



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.

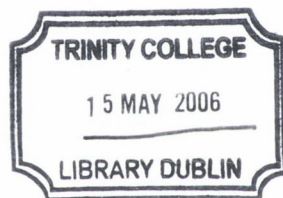
A Candidate Gene Search for Autism

Judith Conroy



A thesis submitted to the University of Dublin
for the degree of Doctor of Philosophy

Department of Psychiatry
University of Dublin
Trinity College – September 2005



THESIS
7791

Declaration

I hereby certify that this thesis has not been previously submitted for examination to this or any other university. The work described herein has been carried out by the author alone, except where otherwise stated.

This thesis may be made available for consultation within for consultation within the university library and may be photocopied or loaned to other libraries for the purposes of consultation.

Signed: Judith Conway 15/12/06

Summary

This research involved a candidate gene search for autism in the Irish population. It sought to identify candidate genes based on information from linkage studies, reports of chromosomal abnormalities, animal studies and clinical observations, and previous reports of genetic association in other samples diagnosed with autism. Two replication studies were undertaken in genes from the serotonin system. Associations with the long and short promoter variants have been reported in samples with autism, as have no associations¹⁻⁶. In the Irish population there was increased inheritance of the short promoter allele (p value = 0.03), in addition to a number of haplotypes across the gene. However a joint-analysis of the promoter allele highlights the possibility that this, and others may be false positive results. Another serotonin related gene, tryptophan hydroxylase 2, was also tested for association. Association with the A allele of marker rs1117900 was found. Further replications in other samples are required to evaluate this genes contribution to the development of autism.

A genetic and biochemical study of the dopamine- β -hydroxylase gene was carried out. Proband with autism were found to have significantly decreased DBH activity in comparison to their parents (p value = 16.4×10^{-6}) and a previously reported control sample (p value = 1.97×10^{-8}). Transmission disequilibrium testing did not reveal any over-transmission of “low activity” variants from parents to probands. Regression analysis also indicated that the variants analysed contributed less to DBH activity in children in comparison to their parents. The heritability of DBH activity was also ~50% lower than previous estimates^{8,9}. This leads to the possibility that there may be other genes or environmental factors contributing to the observed lower DBH activity levels.

Following the identification of a 2q:9q translocation in a patient within this study, and a subsequent fine-mapping linkage disequilibrium experiment, the integrin- α 4 gene was selected for fine-mapping. TDT and haplotype transmission testing found an increased transmission of a number of markers surrounding exons 16 and 17 (rs3770112-rs1551031

haplotype, OR = 1.92, p value = 0.0025). These exons were screened for mutations, leading to the identification of a variant (rs12690517) within a splice site sequence. Preferential transmission of the C allele to affected probands was observed (OR = 1.83, p value = 0.0046). An attempted replication in two samples, an AGRE and Vanderbilt sample, was undertaken. In the AGRE sample, the same two marker haplotype (rs3770112-rs1551031) was observed to be over-transmitted to affected probands (p value = 0.05), but in the Vanderbilt sample there was over-transmission of a number of different alleles / haplotypes. It is possible that within the *ITGA4* gene, there are a number of variants leading to the development of autism. A joint analysis of all three samples revealed an increased OR associated with the C allele of marker rs3770112, although this OR of 1.19 (p value = 0.0295) shows that inheritance of this variant would play a minor role in the development of autism.

Analysis of 4 genes in the Wnt pathway was carried out. A number of significant associations were observed with the *FRZB* gene (rs288326 p value = 0.04566, rs2242070 p value = 0.0072), the *DVL* gene (rs307354 p value = 0.021). There were also trends towards significance observed with markers in the *WNT2* gene (rs2024233 p value = 0.0556) and the *EN2* gene (rs3757846 p value = 0.107). These preliminary results suggest that the WNT pathway may play a role in autism. A large number of other genes in this pathway remain to be tested.

