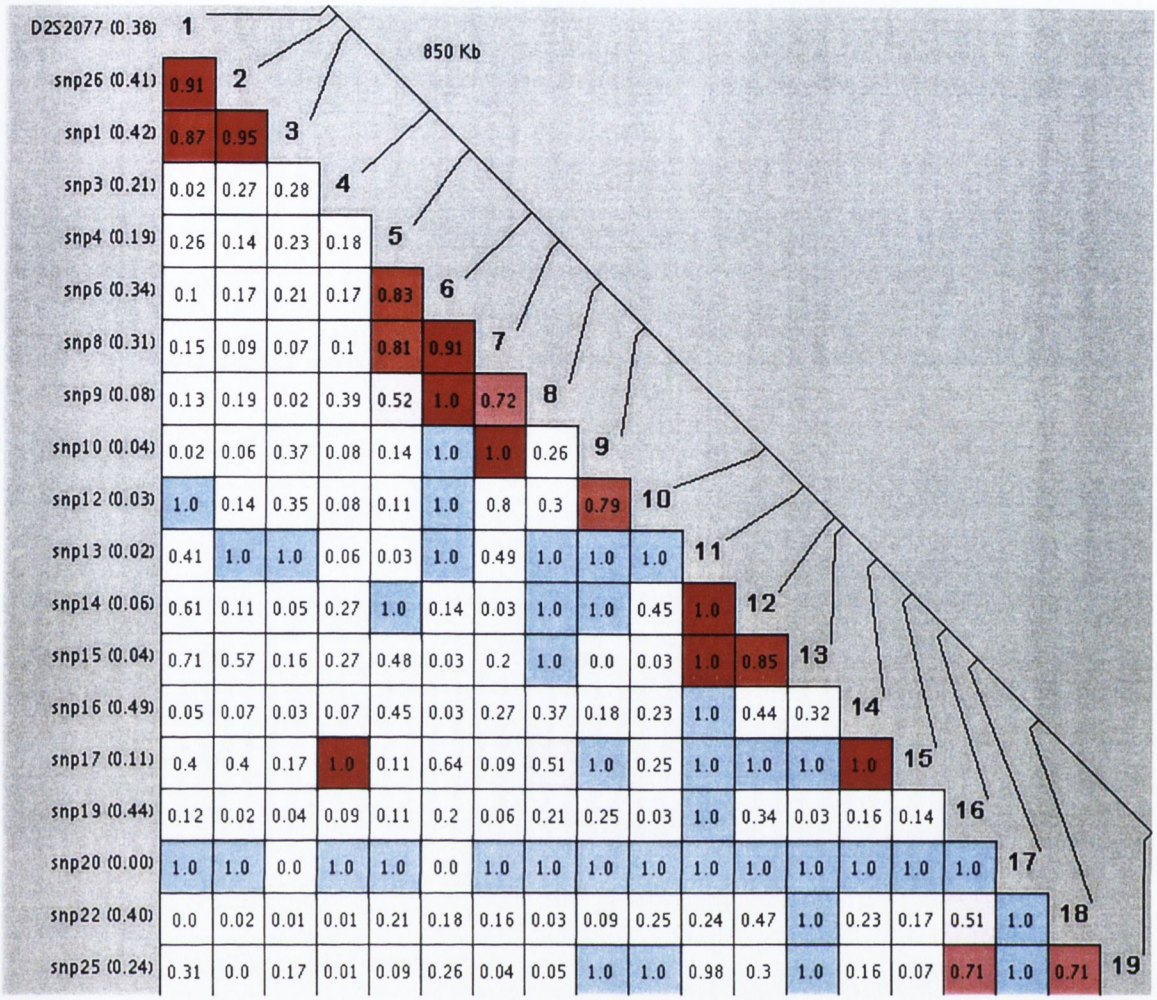


Figure 6.11a Haplotype blocks based on single nucleotide polymorphisms in chromosome 2q region

This figure shows the haplotypic structure in the region that was fine-mapped using SNPs. The haplotypes are derived from the LD relationships in the region based on D' values within the Haploview program. Positional bar on the right of the graph represents the positional relationship of the SNPs based on the physical position for each SNP.

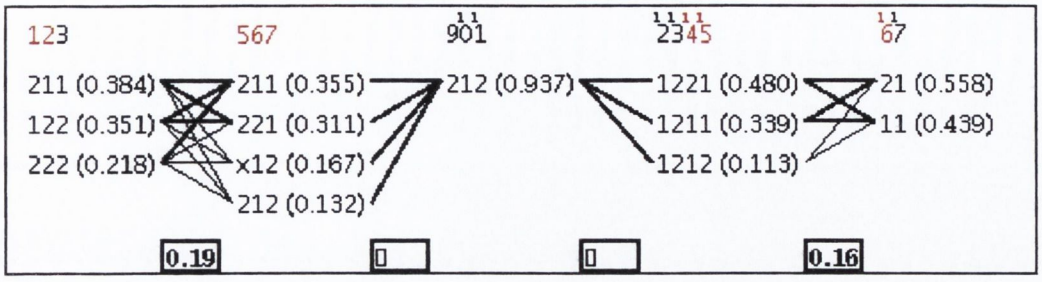


Figure 6.11b Relationship between 2 block haplotypes in region

This figure shows the crossover percentages of haplotypes observed in the sample. Dark lines indicate crossovers with a frequency greater than 10%. The lighter lines show inter-block relationship with a frequency less than 1%. The figures in boxes under the crossovers represents the multi-allelic D' values.

6.3.8 Association analysis of haplotypes using TRANSMIT

Association analysis of the underlying haplotypes in the region was undertaken using the TRANSMIT program which tests for association between multi-locus haplotypes and disorder by examining the transmission from parents to affected offspring. TRANSMIT haplotypes were constructed across the entire region comprised of two to seven markers. In addition, testing of association around the haplotypes generated by Haploview was also conducted. Haplotype data are presented as the results of association testing by TRANSMIT. TRANSMIT generates a global χ^2 for haplotypes (where the numbers of haplotypes are not excessive). Haplotypes showing significant transmissions in the autism sample are shown in Tables 6.13-6.17. To control for multiple testing, all transmissions were also analysed using bootstrap testing in TRANSMIT. This test performs simulations of the haplotype data based on a specified number of replications. In this case all significant haplotypes were subject to bootstrap testing of 1000 replicates. The p-value for bootstrap testing is also indicated in these tables.

6.3.8.1 Two-marker haplotypes

Association analysis of two-marker haplotypes is presented in Table 6.13. Significant association between the autistic sample and the following two haplotypes was observed, D2S2310/ D2S2077 (Global $\chi^2 = 24.98$, 14df, p=0.035) and D2S2077/ SNP26 (Global $\chi^2 = 8.06$, 3df, p=0.0447)). For markers D2S2310/ D2S2077, increased transmission of allele 2 in both cases was observed. Both of these alleles show evidence of increased transmission for the individual TDT analysis of these markers. Within the D2S2077/ SNP26 haplotype, alleles 2 and 1 were over transmitted respectively ($\chi^2 = 2.95$, p=0.0859, OR=1.16). The SNP4/ SNP6 haplotype did not show evidence of significant transmission on the global χ^2 test, however the allele 2/ allele 1 combination was transmitted with significantly greater frequency ($\chi^2 = 5.67$, p=0.0172, OR=1.3). The alleles in this haplotype were also those showing increased transmission in the TDT association testing for the individual loci (SNP4: $\chi^2 = 3.314$, p=0.069, OR=1.68; SNP6: $\chi^2 = 1.067$, p=0.302, OR=1.3). Bootstrap significance testing showed that the D2S2077/ SNP26 haplotype retained a significant p value on the global χ^2 test and the individual χ^2 for the 2.1 haplotype (p=0.043, p=0.05 respectively).

Table 6.11 Two-marker haplotypes with increased transmission in autism sample

Haplotype	Alleles	Bootstrap p value	Global χ^2 (df) (p value)	χ^2 indv(1df) (p value)	Bootstrap p value	OR
D2S2310/ D2S2077	2.2	0.26	24.98 (14df), (p=0.035)	6.2621 (p=0.012)	0.42	2.08
D2S2077/ SNP26	2.1	0.043	8.06 (3df) (p=0.045)	2.95 (p=0.086)	0.05	1.16
SNP4/ SNP6	2.1	0.073	6.76 (3df) (p=0.08)	5.67 (p= 0.0172)	0.061	1.3

Shaded cells indicate haplotypes that were significantly transmitted to the autism cases

6.3.8.2 Three-marker haplotypes

A significantly increased transmission of a haplotype (comprised of allele 2 (148bp) of D2S2077, allele 1 of SNP3 and allele 2 of SNP4) was observed in the autism sample (χ^2 = 5.64, p=0.018, OR=1.32). The Global χ^2 was also significantly associated (χ^2 =15.977, 7df, p=0.0253). TDT analysis of these loci individually (presented in section 6.3.6), also showed increased transmission of these alleles in the autism sample.

A SNP haplotype consisting of SNPs 4, 6 and 8 also showed a global significance (χ^2 = 15.68, 7df, p=0.028). A haplotype consisting of allele 2 for SNP4, allele 1 for SNP6 and allele 2 of SNP8 was significantly over transmitted (χ^2 = 7.2526, p=0.0071, OR=2.06). Once again, these were alleles that were found to be overtransmitted when the loci were tested individually with TDT analysis. The results of bootstrap testing was significant for two haplotypes D2S2077/ SNP3/ SNP4 (global χ^2, p=0.01, individual χ^2 , p=0.043) and D2S2077/ SNP4/ SNP6 (global χ^2 , p= 0.023, individual χ^2, p =0.057).

Haplotypes of D2S2310/ D2S2077/ SNP26, D2S2077/SNP4/SNP6 and SNP3/ SNP4/ SNP6 all showed borderline significance on the global chi-squared test. The results of these analyses are shown in Table 6.14.

Table 6.12 Three-marker haplotypes with increased transmission in the autism sample.

Haplotype	Alleles	Global χ^2 (df) p value	Bootstrap p value	χ^2 indv(1df) (p value)	Bootstrap p value	OR
D2S2310/ D2S2077/ SNP26	2.2.1	30.1 (20df) p=0.07	0.35	5.5 (p=0.02)	0.68	1.97
D2S2077/ SNP3/SNP4	2.1.2	15.98 (7df) p=0.03	0.01	5.6 (p=0.02)	0.04	1.32
D2S2077/ SNP4/ SNP 6	2.2.1	13.99 (7df) p=0.05	0.02	7.6 (p=0.006)	0.06	1.5
SNP3 / SNP4/ SNP6	1.2.1	9.34 (7df) p=0.23	0.1	6.2 (p=0.01)	0.06	1.36
SNP4/ SNP6/ SNP8	2.1.2	15.68 (7df) p=0.03	0.08	7.3 (p=0.007)	0.07	2.06

Shaded cells indicate those haplotypes that were significantly transmitted to the autism cases.

6.3.8.3 Four-, five- and six- marker haplotypes

It was less likely that Global χ^2 values were derived for haplotypes consisting of four or more alleles. This was because of the increase in the permutations of possible haplotypes. One four-marker haplotype was observed to show a global χ^2 significance. This was comprised of SNP3/ SNP4/ SNP6/ SNP8 (χ^2 =24.32, 20df, p=0.04). This was not significant on bootstrap testing (global χ^2, p= 0.361, individual χ^2, p= 0.23) (Table 6.15). The alleles that were significantly transmitted (allele 1 for SNP3, allele 2 for SNP4, allele 1 for SNP6 and allele 2 for SNP8) were again those that showed evidence for increased transmission when the loci were tested individually.

Table 6.13 Four-marker haplotypes with increased transmission in the autism sample

Haplotype (Alleles)	Alleles	Global χ^2 (df)	p-value	Bootstrap p value	χ^2 indv(1df) (p value)	Bootstrap p value	OR
D2S2310/ D2S2077/ SNP26/ SNP1	2.2.1.1	**Not calculated**	_	_	4.75 (0.023)	_	2.08
D2S2077/ SNP3/ SNP4/SNP6	2.1.2.1	**20.8 (14)**	**0.11**	**0.20**	6.68 (0.01)	**0.33**	1.6
SNP1/SNP3/ SNP4/SNP6	2.1.2.1	**19.24 (14)**	**0.16**	**0.055**	4.62 (0.032)	**0.426**	1.4
SNP3/SNP4/ SNP6/SNP8	1.2.1.2	**24.33 (14)**	**0.04**	**0.361**	9.48 (0.002)	**0.23**	2.5
SNP4/SNP6/ SNP8/SNP9	2.1.2.1	**17.23 (11)**	**0.1**	**0.406**	5.47 (0.02)	**0.138**	2.02

Several other haplotypes made up of four to six alleles were observed to show significant transmission to the autism sample. The odds ratios for these haplotypes were calculated and are presented in Table 6.18 together with odds ratios for two and three marker haplotypes. Graphical representation of the data are also presented in Figure 6.13. As can be seen from the graph, the haplotype showing the strongest odds ratio within the autism sample was a six-marker haplotype consisting of markers SNP1/SNP3/ SNP4/ SNP6/ SNP8/ SNP9 (OR= 6.71). Another six-marker haplotype, SNP26/ SNP1/SNP3/ SNP4/ SNP6/ SNP8, had an odds ratio of 4.15. A five-marker haplotype, SNP1/SNP3/ SNP4/ SNP6/ SNP8, had an odds ratio of 4.49. These haplotypes are all comprised of similar markers and the alleles transmitted within the haplotypes are the same.