This thesis presents evidence supporting the role of genetics in the development of the autism. A number of candidate genes, as above, have shown association with autism. Eight, out of a total of fifty-four, variants showed significant association with autism. This was significantly greater than the number expected by chance given a false positive rate of 5% ($\chi^2 = 10.4$, p value = 0.0013). One would expect that at least some of the genes reported here do play a role in the development of autism. Further investigations following these important findings are ongoing.

Acknowledgements

I would like to thank Prof. Michael Gill and Dr. Louise Gallagher for giving me the opportunity to undertake this research. Their support and enthusiasm was especially important towards the latter stages, particularly while I was writing the thesis. Also, the constant day-to-day advice, support (emotional and computer) and humour of Dr. Ricardo Segurado, helped keep what little was left of my mind. Thanks also to Lynne for the constant companionship over the past year. Your frequent injections of enthusiasm regarding autism research always came at the right time.

Thanks to the other members of the lab, Dr Ziarah (the Great), Dr David Lambert, Dr. Naomi Lowe, Karen Shaheen, Dr Brian McEvoy and Dr Katharina Domschke, for making the lab such a great place to work. Also importantly are the members of our second lab, Dr Derek Morris, Geraldine, Irish Kevin, Sarah, Eleanor and Niamh (and her alter ego Stiamh), and the neuropsychiatric research group as a whole. My thanks also goes to Kevin McGee, who was always there when I needed to complain to someone, and who always cheered me up. My appreciation also goes to the individuals and groups who have collaborated with work on this project. Without your help, Dr Sean Ennis, Dr George Anderson, Dr. James Sutcliffe and Dr Marylyn DeRiggi-Ritchie, the results of this research would not have been half as interesting.

To Dave, Paul, Sue, the entire Genetics Department, and especially Brenda, thanks for helping to solve all the mini day-to-day crisis that occurred throughout my PhD, and for being so friendly and never seeming to mind that these things happened so frequently.

This project would never have been possible without the generous donation of the families who contributed their time, effort and DNA to the study. The entire group appreciates your efforts.

Most importantly, a very special thanks goes to my parents, Colette and Seamus, who have put up with the highs and the lows of having someone in the house trying to complete a PhD. To my sisters, Charlotte and Genevieve, for humorously reminding me that life does exist beyond the lab, thank you also.

Statement of Work

This work was the product of the autism genetics group, Department of Psychiatry, Trinity College Dublin. The clinical recruitment and collection of samples for DNA were performed by Dr. Louise Gallagher, Geraldine Kearney (research nurse) and Eleanor Meally (research nurse).

DNA extraction and quantification was performed by the author and Dr. Louise Gallagher. All genotyping and analysis for the Serotonin Transporter, TPH2, DBH, DVL, EN2 genes and one SNP in the ITGA4 gene (rs12690517) was performed by the author. Genotyping of the ITGA4, WNT2 and FRZB SNPs were carried out commercially by K-Bioscience Genotyping Services. The analysis of all data, with the exception of the gene-gene interactions, was undertaken by the author.

DBH activity was measured by Dr. George Anderson, Yale Child Study Centre, New Haven, Connecticut, USA.

Initial studies undertaken (including mapping and the highlighting of the ITGA4 gene as a potential candidate gene) on chromosome 2q and the associated analysis were performed by Dr. Sean Ennis, Lecturer in Medical Genetics at the National Centre for Medical Genetics and Dr. Louise Gallagher, Co-ordinator of the Autism Research Group, Departments of Genetics and Psychiatry, Trinity College Dublin.

Publications

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* 2004 9(6): 587-93.

Abbreviations

5-HIAA	5-hydroxyindoleacetic acid
5-HT	5-hydroxytryptamine , serotonin
5-HTP	5-hydroxytryptophan
5-HTT	Serotonin transporter
A	Adenosine
ADHD	Attention deficit hyperactivity disorder
ADI-R	Autism Diagnostic Interview (Revised)
ADOS-G	Autism Diagnostic Observation Schedule-Generic (ADOS-G)
AGRE	Autism Genetic Resource Exchange
AS	Angelman syndrome
ASP	Affected sib pairs
bp	Base-pair
C	Child/Proband (in relation to sample structure)
C	Cytosine (in relation to DNA chemical structure, mutations or chemical reactions)
cM	Centimorgans
CNS	Central Nervous System
COMT	Catechol-O-methyltransferase
CSF	Cerebrospinal fluid
dATP	Deoxy-adenosine triphosphate
DBH	Dopamine- β -hydroxylase
dCTP	Deoxy-cytosine triphosphate
ddNTPs	Dideoxy nucleotide triphosphate
del	Deletion
df	Degrees of freedom
dGTP	Deoxy-guanosine triphosphate
dHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
dNTP	Deoxy nucleotide triphosphate
<i>Dsh</i>	Dishevelled (Drosophila gene)
DSM-IV	Diagnostic and Statistical Manual, 4 th Edition
DSP	Discordant sib pairs
dTTP	Deoxy thymidine triphosphate
DVL	Dishevelled (Human gene)
Dvl	Dishevelled (Mouse gene)
E6-AP	E3 ubiquitin protein ligase
EN2	Engrailed 2
F	Father

FISH	Fluorescent In-Situ Hybridisation
FMR-1	Fragile X mental retardation gene 1
fMRI	Functional magnetic resonance imaging
FRAXA	Fragile X site A
FRAXE	Fragile X site E
Frzb	Frizzled Related Protein
Fz	Frizzled
GDB	Genome Database
GRR	Genotype Relative Risk
HHRR	Haplotype Based Haplotype Relative Risk
HTR2A	Serotonin receptor 2A
htSNP	Haplotype tagging SNP
ICD-10	International Classification of Diseases, 10 th Edition
IMGSAC	International Molecular Genetics Consortium in Autism
In/Del	Insertion/Deletion
IQ	Intelligent quotient
ITGA4	Integrin-a4
Kb	Kilobase
LD	Linkage disequilibrium
LOD	Log of the odds
M	Mother
MAO	Monoamine oxidase
Mb	Megabase
MDR	Multifactor dimensional reduction
MeDa	Mesencephalic dopaminergic
MLS	Maximum LOD scores
MMLS	Maximum multipoint lod score
MMR	Measles-Mumps-Rubella
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MZ	Monozygotic
NCBI	National Centre for Biotechnology Information
NE	Norepinephrine
NF	Neurofibromatosis
OMIMI	Online Mendelian Inheritance in Man
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PDD	Pervasive Developmental Disorder
PDD-NOS	Pervasive Developmental Disorder - not otherwise specified
PDT	Pedigree Disequilibrium Test
PET	Positron emission tomography

PIC	Polymorphic information content
PKU	Phenylketonuria
Pu	Purine
PWS	Prader-Willi syndrome
RAY1	Suppressor of tumorigenicity 7
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
SAP	Shrimp Alkaline Phosphatase
SNP	Single nucleotide polymorphism
SSCP	Single strand conformational polymorphism
SSRI	Selective serotonin reuptake inhibitor
T	Thymidine
TDO2	Tryptophan 2,3-dioxygenase
TDT	Transmission Disequilibrium Test
TPH	Tryptophan hydroxylase
TSC1	Hamartin gene
TSC2	Tuberin gene
UBE3A	Ubiquitin-protein-ligase E3A
UTR	Untranslated region
UV	Ultraviolet
VNTR	Variable Nucleotide Tandem Repeat
WNT	Wingless-type MMTV Integration