Table 6.14 Five-Marker haplotypes with evidence of increased transmission in autism sample.

Haplotype	Alleles	χ^2 indv(1df) (p value)	P-value	Bootstrap p-value	OR
SNP26/SNP1/ SNP3/ SNP4/ SNP6	1.2.1.2.1	4.00	0.05	0.149	1.41
SNP1/ SNP3/ SNP4/ SNP6/ SNP8	2.1.2.1.2	11.56	0.0007	0.243	4.49
SNP3/SNP4/ SNP6/ SNP8/ SNP9	1.2.1.2.1	7.02	0.01	0.086	2.39
SNP4/ SNP6/ SNP8/ SNP9/ SNP10	2.1.2.1.2	5.68	0.02	0.145	2.16

Table 6.15 Six-marker haplotypes with evidence of increased transmission in the autism sample

Haplotype	Alleles	χ^2 indv(1df)	p-value	Bootstrap p value	OR
SNP26/ SNP1/ SNP3/ SNP4/ SNP6/ SNP 8	1.2.1.2.1.2	11.15	0.0008	0.6	4.15
SNP1/ SNP3/ SNP4/ SNP6/ SNP 8/ SNP9	2.1.2.1.2.1	12.06	0.0005	0.31	6.71

Table 6.16 Odds Ratios for Haplotypes observed to be significantly transmitted.

Haplotypes	Alleles transmitted	No. of haplotypes transmitted	No. of haplotypes not-transmitted	Odds Ratio
2 Marker Haplotypes:				
D2S2310/ D2S2077	2.2	25	12	2.09
SNP4/ SNP6	2.1	95	73	1.3
3 Marker Haplotypes:				
D2S2310/ D2S2077/SNP26	2.2.1	24	12	1.97
D2S2077/ SNP3/ SNP4	2.1.2	68	65	1.32
D2S2077/ SNP4/ SNP6	2.2.1	67	45	1.5
SNP3/ SNP4/ SNP6	1.2.1	82	60	1.36
SNP4/ SNP6/ SNP8	2.1.2	33	16	2.06
4 Marker Haplotypes:				
D2S2310/ 2S2077/SNP26/ SNP 1	2.2.1.1	19	9	2.11
D2S2077/SNP3/ SNP4/ SNP6	2.1.2.1	51	32	1.59
SNP1/SNP3/ SNP4/ SNP6	2.1.2.1	53	37	1.43
SNP3/ SNP4/ SNP6/ SNP8	1.2.1.2	30	12	2.51
SNP4/ SNP6/ SNP8/ SNP9	2.1.2.1	25	12	2.08
SNP6/ SNP8/ SNP9/ SNP10	1.1.1.2	72	62	1.17
5 Marker Haplotypes:				
SNP26/ SNP1/SNP3/ SNP4/ SNP6	1.2.1.2.1	41	29	1.41
SNP1/SNP3/ SNP4/ SNP6/ SNP8	2.1.2.1.2	19	4	4.49
SNP3/ SNP4/ SNP6/ SNP8/ SNP9	1.2.1.2.1	23	10	2.39
SNP4/ SNP6/ SNP8/ SNP9/ SNP10	2.1.2.1.2	23	11	2.16
SNP6/ SNP8/ SNP9/ SNP10/ SNP 11	1.1.1.2.1	73	60	1.21
6 Marker Haplotypes:				
SNP26/ SNP1/SNP3/ SNP4/ SNP6/ SNP8	1.2.1.2.1.2	17	4	4.15
SNP1/SNP3/ SNP4/ SNP6/ SNP8/ SNP9	2.1.2.1.2.1	15	2	6.71

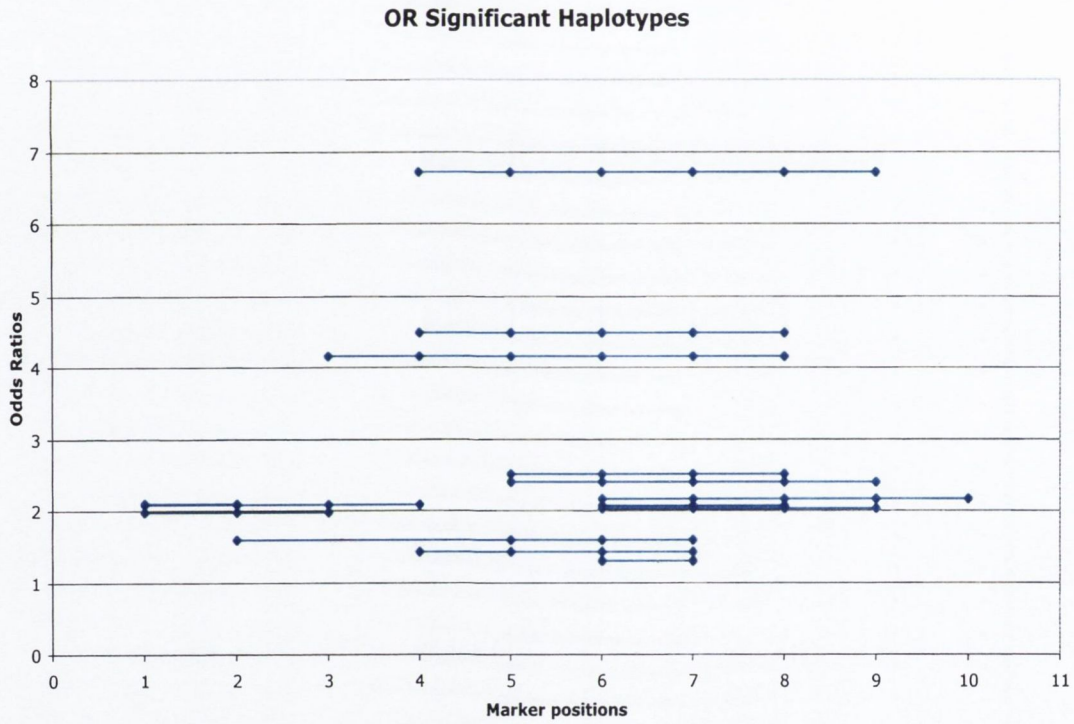


Figure 6.12 Comparisons of the odds ratios for associated haplotypes

This diagram shows a comparison of the Odds Ratios for marker haplotypes. Two- to six-marker haplotypes are illustrated. Markers are labelled 1-10.

1=D2S2310 3= SNP26 5= SNP3 7=SNP6 9= SNP9

2= D2S2077 4= SNP1 6= SNP4 8=SNP8 10=SNP10

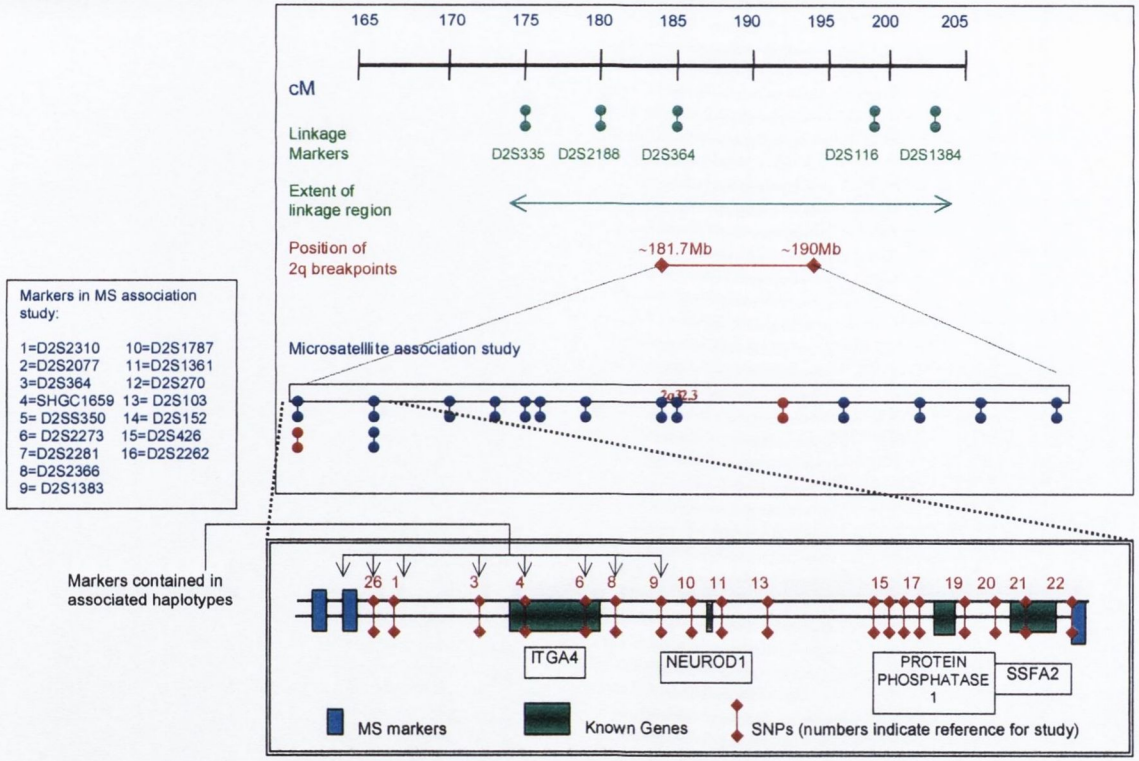


Figure 6.13 Diagram illustrating the relationship between the linkage findings, the cytogentic breakpoints and results of the microsatellite and SNP-based association mapping experiments

This figure illustrates the relationship between the linkage findings on chromosome 2q, the translocation described in section 6.2 and the results of the association findings from section 6.3. The relative positions of the markers from linkage studies are shown in green located between 175cM and 203cM. The position of the translocation is shown in red at 181.7-190Mb. The LD association experiment is shown underneath. The two MS markers shown in red are those showing evidence of association with autism. The region between markers D2S2310 and D2S364 is shown underneath with the relative positions of the SNP markers used for fine-mapping. The SNPs located haplotypes found to be associated with autism are indicated by arrows and these are located around the gene of interest, ITGA4.

6.4 Discussion

The results presented in this chapter represent the characterization of a region on chromosome 2q32 with converging evidence supporting the presence of a risk locus for autism. The evidence derived from several lines of investigation. The first line of evidence was from the characterization of a region on chromosome 2q that was translocated in an individual with autism. The translocation lay within a region of interest in autism as highlighted by the results of three published genomewide linkage studies. The second piece of evidence comes from the investigation of the region in a well-characterized sample of individuals with autism recruited in the Irish population. The relationship between the linkage findings, the cytogenetic translocation and the results of association testing are shown in Figure 6.14. This figure illustrates the position of the linkage markers on chromosome 2q, located between 175cM and 203cM. The translocation was mapped to 181.7Mb-190Mb, which as shown in the figure is located within the linkage region. The positions of the microsatellelite markers that were used for the first association experiment are shown. The markers shown in red showed evidence of association. An association finding around D2S2077 was followed up by a SNP-based fine mapping experiment that is also illustrated in the diagram. The positions of these SNPs relative to genes in the region are illustrated and the arrows indicate the SNPs contained on haplotypes that were found to be associated with autism. These SNPs are concentrated around the gene Integrin-Alpha 4, proposed in this work as a candidate gene for autism.

There are some aspects of the work that have not as yet been addressed which will be discussed briefly, followed by a discussion of the implications of the findings and future plans for investigations of this region in autism.

6.4.1 Difficulties and anomalies in characterization of the cytogenetic region.

As discussed above (section 6.2), a considerable effort was devoted to the characterization of the chromosomal translocation. The translocation of chromosome 2q32.3 to chromosome 9q31.1q31.2 was cryptic and only noticeable on hindsight following the investigations outlined above. The abnormality was perceived to be a deletion at first and the microsatellite-based association experiment was undertaken in tandem with the further

attempts to define the region that subsequently showed that the anomaly was a translocation. A disadvantage of the late characterisation of the breakpoints of the region is that the region defined for fine mapping was shorter at the 5' end (630kb) than the actual size of the translocation. Better coverage in the region would have been achieved if this had been included at the centromeric end of the deletion.

One disappointing result is the failure to identify a known gene that was disrupted by the chromosome 2q breakpoint. A predicted gene, 'Weakly Similar to Vegetable Incompatibility Protein', located at the telomeric end of the region was disrupted but does not present as a likely candidate gene for autism. No known function has been identified for this predicted gene nor is there evidence to suggest that it might have a function in the CNS or in neurodevelopment. At present the breakpoints on chromosome 9 are not exactly defined and it is as yet unknown if a gene either on chromosome 9 or within the chromosome 2q translocated region has been disrupted. This latter case would occur if a small portion of the translocated region was also deleted.