Table of Contents

	Page
Declaration	i
Summary	ii
Acknowledgements	iv
Statement of work	v
Publications	vi
Abbreviations	vii
Index of Tables	xvi
Index of Figures	xviii
Chapter 1.....	1
Introduction.....	1
1.1 General Overview of Autism.....	2
1.1.1 Definition of Autism.....	2
1.1.2 Epidemiology.....	6
1.1.3 Associated Medical Disorders.....	7
1.1.3.1 Tuberous Sclerosis.....	8
1.1.3.2 Phenylketonuria.....	8
1.1.3.3 Fragile X.....	9
1.1.3.4 Neurofibromatosis.....	10
1.1.4 Neuroanatomical Observations.....	10
1.2 Genetic Studies and Autism.....	11
1.2.1 Family Studies and Autism.....	11
1.2.2 Twin Studies.....	13
1.2.3 Mode of Inheritance.....	14
1.2.4 Environmental Factors.....	15
1.3 The Search for Autism Candidate Genes.....	17
1.3.1 Linkage Studies.....	17
1.3.2 Chromosomal Abnormalities.....	20
1.3.3 Animal Models.....	25
1.3.4 Clinical Observations.....	27
1.4 Testing Potential Autism Susceptibility Genes.....	29
1.4.1 Linkage Disequilibrium Mapping, Association Testing and Haplotype Analysis.....	29
1.4.1.1 The Theory.....	29
1.4.1.2 Potential Candidate Genes Highlighted by Association Studies ...	33

1.4.2	Mutation Screening.....	38
1.4.2.1	The Theory	38
1.4.2.2	Potential Candidate Genes Highlighted by Mutation Screens.....	39
1.4.3	Expression Profiling.....	43
1.5	Aims of thesis	45
chapter two		46
Materials and Methods		46
2.1	Sample Description.....	47
2.2	Biological Techniques	48
2.2.1	DNA Extraction.....	48
2.2.1.1	DNA Extraction from Blood	48
2.2.1.2	DNA Extraction of DNA from Cheek Cells.....	49
2.2.1.3	DNA Quantification	50
2.2.2	Polymerase Chain Reaction (PCR).....	50
2.2.3	Gel Electrophoresis and Visualisation	51
2.2.3.1	Agarose Gels.....	51
2.2.3.2	Non Denaturing Polyacrylamide Gels.....	51
2.2.3.3	Denaturing Polyacrylamide Gels for Use with the ABI PRISM 377 DNA Sequencer	51
2.2.4	Genotyping	52
2.2.4.1	Insertion / Deletions (InDel) and Variable Tandem Nucleotide Repeats (VNTRs)	52
2.2.4.2	Restriction Fragment Length Polymorphism (RFLP) Analysis	52
2.2.4.3	SNaPshot	53
2.2.4.4.1	Genotyping with Applied Biosystems 3100 Genetic Analyzer (ABI 3100).....	54
2.2.4.5	TaqMan Allelic Discrimination	54
2.2.5	Mutation Detection	55
2.2.5.1	Single Strand Conformational Polymorphism (SSCP) Analysis...	55
2.2.5.2	Denaturing High Performance Liquid Chromatography (dHPLC) Analysis.....	56
2.2.5.3	DNA Sequencing	57
2.2.5.3.1	Post PCR Cleanup.....	57
2.2.5.3.2	Sequencing Reaction.....	57
2.2.5.3.3	Post Sequencing Reaction Cleanup	58
2.2.5.3.4	Sequencing Analysis.....	58
2.2.6	Measurement of DBH Activity	58
2.3	Statistical Tests	59
2.3.1.1	Hardy Weinberg Equilibrium	59
2.3.1.2	Transmission Disequilibrium Testing	60
2.3.1.3	Transmission of Haplotypes	61
2.3.1.3.1	TRANSMIT	61
2.3.1.3.2	UNPHASED	61
2.3.2	Other Statistics.....	62
2.3.2.1	Odds Ratio (OR)	62

2.3.2.2	Graphical overview of linkage disequilibrium (GOLD).....	62
2.3.2.3	Haplotype Block Structure Analysis using Haploview.....	64
2.3.2.4	Heritability.....	64
2.3.2.5	T-Tests.....	65
2.3.2.6	Kruskal-Wallis Testing.....	66
2.3.2.7	Regression Analysis.....	66
2.3.2.8	Selection of Tagging SNPs.....	67
Chapter 3		69
The Serotonergic System and Autism		69
3.1 Introduction		70
3.1.1	The Serotonin System.....	70
3.1.2	Serotonin and Autism	71
3.1.3	Genetic Studies of the Serotonin Related Genes And Autism	75
3.1.4	The Serotonin Transporter Gene	76
3.1.5	The Tryptophan Hydroxylase 2 (TPH2) Gene	78
3.1.6	Aims.....	81
3.2 Materials and Methods		82
3.2.1	Samples	82
3.2.2	Amplification of Serotonin Transporter Variants.....	82
3.2.3	Amplification of TPH2 Variants	82
3.2.4	Statistics	83
3.3 Results		85
3.3.1	The Serotonin Transporter Gene	85
3.3.1.1	LD Measures.....	85
3.3.1.2	TDT Testing.....	86
3.3.1.3	Transmission of Haplotypes	87
3.3.1.4	Power Study.....	89
3.3.1.5	Joint-analysis of published serotonin transporter promoter variants.....	90
3.3.2	TPH2.....	94
3.3.2.1	LD Measures.....	94
3.3.2.2	TDT Testing.....	96
3.3.2.3	Transmission of Haplotypes	96
3.3.2.4	Power Study.....	97
3.4 Discussion		98
3.4.1	Replication of Previous Results?.....	98
3.4.1.1	Serotonin Transporter.....	98
3.4.1.2	Replication of TPH2 Results?	100
3.4.2	Potential Reasons for Differing Results.....	100
3.4.2.1	The Serotonin Transporter Gene.....	100
3.4.2.2	The TPH2 Gene	103
3.4.3	Conclusion.....	103
Chapter 4		105
The Dopamine-β-Hydroxylase Gene and Autism		105

4.1	Introduction	106
4.1.1	Dopamine and Autism	106
4.1.2	The Dopamine Beta Hydroxylase Gene.....	109
4.1.3	Dopamine- β -hydroxylase & Epilepsy	111
4.1.4	Dopamine- β -hydroxylase Animal Models.....	112
4.1.5	The Aims	113
4.2	Materials & Methods.....	114
4.2.1	Measuring DBH Activity	114
4.2.2	Amplification of DBH Variants	114
4.2.3	Statistics	115
4.3	Results	116
4.3.1	Age effect and DBH Activity.....	116
4.3.2	Heritability of DBH Activity.....	119
4.3.3	DBH Activity. Values and Comparisons.	120
4.3.4	Comparison of Irish Proband DBH Activity with Previously Reported Children Control Sample.	123
4.3.5	Comparison of DBH Activity in Individuals with Epilepsy	125
4.3.6	Investigation of the Association of Allelic Variants with DBH Activity.	126
4.3.7	Regression Analysis of DBH Variants in Parents	128
4.3.8	Regression Analysis of DBH Variants in Children.....	129
4.3.9	Linkage Disequilibrium Measurements.....	130
4.3.10	TDT Testing	133
4.3.11	Transmission of Haplotypes.....	134
4.4	Discussion.....	136
4.4.1	DBH Activity with Age	136
4.4.2	Heritability of DBH Activity.....	137
4.4.3	Differences in DBH Activity.....	137
4.4.4	DBH and Epilepsy	139
4.4.4	Transmission of Allelic Variants.....	139
4.4.5	Regression Analysis in Parents and Children.....	140
4.4.6	Possible Reasons for Biochemical/Genetic Discrepancy.....	141
4.4.7	Conclusions	143
Chapter 5.....	144
Integrin-α4 and Autism	144
5.1	Introduction	145
5.1.1	Chromosome 2q and Genome Scans	145
4.1.2	Cytogenetic Study in Irish Patient Suffering Autism	146
5.1.3	Molecular Studies Involving Chromosome 2q32.....	149
5.1.4	The Integrin Alpha 4 (ITGA4) Gene	152
5.1.5	The HapMap Project.....	156
5.1.6	Aims.....	157
5.2	Materials And Methods.....	158
5.2.1	Samples	158
5.2.2	Selection and Genotyping of SNPs.....	159