A further consideration is that the translocated portion of chromosome 2q was inverted on chromosome 9q. The significance of this is the potential loss of regulatory regions for genes within the region causing a haploinsufficiency of a gene, or the opposite case that a gene within the region is expressed (either through loss or gain of regulatory regions). This is most likely to occur if the region was imprinted. There is evidence from other species that the region is imprinted (Cockett et al. 1996) and the presence of a regulatory region for $MeCP_2$ has been identified in an overlapping syntenic mouse region from the published databases. Preliminary TDT analysis on microsatellite markers and single nucleotide polymorphisms however, does not indicate overtransmission of alleles from either parent to their affected offspring. Thus further investigations of the imprinting hypothesis are required.

6.4.2 Fine mapping of the region using microsatellite markers and single nucleotide polymorphisms (SNPs)

An exploratory fine mapping experiment using microsatellite markers was conducted across the region defined by microsatellite mapping and gene dosage analysis. A 8.5Mb region was mapped and the average density of microsatellite markers was approximately 500kb. Two microsatellites (D2S2077 and D2S270) were identified in the region that showed evidence for association in the autism sample. The association observed around D2S2077 is more easily interpreted as this was a biallelic marker. Transmission of the less frequent marker in the autism sample was significantly increased (χ^2=5.8, p=0.013, OR = 1.75). On the basis of this finding, a region of approximately 868kb was defined for fine mapping using SNPs at a density of 40kb on average. The association testing of SNP genotypes using the TDT revealed a non-significant increased transmission of alleles at two loci (SNP3 and SNP4) (SNP 3: χ^2=2.88, p=0.0897, OR=1.63; SNP4: χ^2= 3.314, p=0.069, OR=1.68). HHRR analysis of the transmission of SNP4 was significant (χ^2= 5.16, p=0.023, RR=1.48, 1.01<RR<2.16). Haplotypes were constructed using the SNP genotypes and genotypes from the flanking microsatellites and D2S2077. Association testing of these haplotypes showed an association that included the transmitted alleles for D2S2077, SNP3 and SNP4 (Global χ^2= 15.98, (7df), p=0.03, bootstrap significance = 0.01, haplotype 2.1.2 χ^2=5.64, p=0.02, bootstrap significance= 0.043, OR=1.32). A second three-marker haplotype (D2S2077/ SNP4/ SNP 6) showed a significantly increased transmission in the sample (Global χ^2=13.99, (7df), p=0.05, bootstrap significance = 0.023, haplotype 2.2.1 χ^2=7.6, p=0.006, bootstrap significance= 0.057, OR=1.5). The most significant haplotype was a six-marker haplotype (SNP1/3/4/6/8/9) (χ^2= 12.06, p= 0.0005, OR=6.71), however this haplotype had a frequency of 8.5% and was not significant when bootstrap testing with 1000 replicates was applied.

Analysis of the data for D2S2077 and SNPs with the Haploview programme identified the presence of two blocks of linkage disequilibrium (LD) in this region comprised of these markers. While it could be argued that more dense SNP genotyping would be required to fully characterise the underlying haplotypic structure, based on the genotypic data available it appears that the SNP haplotypes found to be associated with the autism sample were

located largely on a haplotypic block of LD as defined by Haploview. It is to be expected that the associated SNP haplotypes occur on the same block of LD defined by Haploview and this is largely what was observed in the data. One three-marker haplotype (D2S2077/ SNP4/ SNP 6) spanned two haplotypic blocks of LD suggesting that the association with autism is present across both haplotype blocks. This might also be expected when the Haploview data are considered. Figure 6.12b showed the presence of LD between these two blocks. The association detected between the SNP haplotypes derived by TRANSMIT and autism might be explained by the presence of a disease-causing variant within it, with which the SNPs are in LD. It is unlikely that any of the SNPs tested are the disease-causing variant. With the exception of SNPs 4 and 6, none of the SNPs occur within a known gene. SNPs 4 and 6 are intronic SNPs and it seems unlikely that either is a causative variant. It is interesting that the haplotype occurs around a known gene within the region, namely the integrin alpha 4 gene (ITGA4). SNP4 is located within the second intron of ITGA4, while SNP6 is located within intron 23. The potential role of this gene as a candidate gene for autism is discussed below.

6.4.3 Integrin Alpha IV (ITGA4):

ITGA4 belongs to a family of genes comprised of cell surface receptors for extracellular matrix components in addition to receptors involved in leukocyte adhesion. ITGA4 was mapped to 2q31-q32 by nonisotopic in situ hybridization and by fluorescence in situ hybridization (Jaspers et al. 1991; Fernandez-Ruiz et al. 1992). It is a 1038 amino acid protein. The protein product normally exists as an alpha-beta heterodimeric transmembrane glycoprotein and it normally bonds with either beta-1 or beta-7 subunits. Both alpha4-beta1 and alpha4-beta7 have a well-defined role as receptors for fibronectin (a glyco-protein located on the surface of fibroblasts) and vascular cell adhesion molecule 1 (VCAM-1) (Lu and Cyster 2002). However, there is good evidence to suggest that integrins influence a wide range of cellular functions including differentiation, cell proliferation, process outgrowth, gene expression and survival. The expression of integrin alpha and beta subunits in adult rat brain has shown that the alpha 4-subunit is almost exclusively expressed in discrete regions of the limbic telencephalon, the olfactory cortical layer II and the hippocampal CA2 (Pinkstaff et al. 1999). Integrins are now thought to play multiple

roles within the CNS during development and in the adult brain, including synaptogenesis, stabilisation of the blood-brain barrier and activation of microglia (Milner and Campbell 2002). Given the evidence supporting the role of Integrins in neuro-development this gene presented as a good candidate for autism within the region that was fine mapped using SNP markers. The evidence is further supported by the expression profile of ITGA4 in the limbic system as this fits with one of the neuropychological theories of impaired functioning within the Limbic system associated with impaired emotional behaviour and social interactions in autism (Kemper and Bauman 1998).

6.4.4 Ubiquitin-conjugating enzyme E2E 3 (UBE2E3):

This gene was identified at the 5' end of the deleted region on chromosome 2q and thus none of the markers genotyped were located in or near to this gene. However given the proximity of this gene to the 5' breakpoint of the translocation, it is worthy of further investigation in future studies. This gene is involved in the ubiqutination of a protein substrate. The ubiquitination pathway is a cyclical 4-step process, targeting specific cellular proteins for degradation and regulating their activity. It operates in all cell-types and involves three types of enzymes: E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin protein ligases. UBE2E3 belongs to the Class II family of Ubiquitin conjugating enzymes. The cDNA for this gene has been isolated and the predicted product is a 207-amino acid protein. Northern blot hybridization revealed that UBE2E3 is expressed as a 1.9-kb mRNA primarily in skeletal muscle (Ito et al. 1999). While the known expression profile of this gene does not make it a likely candidate for autism, it is noteworthy that another gene involved in the ubiquitination pathway, namely UBE3A, on chromosome 15q11 has been regarded as a potential candidate for autism (Nurmi et al. 2001). UBE3A appears to have a different function to UBE2E3. It is a protein ligase and belongs to a family of genes (HECT proteins) that appear to have a role in substrate recognition and protein transfer. It also has been demonstrated to have CNS expression, unlike UBE2E3 (Herzing et al. 2002). Given the relationship between the two genes it could be considered as a potential candidate for autism, although based on what is known about the function of the gene product to date, it is difficult to speculate on its potential role in the disorder.

6.4.5 Neurogenic Differentiation 1 (NEUROD1):

As illustrated in Figure 6.14, this gene is located within the region that underwent SNP based association testing. It is located approximately 200kb from the MS marker, D2S2077, that was found to be associated. It is a small gene spanning 4kb and containing 2 exons (Tamimi et al. 1996). It is a transcription factor that belongs to the basic helix loop helix group of proteins. It functions during neurogenesis and is responsible for determining cell type. It is expressed in both the central and peripheral nervous systems and the predicted product is a 357 amino acid polypeptide (Tamimi et al. 1996). Mutations within the gene have been described and are associated with Diabetes Mellitus Type II/ Maturity Onset Diabetes of the Young. (Malecki et al. 1999). While the function of the gene in the human CNS is unknown, the evidence from animal models is interesting. Mouse models that are homozygous for a deletion within the gene showed abnormalities of the development of the dentate gyrus within the hippocampal formation and developed limbic seizures (Liu et al. 2000). This is of interest to autism because of the abnormalities within the limbic system and the association with seizures, also observed in autism (Fombonne 2002). In addition Liu et al, 2000 also report the presence of abnormalities in the cerebellum of these mice, with an apparent increase in the number of glial cells with a consequent increase in the size of the cerebellum. This is also an interesting finding as it ties in with a similar phenomenon of increased cerebellar size in autism (Bailey et al. 1998; Carper and Courchesne 2000).

6.4.6 Sperm Specific Antigen 2 (SSFA2)

This gene is also located within the region of interest on chromosome 2q at approximately 500kb from the associated marker, D2S2077. It spans 39.7kb and has 18 exons. It undergoes alternative splicing and has 14 known transcripts. It has seven alternative promoters and five non-overlapping last exons. This is a sperm surface antigen involved in the early cleavage of the fertilized embryo. However it has also been shown to be expressed in the retina, colon and lung. It has been shown that it is translated on the reverse DNA strand from 3' to 5' and alternative splicing may influence the expression of neighbouring genes on the 5' strand. From the published databases (AceView: http://www.ncbi.nih.gov/IEB/Research/Acembly/) there appears to be overlap between this gene on the 3' strand and the NEUROD1 gene on the 5' strand. While the function of the

gene and its role in autism is not obvious, it will be considered in future studies in the chromosome 2q region.

6.4.7 Homosapiens Protein Phosphatase I, Regulatory Subunit 1A

This gene is contained in the associated region at approximately 400kb from D2S2077. It spans 1.18kb. It has an inferred protein product and the function is not well elucidated. It appears to inhibit protein phosphatase and is expressed primarily in skeletal muscle. This gene will not be prioritized for investigation in autism unless a clear role in the CNS or neurodevelopment emerges.

6.4.8 Information from investigation of syntentic regions in the mouse

The similarities between autism and Rett disorder have been discussed in chapter one. Several mutations seem to account for differences in the phenotypic features of the condition. Rett disorder was previously thought to affect only girls but a lesser variant in males and a preserved speech variant have been described. It has been proposed that some of these clinical presentations may be mistaken for autism and that screening for these mutations in autism is warranted. This sample has been screened for known methyl CPG binding protein 2 (MeCP2) mutations, and none were identified. It is likely that there are other as yet unidentified mutations within the MeCP2 gene that have a similar effect. A gene on the X chromosome that influences gene expression may account for the sex ratio differences in autism particularly if a mutation effects its expression, since males with such a variant are by necessity hemizygous.

MeCP2 acts on sites within the genome influencing gene expression. A search of syntenic regions in the mouse genome revealed the presence of a MeCP2 regulatory region overlapping at the 3' end of the translocated region on chromosome 2q32. It is not known if this is also present in the human genome, however the region appears highly conserved between mouse and human and it is not unreasonable to assume that a similar region may exist in the human genome. The presence of such a regulatory region would not explain the findings of the linkage disequilibrium study, since the region in located downstream of the

ITGA4 gene. However if alternative mechanisms or genes are involved within the region it is possible that this may play a role in predisposing to autism.

6.4.9 Weakness in the linkage disequilibrium mapping experiments:

There are some limitations to the work that has been presented here. Firstly, the chromosomal breakpoints had not been clearly defined at the time that the maps for the fine mapping experiments were constructed. The translocated portion of chromosome 2q was approximately 630kb larger at the 5' end of the deletion than was originally estimated. It included a known gene (UBE2E3) that was not included in the fine-mapping experiments. Therefore none of the data here has investigated any linkage disequilibrium testing for this region or gene and this will need to be addressed in future work.

A major problem was created by changes that occurred in the published maps of this region over the course of this work. The most significant change involved the orientation of a BAC clone in the region that shifted the positions of genes and markers considerably. The outcome was that the markers were not as uniformly distributed as they had been designed to be. For example, on the UCSC Genome Browser July 2002, D2S2310 and D2S2077 were spaced 100kb apart at the time when the experiment was undertaken, but currently (July 2003) are positioned only 15kb apart. Map changes also had a significant effect on the positions of SNPs relative to the microsatellite markers for the SNP mapping experiment. Due to a change in the position of D2S2077, it was located at the 5' end of the region being mapped whereas when the experiment was designed it was positioned between SNPs 4 and 6. The implication is that there is probably insufficient information about the LD in the region upstream of D2S2077 and the region remains untested for association in the autism sample. While the inferred haplotypic structure in the region seems to imply that there is a break in LD around D2S2077 (Fig. 6.12a), more SNP genotyping would be required to confirm this.

While the haplotypic structure of the region has been inferred, the density of the SNP map is insufficient to confirm the underlying block structure within the region. Based on recent investigations it has been estimated that blocks of LD in the genome exist over distances

between 10-100kb (Daly et al. 2001). The haplotype defined by Haploview represents a distance of approximately 160kb. While it has been observed that in some regions of the genome, haplotype blocks can extend over larger distances (Dawson et al. 2002), further SNP mapping at a higher density would be required to confirm if this was true in the current case. The density of the SNP map was constrained by obvious factors such as costs, however given these findings, it now merits further investigation with genotyping of densely spaced SNP markers (at approximately every 5-10kb). This will allow for more accurate assessment of haplotypic structure in the region in addition to the construction and tagging of haplotypes that will reduce the costs of attempted replication in another sample.

Time constraints did not allow for investigations around marker D2S270. This marker showed a genotypewise association in the autism sample. The interpretation of this finding is difficult. It appears that particular alleles are transmitted more frequently than expected in the sample. This may indicate the presence of a haploytpe around this marker where these alleles co-occur with a causative variant. One of the main difficulties in addressing this was the difficulty in identifying other informative microsatellite markers in the region. Indeed the D2S270 marker is located approximately 1.5Mb on either side from the flanking markers. It will be necessary to refine the map further in the region and SNP genotyping around the marker is currently in progress.

6.4.10 General difficulties

The linkage disequilibrium experiments in this investigation were based on a sample of 96 autistic trios. Power calculations are largely dependent on factors that are unknown at the start of an investigation such as this, e.g. the size of the gene effect An estimate of the power of this study to detect a susceptibility allele under variable parameters was conducted using the Genetic Power Calculator (Purcell et al. 2003) (http://statgen.iop.kcl.ac.uk/gpc/). Based on the assumptions of an allele frequency of 25%, population prevalence of 1/1000 and a relative risk of 2 and 3 respectively for heterozygotes and homozygotes with the high risk allele, A study of 100 trios should have a power of 76% to detect an association at the 0.05 significance level. For a 0.01 significance level this number of trios has a power of 54% to detect an effect. This study should be

considered therefore as exploratory and any detected association would require replication in a larger sample set. Given the small sample size, it was less likely that a positive association would have been detected, and so the positive findings reported merit more in-depth investigation. There are a number of approaches that should be included in further investigation.

6.4.11 Future directions for this study of the role of a potential susceptibility locus on chromosome 2q for autism

6.4.11.1 Refining the linkage disequilibrium experiment

As discussed, several gaps exist within the region that underwent linkage disequilibrium testing. Fine mapping of the broader region, showing evidence for linkage is currently in progress. This involves linkage disequilbrium mapping with microsatellite markers at an average distance of 500kb and will include the region that was translocated and not mapped in this investigation.

The gap in the map around marker D2S270 is currently being refined using SNP markers at a density of 50kb. This will allow for linkage disequilibrium mapping at greater resolution and will include markers that are within the Integrin alpha V gene.

6.4.11.2 Defining the haplotypic structure and haplotype tagging

By genotyping SNPs at a higher density in the sample, it will be possible to define the underlying haplotype structure more accurately. Haplotype tagging will identify SNPs that are informative for a haplotype and will reduce the amount of genotyping that will be necessary in attempts to replicate this work. This will allow for linkage disequilibrium studies in an expanded sample of individuals with less genotyping and a consequent reduction in the cost.

6.4.11.3 Replication studies

Clinical sample collection has continued and an additional 80 families are now available for molecular genetic investigations. All individuals have been recruited using the same criteria as described in chapter 3. It will be possible to expand the current sample to approximately

160 well-characterized cases. It is planned to genotype these families for the same markers and present the analysis for the entire sample.

Through collaboration with Dr. Astrid Vicente at the Gulbenkian Institute in Lisbon, genotyping of tagging SNPs will be conducted in a sample of 250 Portuguese autistic trios. This sample has been recruited in a similar manner to the Irish sample and the autistic individuals have been equally well characterized and subject to the same inclusion criteria. One third of this sample was recruited in the Azores, a genetically isolated population. The advantage of using this population lies in the homogeneity of the sample and the possibility that potentially fewer genes are involved in complex disorders in isolates. Haplotypes in isolated populations are likely to be larger and show less heterogeneity as the population was founded relatively recently and went through a bottleneck. This is likely to make the identification of association with a haplotype easier.

6.4.11.4 Expression studies

Further investigations in the index case might include assessment of expression of the genes within the region. It would be useful to obtain haploid cell lines for this purpose to be able to determine expression from haploid copies of both chromosomes 2 and 9. It has been determined that the maternal copy of chromosome 2q was translocated. It would be necessary to determine if it was translocated to the maternal or paternal copy of chromosome 9, particularly if imprinting was confirmed within chromosome 2.

6.4.11.5 Candidate gene studies

Candidate gene studies are planned in this region. The focus initially will be on the three genes contained within the region that underwent linkage disequilibrium mapping using SNPs (i.e. Integrin Alpha 4, Neurogenic Differentiation Factor 1 and Sperm Specific Antigen 2). The approach to candidate gene studies will involve a number of methods.

6.4.11.6 Linkage disequilibrium mapping of candidate genes using single nucleotide polymorphisms

Candidate genes within the region defined by single nucleotide polymorphism (SNP) mapping will be investigated by genotyping SNPs at a higher density than the investigations described here. The proposed density of SNPs is 5-10kb and this will be informative regarding the underlying haplotypic structure. As discussed above, haplotypes will be defined in greater detail and tagged to allow for less labour intensive replication.

6.4.11.7 Mutational screening of candidate genes

Candidate genes will undergo mutational screening using dHPLC (Transgenomic Wave machine). The objective is to identify new variants that will be sequenced and characterized based on the potential effects of sequence change, e.g. non-synonymous SNPs leading to altered gene expression. Exons and regions of introns that flank exons will be screened in addition to promotor regions and 3'UTR.

It may not be possible to detect the presence of non-synonymous SNPs within the candidate genes, however new variants may also be utilized in association studies as described above.

6.4.11.8 Further characterization of region on chromosome 9q

The breakpoints on chromosome 9q are still not refined and it is not yet known if a gene on this chromosome has been disrupted. As there was evidence from several sources for the role of this region of chromosome 2q in autism, it was decided to pursue these investigations initially. However studies on chromosome 9q are planned. In addition to better localization of the breakpoints, investigations of the potential effects of the inversion of the chromosomal region on 9q is also planned. This will involve the characterization of the region, investigating the presence of potential regulatory regions that may effect the expression of genes within the translocation.

6.4.12 Conclusion

The results of the characterization of a susceptibility region for autism on chromosome 2q have been described here. The studies involve the dual approaches of characterization of a

chromosomal translocation that was identified in an individual with autism and linkage disequilibrium (LD) fine mapping using microsatellite markers and single nucleotide polymorphisms. The LD studies have shown evidence of significant association to autism in the sample described here. The findings are interesting as they represent converging evidence for a susceptibility gene for autism within this region. One gene, Integrin alpha 4, occurs on a haplotype that was found to be associated with autism. This gene has a described function in the CNS and is expressed particularly in a region that has been postulated to be involved in autism. Further investigations are in progress to refine some of the findings described here. Future plans for this research have been outlined. A variety of approaches will be utilised, involving attempts at replication, further linkage disequilibrium studies, candidate gene studies and investigations of expression of the translocated genes in the index case. Owing to the complexity of the genetic mechanisms of complex disorders, such as determining the mode of inheritance, additive effects, determining the presence of imprinting or other epigenetic phenomena, a variety of approaches will be required in the examination of candidate genes.

Chapter 7
General discussion and future directions

This research thesis was concerned with molecular genetic investigations in autism and had two distinct components. The first phase of the research was devoted to the establishment and management of a research group to pursue these investigations and the construction of a well-characterized sample of individuals with autism. The second phase of the work presented here involved a laboratory phase of molecular genetic investigation that was undertaken by the author.

The clinical phase recruited a sample of parent/ child trios with autism with optimal phenotype definition. The procedures for recruiting this sample were outlined in Chapter 3. Briefly, standard research diagnostic assessments (Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule-Generic) were used and rigorous exclusion criteria were applied to minimize phenocopies from the sample. During the first phase of recruitment 114 children were recruited. Fourteen of the recruited families either withdrew or were excluded from the final molecular investigations. The remaining 100 children were comprised of 93 children who met both ADI-R and ADOS-G criteria for autism and seven children who met ADI-R criteria for autism and ADOS-G criteria for autism spectrum disorder. Genetic heterogeneity was minimized in the sample by recruiting an ethnically Irish sample with the exception of four families where one parent was Irish and the other was either North American or European. All individuals recruited were Caucasian.

During the clinical recruitment an individual with autism and a translocation of an interstitial portion of the long arm of chromosome 2q was identified. This was a very interesting discovery as the abnormality occurred within a region shown by three genomewide linkage studies to have putative evidence for linkage and helped to shape the direction of the molecular genetics investigations that were undertaken in phase II of the research.
During the clinical phase, grant funding was obtained by the author to continue clinical recruitment and this is in progress. To date a further 80 parent/ child trios have been recruited using the same procedures as in the clinical phase described herein.

The laboratory phase of the research followed two lines of investigation. The first investigation was a genetic association study between variants in two neurodevelopmental genes, HOXA1 and HOXB1, and autism. This was an attempted replication study of work published by Ingram *et al*, 2000 who reported an association between a variant in the HOXA1 gene and a sample of individuals with autism. This project also provided an opportunity for testing the sample for any potential problems such as sample incompatibility, paternity discrepancies, or poor DNA quality. No significant association was found in the Irish sample for polymorphisms identified by Ingram in either the HOXA1 or HOXB1 genes although a non-significant trend towards increased transmission of the G allele of the HOXA1 variant was identified. Additionally, unlike the Ingram study, no parent-of-origin effects were noted with respect to the transmission of the HOXA1 variants.

The major line of genetic investigation in this work was determined by the identification of an individual with autism and a cytogenetic translocation of a portion of chromosome 2q. This was a very interesting and fortuitous finding as the translocation occurred in a putative region of linkage from genomewide linkage studies (Buxbaum et al. 2001; IMGSAC 2001; Shao et al. 2002a). A collaboration was established with the National Centre for Medical Genetics to characterize the translocation and map the breakpoints. As outlined in Chapter 6, mapping the anomaly required more work than was anticipated and took a number of surprising turns. It was judged to be a deletion on the basis of the cytogenetic spreads of G-banded metaphases and a gene dosage study gave results supporting this hypothesis. Subsequently comparative genome hybridization failed to establish the absence of chromosome 2q material and Flourescent in-situ hybridization (FISH) demonstrated the presence of a chromosome 2q probe on chromosome 9, confirming that the anomaly was a translocation. FISH mapping was used to define the deletion breakpoints on chromosome 2q and the insertion breakpoint on chromosome 9q is currently being narrowed down. This was a highly significant finding as the possibilities arising are that there may be a candidate gene for autism at the breakpoints on either chromosome 2q or on chromosome 9q. Due to the converging evidence for the presence of a susceptibility gene for autism on chromosome 2q, further investigation on 2q was pursued first.

An LD mapping study in the clinical sample described here was conducted across the region defined by the gene dosage investigations. In the first instance an experiment was conducted using microsatellite markers at average intervals of approximately 500kb. Using

transmission disequilibrium test (TDT) analysis LD was detected between a biallelic marker (D2S2077) located at the centromeric end of the translocated region and the autism sample (χ^2=5.8, p=0.013, OR = 1.75). A multiallelic marker (D2S270) located closer to the telomeric end of the deletion was found, using eTDT association testing, to have a genotypewise association with the autism sample (χ^2=43.48, 19 d.f., p=0.001). This provided further evidence for the potential presence of a susceptibility gene. Interestingly the D2S2077 marker was located close to three genes of known function and a further gene with a postulated function. At least two of these appeared to be potential candidates for autism based on their known function (discussed below).

The second investigation conducted on chromosome 2q followed up the linkage disequilibrium findings between D2S2077 and autism. A fine mapping linkage disequilibrium experiment using SNP markers was conducted across a region spanning 868kb that contained the D2S2077 marker. A trend towards increased transmission of one SNP (SNP4: dbSNP ref: 155149) was observed with TDT association testing (χ^2=3.314, p=0.069, OR=1.68). Association testing for this SNP was significant using HHRR analysis (χ^2=5.16, p=0.023, RR=1.48, 1.01<RR<2.16). Association testing of haplotypes constructed from the SNP and flanking microsatellite markers was undertaken using the TRANSMIT program. Significant transmission of several haplotypes comprised of two to six markers were observed. All of these haplotypes contained marker alleles that were found to be overtransmitted in the autism sample. The haplotypes occurred around a known gene, Integrin Alpha 4 (ITGA4) and two of the markers were located in introns of this gene. ITGA4 is expressed in the limbic brain, a region that has been implicated in autism and represents an interesting candidate gene worthy of further study. Two other genes located within the region that underwent SNP mapping on chromosome 2q present as interesting candidates. Neurogenic Differentiation Factor 1 (NEUROD1) is a small gene that is expressed in the limbic brain and cerebellum. Mouse null mutants have been found to have brain abnormalities similar to those found on post-mortem examination of autistic brains. Based on these reports, NEUROD1 also represents an interesting candidate gene for autism. Sperm Specific Antigen 2 (SSFA2) is expressed specifically in the testis, however it may influence the expression of surrounding genes based on positional effects in the genome. These genes are currently subject to further investigation and it is a priority to now conduct fine mapping LD experiments in these genes in an expanded Irish sample

There were several factors that were limiting to the progress of this study. The research commenced following the award of a Wellcome Trust Training Fellowship, however it was apparent from quite early on that it would be difficult to accomplish the aims of the project with one researcher. It was thus necessary for the author to take an active role in procuring grant funding to establish a research group in autism genetics. To date two research nurses have worked alongside the author on the clinical aspects of the project which has been invaluable given the necessarily slow recruitment process, which is a second limitation. The research tools described in Chapter 3, the ADI-R and the ADOS-G, are lengthy to apply (particularly the ADI-R) and the necessity to conduct two assessments is also time consuming. The advantage of having chosen this slow procedure is that the sample that has been collected is well characterised. Good phenotypic characterisation has contributed to the promotion of research collaboration with other groups (particularly the NAAR-AGP collaboration discussed below). From a clinical perspective a more standardised approach might also have included uniform psychological measures of intelligence, however funding was not obtained for a psychologist. Other demands on the author's time included the management of the research group, obtaining training as a clinical trainer in the research assessments and in turn training and supervising the clinical researchers, supervising lab personnel and establishment and management of research collaborations. Obviously this meant that the focus for the author was not always concentrated solely on the work presented here however the advantage is that in addition to having a good autism sample, the research group is now well-trained and is beginning to develop an international reputation.

With respect to phenotypic measures, the clinical data generated by the ADI-R and ADOS-G instruments does not lend itself easily to the investigation of association between candidate genes and phenotypic sub-types. However a proposed approach to this analysis is to perform cluster analysis of ADI-R items for genetic analysis (Tadevosyan-Leyfer et al. 2003). This would allow for association testing between candidate genes and particular aspects of the phenotype, e.g. repetitive behaviours or obsessional features and serotonergic genes. A second approach might include collecting further quantitative clinical measures for a particular sub-phenotype, e.g. hyperactivity features or obsessionality. With additional funding it is planned to add a variety of other clinical measures that will include both quantitative measures of autistic traits and neuropsychological measures.

The second phase of clinical recruitment to this study is currently underway and to date a further 80 parent/ child trios have been recruited. Ongoing efforts to secure research funding to continue recruitment are being made and it is planned to recruit the maximum number of trios possible in addition to other types of clinical cases such as individuals with cytogenetic abnormalities and extended pedigrees.

Obviously it is a priority to follow up the interesting findings on chromosome 2q. A microsatellite-based linkage disequilibrium experiment on chromosome 2q is currently in progress across a 30cM region defined by the linkage findings reported by IMGSAC, Buxbaum and Shao. This work was not completed in time to be presented here, but preliminary investigations have not identified further evidence of LD elsewhere across the region with the autism sample. Dense SNP mapping of the region around the associated haplotypes is planned. This will provide information regarding the underlying haplotypic structure in the region. With increased SNP density, haplotype tagging will be possible reducing the cost and effort involved in replicating the findings in a further sample. In addition, an attempted replication of the association findings with the D2S2077 marker and SNPs in the region is currently underway in a Portuguese autism sample. While the focus to date has been on the D2S2077 marker, it is also essential to follow up the association findings around D2S270. An LD experiment using SNP markers is currently in progress around this marker.

Efforts to narrow the insertion breakpoints on chromosome 9q are underway. While the chromosome 2q region was more interesting because of the findings from linkage studies there is still a possibility that the breakpoint on chromosome 9q might harbour a susceptibility gene for autism. Any gene that is disrupted on this chromosome will be subject to LD fine mapping in the autism sample. In addition, it was noted that the orientation of the chromosome 2q translocated region is reversed on chromosome 9q. Investigations of methylation are planned on both chromosome 2q and chromosome 9q as this shift in orientation may affect the expression of genes contained in the region.

A number of candidate gene studies are currently being conducted in this autism sample. These include genes involved in the dopaminergic and serotonergic systems (Conroy), replication studies of candidate genes where an association with autism has been reported, e.g., WNT2, Reelin, , GABRB3 and screening for mutations in $MeCP_2$

A number of collaborations have been established in addition to the Irish-Portuguese collaboration discussed above. Importantly a broad international collaboration has been established in autism genetics and the Irish Autism Genetics Research Group will be involved in this. This collaboration is a large international collaboration supported by the National Alliance for Autism Research (NAAR) and the National Institute of Health (NIH) in the United States. All groups that conducted genomewide linkage studies in autism are involved in Phase 1. This will involve genotyping all samples from these studies using a uniform set of markers for linkage analysis. The second phase will target follow up regions identified in the linkage analysis and will involve linkage disequilibrium mapping of these regions in both trios and cases and controls. The Irish sample will be utilized in the second phase of this collaboration. In addition, a cell line repository will be established at NIH consisting of cell lines obtained from all study subjects involved in this broad collaboration. This material will be made available after an initial holdback period to all researchers working in autism genetics research. This collaboration represents a very large autism sample. It is projected that 1000 affected sib-pairs will be included in phase 1 and at least 2000 trios in phase 2. This sample should have adequate power to identify regions of linkage to autism and subsequently narrow these regions down with linkage disequilibrium studies.

In summary, this work has outlined the establishment of a research group in autism genetics with the associated construction of a well-characterised autism sample. The findings of a candidate gene association study were presented in addition to exciting findings showing converging evidence for a susceptibility gene for autism on chromosome 2q. Several active collaborations are in progress and the author is currently pursuing program funding. The general aims for the future will include:

1. Continued recruitment to increase the sample size. This will increase the power of future genetic association investigations.
2. Recruitment of alternative family/ case types, e.g. sib-pairs, extended pedigrees, further cytogenetic abnormalities. This will provide the opportunity to adopt other approaches for the identification of autism susceptibility genes using for example linkage methods in genetically enriched families and physical methods in the characterisation of cytogenetic anomalies.

3. Continued development of collaborations for the purposes of replication of data and investigations in areas in which the group is not experienced (e.g. cytogenetics, functional studies, neuropsychological testing). Neuropsychological assessments will allow for the development of endophentoypes in autism, which may inform future genetic association investigations, e.g. tests of limbic function in risk haplotype carriers of the Integrin alpha 4 gene.
4. Expansion of the clinical data on the sample recruited, including further phenotypic measures, neuropsychological assessments and functional imaging. This will also aid endophenotype analysis and inform genetic association investigations.
5. Extension of investigations to broader phenotypes/ related disorders to facilitate quantitative trait types of analyses. This will allow for the investigation of aspects of the phenotype that are potentially normally distributed in the population, e.g. sociability and affords greater power on the one hand in addition to the opportunity to investigate the role of candidate genes in a specific aspect of the phenotype.
6. In addition to the follow up of the findings on chromosome 2q outlined in detail in Chapter 6, laboratory investigations will pursue other candidate genes and fine mapping of other regions of interest that have emerged from linkage studies. Autism is a polygenetic disorder and thus it will be necessary to investigate a range of genes in the disorder.

gamma-aminobutyric acid receptor subunit genes in autistic disorder. Am J Med Genet 96:43-8

Masi G, Cosenza A, Mucci M, De Vito G (2001) Risperidone monotherapy in preschool children with pervasive developmental disorders. J Child Neurol 16:395-400

McAlonan GM, Daly E, Kumari V, Critchley HD, van Amelsvoort T, Suckling J, Simmons A, Sigmundsson T, Greenwood K, Russell A, Schmitz N, Happe F, Howlin P, Murphy DG (2002) Brain anatomy and sensorimotor gating in Asperger's syndrome. Brain 125:1594-606

McCoy PA, Shao Y, Wolpert CM, Donnelly SL, Ashley-Koch A, Abel HL, Ravan SA, Abramson RK, Wright HH, DeLong GR, Cuccaro ML, Gilbert JR, Pericak-Vance MA (2002) No association between the WNT2 gene and autistic disorder. Am J Med Genet 114:106-9

McCracken JT, McGough J, Shah B, Cronin P, Hong D, Aman MG, Arnold LE, et al. (2002) Risperidone in children with autism and serious behavioral problems. N Engl J Med 347:314-21

McDougle CJ, Holmes JP, Carlson DC, Pelton GH, Cohen DJ, Price LH (1998) A double-blind, placebo-controlled study of risperidone in adults with autistic disorder and other pervasive developmental disorders. Arch Gen Psychiatry 55:633-41

McDougle CJ, Stigler KA, Posey DJ (2003) Treatment of aggression in children and adolescents with autism and conduct disorder. J Clin Psychiatry 64 Suppl 4:16-25

McGuffin P, Owen, M.J., O'Donovan, M.C., Thapar, A., Gottesman, I.I. (1994) Linkage and association. In: Freeman H, Pullen, I., Stein, G., Wilkinson, G. (ed) Seminars in Psychiatric Genetics. Gaskell, London, pp 55-65

Meyer GA, Blum NJ, Hitchcock W, Fortina P (1998) Absence of the fragile X CGG trinucleotide repeat expansion in girls diagnosed with a pervasive developmental disorder. J Pediatr 133:363-5

Miller MT, Stromland K, Gillberg C, Johansson M, Nilsson EW (1998) The puzzle of autism: an ophthalmologic contribution. Trans Am Ophthalmol Soc 96:369-85; discussion 385-7

Milner R, Campbell IL (2002) The integrin family of cell adhesion molecules has multiple functions within the CNS. J Neurosci Res 69:286-91

Moore SJ, Turnpenny P, Quinn A, Glover S, Lloyd DJ, Montgomery T, Dean JC (2000) A clinical study of 57 children with fetal anticonvulsant syndromes. J Med Genet 37:489-97

Murphy DG, Critchley HD, Schmitz N, McAlonan G, Van Amelsvoort T, Robertson D, Daly E, Rowe A, Russell A, Simmons A, Murphy KC, Howlin P (2002) Asperger syndrome: a proton magnetic resonance spectroscopy study of brain. Arch Gen Psychiatry 59:885-91

Murphy P, Hill R (1991) Expression of the mouse labial-like homeobox-containing genes, Hox 2.9 and Hox 1.6, during segmentation of the hindbrain. Development 111:61-74

Narita N, Kato M, Tazoe M, Miyazaki K, Narita M, Okado N (2002) Increased monoamine concentration in the brain and blood of fetal thalidomide- and valproic acid-exposed rat: putative animal models for autism. Pediatr Res 52:576-9

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

Nicolson R, Bhalerao S, Sloman L (1998) 47,XYY karyotypes and pervasive developmental disorders. Can J Psychiatry 43:619-22

Nielsen J, Christensen KR, Friedrich U, Zeuthen E, Ostergaard O (1973) Childhood of males with the XYY syndrome. J Autism Child Schizophr 3:5-26

Nielsen J WM (1991) Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. Human Genetics 87:81-83

Nurmi E, Bradford Y, Chen Y, Hall J, Arnone B, Gardiner M, Hutcheson H, Gilbert J, Pericak-Vance M, Copeland-Yates S, Michaelis R, Wassink T, Santangelo S, Sheffield V, Piven J, Folstein S, Haines J, Sutcliffe J (2001) Linkage disequilibrium at the Angelman syndrome gene UBE3A in autism families. Genomics 77:105-13

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

O'Brien EK, Zhang X, Nishimura C, Tomblin JB, Murray JC (2003) Association of specific language impairment (SLI) to the region of 7q31. Am J Hum Genet 72:1536-43

Olsson I, Steffenburg S, Gillberg C (1988) Epilepsy in autism and autisticlike conditions. A population-based study. Arch Neurol 45:666-8

Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U (1992) Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. Nat Genet 2:265-9

Pai GS, Rogers JF, Sommer A (1983) Identical multiple congenital anomalies/mental retardation (MCA/MR) syndrome due to del(2)(q32) in two sisters with intrachromosomal insertional translocation in their father. Am J Med Genet 14:189-95

Paschen W, Hedreen JC, Ross CA (1994) RNA editing of the glutamate receptor subunits GluR2 and GluR6 in human brain tissue. J Neurochem 63:1596-602

Peroutka SJ (1990) 5-Hydroxytryptamine receptor subtypes. Pharmacol Toxicol 67:373-83

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

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

Phillippe A, Martinec M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zapella M, Aschauer H, Van Malldergerme L, Penet C, Feingold J, Brice A, Leboyer M, Study atPARIS (1999) Genome-wide scan for autism susceptibility genes. Human Molecular Genetics 8:805-812

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

Pickles A, Starr E, Kazak S, Bolton P, Papanikolaou K, Bailey A, Goodman R, Rutter M (2000) Variable expression of the autism broader phenotype: findings from extended pedigrees. J Child Psychol Psychiatry 41:491-502

Pinkstaff JK, Detterich J, Lynch G, Gall C (1999) Integrin subunit gene expression is regionally differentiated in adult brain. J Neurosci 19:1541-56

Piven J, Palmer P (1997) Cognitive deficits in parents from multiple-incidence autism families. J Child Psychol Psychiatry 38:1011-21

Piven J, Palmer P (1999) Psychiatric disorder and the broad autism phenotype: evidence from a family study of multiple-incidence autism families. Am J Psychiatry 156:557-63

Povey S, Burley MW, Attwood J, Benham F, Hunt D, Jeremiah SJ, Franklin D, Gillett G, Malas S, Robson EB, et al. (1994) Two loci for tuberous sclerosis: one on 9q34 and one on 16p13. Ann Hum Genet 58 (Pt 2):107-27

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1-14

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

Quintana PJ, Neuwirth EA, Grosovsky AJ (2001) Interchromosomal gene conversion at an endogenous human cell locus. Genetics 158:757-67

Ramamoorthy S, Bauman AL, Moore KR, Han H, Yang-Feng T, Chang AS, Ganapathy V, Blakely RD (1993) Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. Proc Natl Acad Sci U S A 90:2542-6

Ramer JC, Mowrey PN, Robins DB, Ligato S, Towfighi J, Ladda RL (1990) Five children with del (2)(q31q33) and one individual with dup (2)(q31q33) from a single family: review of brain, cardiac, and limb malformations. Am J Med Genet 37:392-400

Ratcliffe S (1999) Long-term outcome in children of sex chromosome abnormalities. Arch Dis Child 80:192-195

Reed ML, Leff SE (1994) Maternal imprinting of human SNRPN, a gene deleted in Prader-Willi syndrome. Nat Genet 6:163-7

Rinehart NJ, Bradshaw JL, Moss SA, Brereton AV, Tonge BJ (2001) A deficit in shifting attention present in high-functioning autism but not Asperger's disorder. Autism 5:67-80

Risch N (2000) Searching for genes in complex diseases: lessons from systemic lupus erythematosus. J Clin Invest 105:1503-6

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

Ritvo ER, Freeman BJ, Scheibel AB, Duong T, Robinson H, Guthrie D, Ritvo A (1986) Lower Purkinje cell counts in the cerebella of four autistic subjects: initial findings of the UCLA-NSAC Autopsy Research Report. Am J Psychiatry 143:862-6

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. Am J Psychiatry 147:1614-21

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

Robinson A, Bender BG, Borelli JB, Puck MH, Salbenblatt JA, Winter JS (1986) Sex chromosomal aneuploidy: prospective and longitudinal studies. Birth Defects Orig Artic Ser 22:23-71

Robinson A, Bender BG, Linden MG, Salbenblatt JA (1990) Sex chromosome aneuploidy: the Denver Prospective Study. Birth Defects Orig Artic Ser 26:59-115

Robinson PD, Schutz CK, Macciardi F, White BN, Holden JJ (2001) Genetically determined low maternal serum dopamine beta-hydroxylase levels and the etiology of autism spectrum disorders. Am J Med Genet 100:30-6

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

Rodier PM, Ingram JL, Tisdale B, Nelson S, Romano J (1996) Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. J Comp Neurol 370:247-61

Rolf LH, Haarmann FY, Grotemeyer KH, Kehrer H (1993) Serotonin and amino acid content in platelets of autistic children. Acta Psychiatr Scand 87:312-6

Rossel M, Capecchi MR (1999) Mice mutant for both Hoxa1 and Hoxb1 show extensive remodeling of the hindbrain and defects in craniofacial development. Development 126:5027-40

Ruat M, Traiffort E, Leurs R, Tardivel-Lacombe J, Diaz J, Arrang JM, Schwartz JC (1993) Molecular cloning, characterization, and localization of a high-affinity serotonin receptor (5-HT7) activating cAMP formation. Proc Natl Acad Sci U S A 90:8547-51

Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murtha MT, Pendleton JW (1994) Evolution of Hox genes. Annu Rev Genet 28:423-42

Rutter M, Yule B, Quinton D, Rowlands O, Yule W, Berger M (1975) Attainment and adjustment in two geographical areas: III--Some factors accounting for area differences. Br J Psychiatry 126:520-33

Rutter ML (1970) Psycho-social disorders in childhood, and their outcome in adult life. J R Coll Physicians Lond 4:211-8

Schain R, Freedman D (1961) Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. J Paediatrics 58:315-320

Schanen C (2001) Rethinking the fate of males with mutations in the gene that causes Rett syndrome. Brain Dev 23 Suppl 1:S144-6

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

Schroer R, Phelan M, Michaelis R, Crawford E, Skinner S, Cuccaro M, Simensen R, Bishop J, Skinner C, Fender D, Stevenson R (1998) Autism and maternally derived aberrations of chromosome 15q. Am J Med Genet 76:327-36

Sham P, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann Hum Genet 59:323-36

Shao Y, Raiford K, Wolpert C, Cope H, Ravan S, Ashley-Koch A, Abramson R, Wright H, DeLong R, Gilbert J, Cuccaro M, Pericak-Vance M (2002a) Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. Am J Hum Genet 70:1058-61

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

Shimizu E, Tang YP, Rampon C, Tsien JZ (2000) NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. Science 290:1170-4

Sims MA, Field SD, Barnes MR, Shaikh N, Ellington K, Murphy KE, Spurr N, Campbell DA (2000) Cloning and characterisation of ITGAV, the genomic sequence for human cell adhesion protein (vitronectin) receptor alpha subunit, CD51. Cytogenet Cell Genet 89:268-71

Smalley S, Asarnow R, Spence M (1988) Autism and genetics: a decade of research. Arch. Gen. Psych. 45:953-961

Smalley SL, Tanguay PE, Smith M, Gutierrez G (1992) Autism and tuberous sclerosis. J Autism Dev Disord 22:339-55

Sparks BF, Friedman SD, Shaw DW, Aylward EH, Echelard D, Artru AA, Maravilla KR, Giedd JN, Munson J, Dawson G, Dager SR (2002) Brain structural abnormalities in young children with autism spectrum disorder. Neurology 59:184-92

Sparrow S.S BDACDV (1984) Vineland Adaptive Behaviour Scales. In: Service AG (ed). Australian Council for Educational Research

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

Steffenburg S (1991) Neuropsychiatric assessment of children with autism: a population-based study. Dev Med Child Neurol 33:495-511

Stewart DA, Bailey JD, Netley CT, Park E (1990) Growth, development, and behavioral outcome from mid-adolescence to adulthood in subjects with chromosome aneuploidy: the Toronto Study. Birth Defects Orig Artic Ser 26:131-88

Stromland K, Nordin V, Miller M, Akerstrom B, Gillberg C (1994) Autism in thalidomide embryopathy: a population study. Dev Med Child Neurol 36:351-6

Studer M, Lumsden A, Ariza-McNaughton L, Bradley A, Krumlauf R (1996) Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb-1. Nature 384:630-4

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

Sutherland GR (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. Science 197:265-6

Tadevosyan-Leyfer O, Dowd M, Mankoski R, Winklosky B, Putnam S, McGrath L, Tager-Flusberg H, Folstein SE (2003) A principal components analysis of the Autism Diagnostic Interview-Revised. J Am Acad Child Adolesc Psychiatry 42:864-72

Talebizadeh Z, Bittel DC, Miles JH, Takahashi N, Wang CH, Kibiryeva N, Butler MG (2002) No association between HOXA1 and HOXB1 genes and autism spectrum disorders (ASD). J Med Genet 39:e70

Tamimi R, Steingrimsson E, Copeland NG, Dyer-Montgomery K, Lee JE, Hernandez R, Jenkins NA, Tapscott SJ (1996) The NEUROD gene maps to human chromosome 2q32 and mouse chromosome 2. Genomics 34:418-21

Taysi K, Dengler DR, Jones LA, Heersma JR (1981) Interstitial deletion of the long arm of chromosome 2: case report and review of literature. Ann Genet 24:245-7

Thompson Ta (2001) Genetics in Medicine. W.B. Saunders, Philedelphia

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

Tueting P, Costa E, Dwivedi Y, Guidotti A, Impagnatiello F, Manev R, Pesold C (1999) The phenotypic characteristics of heterozygous reeler mouse. Neuroreport 10:1329-34

Uusitalo M, Heikkila M, Vainio S (1999) Molecular genetic studies of Wnt signaling in the mouse. Exp Cell Res 253:336-48

Veenstra-VanderWeele J, Kim SJ, Lord C, Courchesne R, Akshoomoff N, Leventhal BL, Courchesne E, Cook EH, Jr. (2002) Transmission disequilibrium studies of the serotonin 5-HT2A receptor gene (HTR2A) in autism. Am J Med Genet 114:277-83

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

Vu TH, Hoffman AR (1997) Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. Nat Genet 17:12-3

Wainwright BJ, Scambler PJ, Stanier P, Watson EK, Bell G, Wicking C, Estivill X, Courtney M, Boue A, Pedersen PS, et al. (1988) Isolation of a human gene with protein sequence similarity to human and murine int-1 and the Drosophila segment polarity mutant wingless. Embo J 7:1743-8

Walzer S, Bashir AS, Silbert AR (1990) Cognitive and behavioral factors in the learning disabilities of 47,XXY and 47,XYY boys. Birth Defects Orig Artic Ser 26:45-58

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

Wassink T, Piven J, Vieland V, Huang J, Swiderski R, Pietila J, Braun T, Beck G, Folstein S, Haines J, Sheffield V (2001) Evidence supporting WNT2 as an autism susceptibility gene. Am J Med Genet 8;105:406-13

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

Weaving LS, Williamson SL, Bennetts B, Davis M, Ellaway CJ, Leonard H, Thong MK, Delatycki M, Thompson EM, Laing N, Christodoulou J (2003) Effects of MECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. Am J Med Genet 118A:103-14

Weidmer-Mikhail E, Sheldon S, Ghaziuddin M (1998) Chromosomes in autism and related pervasive developmental disorders: a cytogenetic study. J Intellect Disabil Res 42 (Pt 1):8-12

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

Wing L (1976) Diagnosis, clinical description and prognosis. In: Early Childhood Autism. Pergamon, Oxford, pp pp 15-48

Wing L, Gould J (1979) Severe impairments of social interactions and associated abnormalities in children: epidemiology and classification. Journal of Autism and Developmental Disorders 9:11-29

Wolff DJ, Clifton K, Karr C, Charles J (2002) Pilot assessment of the subtelomeric regions of children with autism: detection of a 2q deletion. Genet Med 4:10-4

Wolpert C, Menold M, Bass M, Qumsiyeh M, Donnelly S, Ravan S, Vance J, Gilbert J, Abramson R, Wright H, Cuccaro M, Pericak-Vance M (2000) Three probands with autistic disorder and isodicentric chromosome 15. Am J Med Genet 96:365-72

Woodhouse W, Bailey A, Rutter M, Bolton P, Baird G, Le Couteur A (1996) Head circumference in autism and other pervasive developmental disorders. J Child Psychol Psychiatry 37:665-71

Woolf BC (1955) On estimating the relation between blood group and disease. Annals of Human Genetics (Lond) 19:251-253

World Health Organization W (1992) nternational Statistical Classification of Diseases and Related Health Problems, Geneva

Yan H, Papadopoulos N, Marra G, Perrera C, Jiricny J, Boland CR, Lynch HT, Chadwick RB, de la Chapelle A, Berg K, Eshleman JR, Yuan W, Markowitz S, Laken SJ,

Lengauer C, Kinzler KW, Vogelstein B (2000a) Conversion of diploidy to haploidy. Nature 403:723-4

Yan WL, Guan XY, Green ED, Nicolson R, Yap TK, Zhang J, Jacobsen LK, Krasnewich DM, Kumra S, Lenane MC, Gochman P, Damschroder-Williams PJ, Esterling LE, Long RT, Martin BM, Sidransky E, Rapoport JL, Ginns EI (2000b) Childhood-onset schizophrenia/autistic disorder and t(1;7) reciprocal translocation: identification of a BAC contig spanning the translocation breakpoint at 7q21. Am J Med Genet 96:749-53

Yirmiya N, Pilowsky T, Nemanov L, Arbelle S, Feinsilver T, Fried I, Ebstein R (2001) Evidence for an association with the serotonin transporter promoter region polymorphism and autism. Am J Med Genet 8;105:381-6

Young RS, Shapiro SD, Hansen KL, Hine LK, Rainosek DE, Guerra FA (1983) Deletion 2q: two new cases with karyotypes 46,XY,del(2)(q31q33) and 46,XX,del(2)(q36). J Med Genet 20:199-202

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

Zneimer S, Schneider N, Richards C (1993) In situ hybridization shows direct evidence of skewed X inactivation in one of monozygotic twin females manifesting Duchenne muscular dystrophy. Am J Med Genet 45:601-605

APPENDIX A

Information Sheets Regarding Autism Study

RESEARCH IN THE GENETICS OF AUTISM

Autism:
Autism is a condition that affects 1 in 1000 children every year. A paediatrician Leo Kanner first described it in the 1940s. Children with autism are usually affected before the age of three-and-half years. They typically have abnormalities in their speech, ability to communicate and behaviour. They are less socially interactive than other children and often have difficulties communicating their needs. They often engage in rituals and repetitive behaviour. Many autistic children have also learning difficulties although some may have normal intelligence.

Genetics and autism:
Since the 1970s it has been noted that autism tends to run in families. Family members of people with autism have a higher risk of having the condition. In addition to this it has been found that where identical twins are affected their co-twin has a much higher chance of being affected than another brother or sister. Because identical twins share the same pattern of genetic material this finding suggests strongly that there is a genetic influence on the development of autism.

Many conditions nowadays are reported to be inherited. In some cases the gene that causes the condition has been identified, e.g. Cystic Fibrosis, Huntington's Disease. But in most other inherited conditions, single genes have not been identified as causing the condition. In these cases it is thought that there are several different mechanisms that could contribute to the development of the condition. Most likely is that there is an interaction between a number of genes and environment which results in the condition. In this case we often try to identify genes that may make an individual susceptible to a condition, so-called *susceptibility* genes.

A number of studies have been conducted worldwide in the molecular genetics of autism. The results so far have indicated possible linkage between autism and regions on chromosomes 1, 2, 6, 7, 13 and 19 with the strongest evidence on chromosomes 2 and 7. The next step in this research is to narrow down the regions further before it is possible to identify genes within these regions. This can be difficult to do in populations where there is a large degree of racial diversity. We are at an advantage in Ireland because there is very little diversity in the Irish population and this increases our chances of finding something significant.

This is one of the main reasons why we have decided to do this research in Ireland now.

About our research project:
A team of researchers in Trinity College Dublin is conducting this project. The main researchers are:

- Dr. Louise Gallagher, Research Psychiatrist

- Ms Eleanor Meally, Research Assistant
- Prof. Michael Gill, Head of Department of Psychiatry
- Prof. Michael FitzGerald, Professor of Child Psychiatry.

The study began in March 1999 and we initially focussed on the Eastern Health Board region. We are now intending to extend the study nationwide. We intend to recruit 200 children with autism and their parents and are inviting all those parents who are interested to participate.

Participation in the project will involves the following

1. Diagnostic interview for one or both parents conducted in your home.
2. Diagnostic assessment of your child, usually conducted in the school and lasting approximately 30-45 minutes.
3. A brief physical examination of your child including examination of their skin to exclude medical causes of autism, height and head measurements.
4. A blood sample from the child and both parents. If a blood sample cannot be obtained, a cheek swab will be taken.

Genetic material from the blood samples will be examined first for Fragile X syndrome and chromosomal abnormalities, which are known to cause autistic like symptoms. Any child who has detectable genetic abnormalities will not be included further in the study. The results of these tests will be sent to your general practitioner. Genetic material of those included in the study will be examined for the presence of *susceptibility* genes. All the genetic material obtained for the study will be pooled so it will not be possible to relate information to a particular individual.

What we hope to achieve:
Identifying *susceptibility* genes in autism would help us to understand more about the condition and how it develops. This may help us in the future to develop preventative measures or treatments that best help the condition.

Obviously this goal is in the future and this research may not benefit your child. However by doing this research we may be able to help children in the future who are affected.

We are relying on people to volunteer to participate in the project therefore we would be very grateful to parents and children who would be willing to be involved. If you are interested in knowing more about this study, there is a name and address at the end that you could contact. Alternatively you can return the attached response form and Louise Gallagher or Eleanor Meally will contact you by telephone over the next few weeks.

Information back to you:
As parents who participate in the study will be keen to hear about the outcome of the research, we will send them a newsletter periodically to keep them informed of developments with our study and interesting results from other research groups working on the genetics of autism.
Thank you for taking the time to read this information leaflet.

APPENDIX A (contd)

Contact:

Eleanor Meally
Dept. of Psychiatry
Trinity Centre for Health Sciences
St. James' Hospital
Dublin 8

Tel: (01) 608 2144
E-mail: meallye@tcd.ie

RESPONSE FORM

Yes - I am interested in hearing more about this research. Eleanor Meally may contact me at the following address/ telephone number:

NAME: _____

ADR: _____

_____

TEL: _____

Best time for telephone contact: _____

APPENDIX B

PHONE-CALL INTERVIEW: INITIAL CONTACT

Name:

G.P (Name and address):

Date of contact:

Researcher:

Information about case:

Childs name:

Age:

Date of birth:

School/ Day placement:

Principal:

Resource teacher:

APPENDIIX B (contd.)

CHECKLIST

Background to study outlined	Y/N
Language: Norn-verbal Words (give example) Phrases (give example) Age at first words: Ages at phrase speech:	Y/N
Medical history: Does he/ she have any of the following? 1.. physical disabilities 2.. epilepsy 3.. exposure to rubella during pregnancy 4.. Was he/ she born prematurely and if so at how many weeks of pregnancy 5.. Does he/ she have any other medical disorders e.g. cerebral palsy, tuberous sclerosis, neurofibromatosis 6.. Does he/ she have any history of meningitis or a head injury.	

APPENDIX B (contd.)

<u>Pregnancy history:</u> 1. Were there any complications during the pregnancy, e.g. bleeding, baby not growing. 2. Did you take any medications? 3. How much alcohol (if any) did you consume during your pregnancy 4. Did you smoke, if so how much? 5. Was the delivery normal, if not did you have a forceps/ vacuum delivery or c-section? 6. Birth weight? 7. Was he/ she well after he/ she was born, e.g. did he go to the ward with you or was he in special care for a while <u>Family history:</u> 1. Are there any other family members affected? If yes: • Name • Relationship to _____ • Has this person(s)received a formal diagnosis?	

APPENDIX B (contd.)

<u>Investigations:</u> Has he/ she had the following investigations: • Fragile X • Chromosomal analysis • Psychological assessment • Paediatric assessment • Metabolic screen • CT scan • EEG	
Is he/ she on medication? If yes, what medication?	
<u>Willingness to participate:</u> • Are both parents willing to give consent to participate? • Will both parents give blood samples • Will it be possible to obtain a sample of blood from _____	

DATE ARRANGED TO VISIT:_____

APPENDIX C: CONSENT TO PARTICIPATE

CONSENT FORM

Title of project:
Molecular genetics study in Autism. Candidate gene and Linkage Disequilibrium Screen.

Researchers:
Dr. Louise Gallagher, Eleanor Meally, Dr. Michael Gill, Prof. Michael Fitzgerald. Trinity Centre for Health Sciences, St. James's Hospital, Dublin 8. Tel (01) 608 2241.

Information about the project:
Autism is a developmental disorder of childhood that affects 5 in every 10,000 children. Children and adults with autism show abnormalities in development of speech and language and problems with behaviour and social skills development. Although it is a rare disorder and the risk of brothers and sisters being affected is also low, the risk is substantially higher than in the general population. Many studies of families and twins affected with autism suggest that there is a major genetic component that may determine which individuals develop autism. The studies conducted to date suggest that there are probably many genes that are involved in the condition. These genes are susceptibility genes that may predispose the individual to the development of autism. The discovery of these genes may also help us to understand more about what causes autism. A number of research groups worldwide are currently involved in the search for genes involved in the disorder. Molecular genetic techniques are used to identify a possible pattern of genes that may be involved. Our research will look certain regions, which some studies have identified as possibly being involved in autism. To conduct this study we need to obtain blood samples or cheek swabs to obtain the genetic material. We need samples from the affected child and both parents.

Your involvement in this study will not influence the care provided to your child in any way.

Procedure:
There are four parts to your involvement in this study.

1. Parents will be interviewed with a questionnaire devised to make the diagnosis of autism. This questionnaire may take about three hours to apply.
2. Your child will be observed by a trained researcher to further confirm the diagnosis.
3. Your child will have a brief physical examination by a doctor or nurse. This is to rule out any medical cause of autism.
4. A blood sample will be obtained by a doctor or nurse from both your child and both parents. This is a very safe procedure but may cause some localised bruising or, rarely, infection.

In the event that blood samples cannot be taken, a cheek swab will be obtained. This involves rubbing the inside of the mouth with a spatula. It is a very safe procedure.

This process should take four hours in total. It is possible that in some cases it may take longer. There will also be coffee breaks during this time to ensure that you and your child don't find it too tiring.

Parents and children may choose to withdraw from the study at any time. This will not effect the standard of care provided to them in any way.

Parents' consent:
I agree to participate in this study and to provide a blood sample. I also agree to my child taking part in this study. The research has been fully explained to me by Dr. Louise Gallagher/Eleanor Meally I understand that it may not bring any immediate or direct benefit to my child or to me personally.

Signature: _____

Signature: _____

Where child is considered competent:

Child's consent:
I agree to have a blood sample taken from me. The research has been explained to me by Dr. Louise Gallagher / Eleanor Meally.

Signature: _____

Witnessed by:

Signature: _____

APPENDIX D: CONSENT TO VIDEOTAPING

GENETICS STUDY IN AUTISM:

CONSENT TO AUDIO-/VIDEO-TAPING

The assessments carried out as part of this research in autism are normally recorded either by audio-/videotaping. This includes your interview and your child's assessment. The purpose of this is to ensure that the researcher who carries out the assessment is doing it properly. The tapes may also be used on occasion to teach other researchers how to do these assessments. Only researchers working on the autism research in the Department of Psychiatry in Trinity College Dublin and our collaborators will have access to the tapes. Although you/your child will appear on the videotape, no other identifying information will appear.

If you do not wish for either yourself or your child to be audio-/ videotaped you may still participate in the study. You may also choose between audio- and videotaping, as you prefer.

I consent to either audio- or videotaping of myself/my child during my/his/her assessment/interview. I would prefer _____
(Please state preference here)

Signed_____

Signed_____

I consent to the use of the audio/videotape for training purposes.

Signed_____

Signed_____

APPENDIX E: ADI-R data

STUDY NO.	AGE	SEX	V/NV	S/I	COMM	R/B	D/A<3y	Delay (S/W)	Delay (P/s)	Item 19	Dx<12y
AS001	68	F	V	21	16	1	4	No	Yes	0	Yes
AS002	68	M	V	19	14	3	5	Yes	Yes	0	Yes
AS003	199	F	V	27	18	8	5	Yes	Yes	0	Yes
AS004	71	M	V	22	18	8	2	No	No	0	Yes
AS005	53	M	V	9	12	3	4	No	Yes	0	Yes
AS006	104	M	V	20	17	5	1	No	No	0	Yes
AS007	265	M	V	20	17	2	2	No	No	0	Yes
AS008	59	M	V	8	8	3	3	No	No	0	Yes
AS009	107	M	V	16	14	7	5	Yes	Yes	0	Yes
AS011	60	M	V	18	16	4	4	No	Yes	0	Yes
AS012	132	F	NV	27	11	8	5	Yes	Yes	2	Yes
AS014	93	M	V	18	18	3	4	No	Yes	0	Yes
AS015	312	M	V	28	18	9	5	Yes	Yes	0	Yes
AS017	372	F	V	23	15	8	4	No	Yes	0	Yes
AS019	408	M	NV	27	14	4	3	Yes	Yes	2	Yes
AS020	85	M	V	25	20	7	5	Yes	Yes	0	Yes
AS021	65	M	V	21	19	4	4	No	Yes	0	Yes
AS022	105	F	NV	24	14	6	3	No	No	2	Yes
AS024	91	M	NV	24	14	7	3	No	No	0	Yes
AS025	76	M	NV	23	14	4	3	No	Yes	2	Yes
AS028	103	M	V	13	8	3	3	No	Yes	1	Yes
AS029	180	M	NV	20	12	6	5	No	No	0	Yes
AS030	73	M	V	16		4	5	yes	yes	2	yes
AS031	84	M	NV	27	14	5	3	yes	yes	0	yes
AS032	73	M	V	21	19	10	3	Yes	Yes	1	Yes
AS033	107	M	V	23	18	6	5	No	No	0	Yes
AS035	138	F	V	29	15	6	3	yes	yes	0	Yes
AS036	61	M	V	17	15	7	3	No	No	0	Yes
AS037	116	F	NV	28	14	6	4	No	Yes	0	Yes

APPENDIX E (contd.)

STUDY NO.	AGE	SEX	V/NV	S/I	COMM	R/B	D/A<3y	Delay (S/W)	Delay (P/s)	Item 19	Dx<12y
AS038	68	M	V	16	17	9	4	Yes	Yes	2	Yes
AS040	108	M	NV	21	14	5	5	Yes	No	0	Yes
AS041	168	M	NV	25	14	7	4	Yes	Yes	2	Yes
AS042	240	F	V	27	21	3	5	No	Yes	2	Yes
AS043	60	M	V	19	16	9	4	Yes	Yes	0	Yes
AS044	75	M	V	24	14	4	3	No	Yes	0	Yes
AS045	324	M	V	22	20	5	3	Yes	Yes	0	Yes
AS046	65	M	NV	24	13	4	5	No	No	0	Yes
AS047	85	M	V	17	19	6	5	Yes	Yes	2	Yes
AS048	53	M	NV	15	8	8	4	Yes	Yes	0	Yes
ÀS049	56	M	NV	15	12	3	5	Yes	Yes	1	Yes
AS050	56	M	NV	15	13	3	5	Yes	Yes	1	Yes
AS051	93	M	V	22	24	10	3	Yes	Yes	2	Yes
AS052	172	M	V	27	20	7	2	No	No	0	Yes
AS053	249	M	V	17	14	6	3	No	Yes	0	Yes
AS054	166	M	NV	29	13	6	5	No	Yes	0	Yes
AS055	71	M	V	22	16	10	5	Yes	Yes	2	Yes
AS056	141	M	V	27	21	3	4	Yes	Yes	0	Yes
AS057	124	M	V	21	22	6	2	No	Yes	0	Yes
AS058	124	F	NV	28	13	6	2	No	Yes	0	Yes
AS059	123	M	V	24	21	6	4	No	Yes	2	Yes
AS060	94	M	NV	28	13	4	5	Yes	Yes	0	Yes
AS062	221	M	V	24	11	3	5	Yes	Yes	2	Yes
AS063	299	M	V	26	23	7	4	Y	Yes	0	No
AS064	143	F	NV	24	12	4	5	No	Yes	0	Yes
AS065	93	M	V	25	24	8	5	Yes	Yes	2	Yes
AS066	71	M	V	20	22	8	4	Yes	Yes	0	Yes
AS067	61	M	NV	23	12	7	5	No	Yes	0	Yes
AS068	70	M	V	12	23	6	4	Yes	Yes	2	Yes
AS069	206	M	NV	23	14	6	3	Yes	No	0	Yes

APPENDIX E (contd.)

STUDY NO.	AGE	SEX	V/NV	S/I	COMM	R/B	D/A<3y	Delay (S/W)	Delay (P/s)	Item 19	Dx<12y
AS070	164	M	V	22	19	7	4	Yes	Yes	2	No
AS071	47	F	NV	20	13	8	5	No	Yes	0	No
AS075	113	M	V	27	14	5	5	Yes	Yes	1	Yes
AS076	98	M	V	20	19	8	5	Yes	Yes	1	Yes
AS077	49	F	V	22	19	7	5	Yes	Yes	0	Yes
AS078	55	M	NV	13	7	4	4	Yes	Yes	0	Yes
ASO80	117	M	V	10	18	4	4	No	Yes	1	Yes
AS081	104	M	V	23	19	7	3	No	Yes	1	Yes
AS082	59	F	V	16	19	5	5	No	Yes	0	Yes
AS083	65	F	V	28	14	4	5	Yes	Yes	0	Yes
AS084	70	M	V	15	18	6	5	Yes	Yes	2	Yes
AS085	83	M	V	19	25	6	4	Yes	Yes	0	Yes
AS086	107	F	V	22	20	4	4	No	Yes	0	Yes
AS087	324	M	NV	24	14	5	5	No	Yes	0	Yes
AS088	99	M	V	16	12	3	3	Yes	Yes	2	No
AS089	127	M	V	10	10	3	4	No	Yes	0	Yes
AS090	162	M	V	19	11	4	5	No	Yes	0	Yes
AS091	117	F	V	22	15	7	5	Yes	Yes	0	Yes
AS092	197	F	V	26	12	4	5	Yes	Yes	0	Yes
AS093	273	F	NV	18	15	5	3	Yes	Yes	0	Yes
AS094	54	M	NV	25	12	7	5	Yes	Yes	2	Yes
AS095	97	M	V	15	12	7	3	Yes	Yes	2	Yes
AS096	145	M	V	23	18	7	3	No	No	0	Yes
AS097	80	M	V	10	19	9	4	No	No	0	Yes
AS098	324	F	NV	26	14	4	5	No	Yes	0	Yes
AS099	60	F	NV	23	13	3	5	Yes	Yes	2	Yes
AS101	70	M	V	15	11	5	4	Yes	Yes	2	Yes
AS102	170	F	V	15	13	3	5	Yes	Yes	0	Yes
AS103	116	M	V	20	14	4	3	Yes	Yes	0	Yes
AS104	193	M	NV	22	10	3	2	No	No	0	Yes
AS105	119	M	V	22	15	8	2	No	No	0	Yes
AS106	58	M	V	21	13	5	3	No	Yes	0	Yes

APPENDIX E (contd.)

STUDY NO.	AGE	SEX	V/NV	S/I	Comm	R/B	D/A<3y	Delay (S/W)	Delay (P/s)	Item 19	Dx<12y
AS107	63	M	V	23	19	7	5	Yes	Yes	0	Yes
AS108	58	F	NV	20	12	4	4	Yes	Yes	1	Yes
AS109	250	M	V	28	21	4	4	No	Yes	0	No
AS110	115	M	V	30	20	4	5	Yes	Yes	0	Yes
AS111	57	M	NV	23	13	4	5	Yes	Yes	2	Yes
AS112	50	M	V	19	12	5	5	Yes	Yes	0	Yes
AS114	114	F	NV	24	13	5	5	Yes	Yes	2	Yes
AS115	79	M	V	23	16	7	5	Yes	Yes	1	Yes
AS116	151	M	V	26	15	3	5	Yes	Yes	0	Yes

V/NV – Verbal/ Non-verbal

S/I – ADI algorithm score for abnormalities in social interaction

Comm – ADI algorithm score for communication abnormalities

R/B – ADI algorithm score for restrictive and repetitive behaviours

D/A < 3 – ADI algorithm score for the presence of developmental abnormalities before 3 years

Delay (S/W) – Presence of delay in onset of single words

Delay (P/S) – Presence of delay in onset of phrase speech

Dx<12y – Occurrence of autism diagnosis before the age of 12 years

APPENDIX F: ADOS-G data

STUDY NO	AGE (MTHS)	MODULE	COMM	SOCIAL	COMM+ SOCIAL	PLAY/ IMAG.	REP. BEHAV.	ADOS DX
AS001	76	1	9	14	23	4	3	Autism
AS002	65	1	7	9	16	3	3	Autism
AS003	199	4	5	12	17	2	0	Autism
AS004	60	2	8	12	16	2	1	Autism
AS005	60	1	5	11	16	2	2	Autism
AS006	86	3	5	9	14		1	Autism
AS007	265	4	3	6	9	1	1	PDD/NOS
AS008	63	2	6	10	16	1	1	Autism
AS009	76	3	6	13	19	2	4	Autism
AS011	62	1	5	8	13	4	3	Autism
AS012	138	1	6	12	18	3	2	Autism
AS014	90	3	5	10	15	2	3	Autism
AS015	Non-compliant/ assessment not completed							
AS017	373	1	4	12	16	1	0	Autism
AS019	72	1	4	11	15	4	3	Autism
AS020	89	3	3	7	10	2	3	Autism
AS021	70	3	4	4	8	1	3	PDD/NOS
AS022	74	1	7	14	21	4	5	Autism
AS023	97	3	2	4	6	0	1	PDD/NOS
AS024	96	1	6	11	17	4	3	Autism
AS025	78	1	6	12	18	2	5	Autism
AS028	105	3	2	7	9	1	1	PDD/NOS
AS029	Non-compliant/ assessment not completed							
AS030	73	2	4	11	15	1	6	Autism
AS031	84	1	7	11	18	3	4	Autism

APPENDIX F (contd.)

STUDY NO	AGE (MTHS)	MODULE	COMM	SOCIAL	COMM+ SOCIAL	PLAY/ IMAG.	REP. BEHAV.	ADOS DX
AS032	77	2	7	10	17	1	2	Autism
AS033	75	3	5	6	11	1	2	Autism
AS035	139	1	9	9	18	4	4	Autism
AS036	62	2	6	12	18	1	2	Autism
AS037	119	1	7	10	17	4	4	Autism
AS038	65	3	4	6	10	1	1	Autism
AS040	109	1	8	13	21	4	6	Autism
AS041	169	1	4	9	13	4	2	Autism
AS042	241	2	8	9	17	1	3	Autism
AS043	64	2	10	14	24	2	5	Autism
AS044	77	2	7	11	18	2	4	Autism
AS045	324	1	7	10	17	4	3	Autism
AS046	66	1	6	13	19	4	4	Autism
AS047	86	2	9	13	22	2	5	Autism
AS048	54	1	3	8	11	3	2	PDD/NOS
ÀS049	55	1	7	9	16	3	2	Autism
AS050	59	1	7	12	19	4	5	Autism
AS051	94	3	7	11	18	1	3	Autism
AS052	173	4	5	6	11	1	4	Autism
AS053	251	4	4	11	15	1	2	Autism
AS054	167	1	5	8	13	4	1	Autism
AS055	75	3	3	13	16	2	1	Autism
AS056	143	3	3	9	12	2	2	Autism
AS057	126	3	4	5	9	1	2	Autism
AS058	125	1	5	11	16	3	2	Autism
AS059	125	3	3	8	11	0	1	Autism
AS060	95	1	7	12	19	4	6	Autism
AS062	221	4	4	7	11	1	1	Autism
AS063	300	4	5	11	16	2	3	Autism

APPENDIX F (contd.)

STUDY NO	AGE (MTHS)	MODULE	COMM	SOCIAL	COMM+ SOCIAL	PLAY/ IMAG.	REP. BEHAV.	ADOS DX
AS064	144	1	3	10	13	3	2	Autism
AS065	104	3	2	7	9	1	1	PDD/NOS
AS066	72	2	3	6	9	0	2	PDD/NOS
AS067	64	1	10	13	23	3	5	Autism
AS068	101	3	3	6	9	1	1	Autism
AS069	207	1	3	10	13	3	2	PDD/NOS
AS070	167	4	5	11	16	2	0	Autism
AS071	48	1	5	9	14	1	1	Autism
AS075	119	1	6	11	17	3	5	Autism
AS076	88	1	8	13	21	2	4	Autism
AS077	48	1	6	7	13	2	1	Autism
AS078	56	1	5	9	14	5	5	Autism
ASO80		3	4	4	8	1	1	PDD/NOS
AS081	105	1	5	9	14	4	5	Autism
AS082	61	1	9	13	22	4	3	Autism
AS083	67	1	7	12	19	3	4	Autism
AS084	71	1	6	8	14	4	4	Autism
AS085	86	3	4	5	9	1	1	PDD/NOS
AS086	109	2	6	8	14	1	3	Autism
AS087	326	1	7	9	16	1	0	Autism
AS088	101	3	2	7	9	1	3	PDD/NOS
AS089	129	3	2	5	7	1	2	PDD/NOS
AS090	163	4	6	12	18	2	0	Autism
AS091	117	2	8	9	17	1	1	Autism
AS092	191	3	6	12	18	2	1	Autism
AS093	65	1	4	11	15	3	4	Autism
AS094	62	1	8	13	21	4	3	Autism
AS095	98	3	4	6	10	1	3	Autism
AS096	146	3	2	6	8	1	1	PDD/NOS

APPENDIX F (contd.)

STUDY NO	AGE (MTHS)	MODULE	COMM	SOCIAL	COMM+ SOCIAL	PLAY/ IMAG.	REP. BEHAV.	ADOS DX
AS097	81	3	2	6	8	1	1	PDD/NOS
AS098	277	1	6	13	19	4	3	Autism
AS099	68	1	7	12	19	4	3	Autism
AS101	76	2	5	6	11	1	3	Autism
AS103			Non-compliant/ assessment not completed					
AS104	202	1	6	12	18	3	0	Autism
AS105	123	3	3	4	7	1	1	PDD/NOS
AS106	62	2	5	7	12	2	2	Autism
AS107	66	1	10	9	19	3	1	Autism
AS108	60	1	5	14	19	4	5	Autism
AS109	250	4	5	9	14	2	0	Autism
AS110	118	2	8	10	18	1	2	Autism
AS111	59	1	8	12	20	4	5	Autism
AS112	51	1	4	9	13	3	2	Autism
AS114	116	1	7	10	17	2	3	Autism
AS115	80	1	8	14	22	4	4	Autism
AS116	153	4	5	9	14	1	1	Autism

COMM - Communication abnormalities score on ADOS algorithm

SOCIAL - Social interaction abnormalities score on ADOS algorithm

COMM+SOCIAL - Combined score for communication and social scores

PLAY/ IMAG - Score for abnormalities in play (Module 1/2) or imagination (Module 3/4) on ADOS algorithm

REP. BEHAV - Score for presence of restrictive or repetitive behaviours on the ADOS algorithm

ADOS DX - Diagnosis generated from scores on ADOS algorithm

APPENDIX G: Excluded cases

STUDY NUMBER	SEX	DIAGNOSIS	REASON FOR EXCLUSION
AS010	Male	Autism	Prematurity
AS013	Male	PDD	Diagnosis
AS016	Male	Autism	One parent deceased
AS018	Male	ADI-R/ ADOS-G not completed	Withdrew
AS026	Male	PDD	Diagnosis
AS027	Male	Autism	Cytogenetic abnormality
AS034	Male	Autism spectrum	47XYY
AS039	Female	PDD	Diagnosis
AS061	Male	PDD	Diagnosis
AS072	Male	PDD	Diagnosis
AS073	Male		Withdrew
AS074	Male	Autism	IQ< 35
AS079	Male	Autism	Withdrew
AS100	Male	Query autism	IQ<20
AS102	Female	PDD	Diagnosis