5.2.3	Mutation Detection	159
5.2.4	Statistics	160
5.3	Results	161
5.3.1	Selection of SNPs	161
5.3.2	LD structure in ITGA4 Gene.....	164
5.3.3	TDT analysis	167
5.3.4	Haplotype Analysis.....	169
5.3.5	Mutation Screening of Exons 16 and 17.....	171
5.3.6	Attempted Replication	177
5.3.6.1	LD Analysis of Replication Samples	177
5.3.6.2	TDT Replication Results.....	179
5.3.6.3	Haplotype Transmissions in Replication Sample	183
5.3.6.4	Comparison of 4-marker haplotypes across all 3 samples	187
5.3.6.5	Joint Analysis.....	190
5.3.4	Testing Validity of htSNPs	194
5.4	Discussion.....	198
5.4.1	LD Structure of ITGA4 Gene.....	198
5.4.2	TDT, Haplotypes and Mutation Detection.....	198
5.4.3	Attempted Replication	200
5.4.3.1	AGRE Sample.....	200
5.4.3.2	Vanderbilt Sample.....	200
5.4.3.3	Joint Analysis.....	202
5.4.4	Possible Theories for Differences in Haplotype / Allele Transmission	203
5.4.5	Biological Hypothesis.....	204
5.4.6	The Use of HapMap Tagging SNPs	208
5.4.7	Conclusions	210
Chapter 6.....	212
The Wnt Pathway.....	212
6.1	Introduction	213
6.1.1	An Overview of the Wnt Pathway.....	213
6.1.2	Expression in the Brain, Animal Models and Autism	218
6.1.3	The Wnt2 Gene.....	222
6.1.4	The Frzb Gene	223
6.1.5	The Dishevelled Gene.....	225
6.1.6	The EN2 Gene	226
6.1.7	Aims.....	227
6.2	Material and Methods	228
6.2.1	Mutation Screen of Wnt2 Gene	228
6.2.2	Genotyping of SNPs	228
6.2.3	Statistics	229
6.3	Results	230
6.3.1	WNT2 Gene	230
6.3.1.1	LD Structure	230
6.3.1.2	TDT Analysis.....	231

6.3.1.3	Haplotype Analysis	232
6.3.2	Frzb Gene	233
6.3.2.1	LD Measurements	233
6.3.2.2	TDT Analysis.....	234
6.3.2.3	Transmission of Frzb Haplotypes	235
6.3.3	Dishevelled Gene.....	237
6.3.3.1	LD Structure	237
6.3.3.2	TDT Testing.....	238
6.3.3.3	Transmission of Haplotypes	239
6.3.4	EN2 Gene	240
6.3.4.1	LD Structure	240
6.3.4.2	TDT Analysis.....	241
6.3.4.3	Haplotype Testing	241
6.4	Discussion.....	243
6.4.1	WNT2 Gene	243
6.4.2	Frzb Gene	244
6.4.2	Dvl Gene	244
6.4.3	EN2 Gene & Autism.....	245
Chapter 7	247
General Discussion	247
7.1	Aims and Findings	248
7.1.1	Findings.....	248
7.2	Strengths and Limitations	251
7.2.1	Strengths of this Study	251
7.2.1.1	Irish Sample	251
7.2.1.2	Design and Analysis.....	253
7.2.1.3	Gene Coverage.....	254
7.2.2	Limitations of Study	257
7.2.2.1	Multiple Testing.....	257
7.3	General Conclusions.....	258
7.4	Future Directions.....	260
Appendix I	294
Appendix II	299
Appendix III	312
REFERENCES	178

Index of Tables

		Page
Table 1. 1	ICD-10 diagnostic criteria.....	4
Table 1. 2	DSM-IV diagnostic criteria for autism.....	5
Table 1. 3A	A list of chromosomal abnormalities reported in individuals with autism.....	23
Table 1. 3B	A list of chromosomal abnormalities reported in individuals with autism.....	24
Table 1.4A	Sample of genes showing positive associations with autism.....	36
Table 1.4B	Sample of genes showing positive associations with autism.....	37
Table 1.5	Sample of genes that have revealed potential mutations associated with autism.....	42
Table 3.1	TDT Results for Serotonin Transporter Markers.....	85
Table 3.2	Haplotype Transmission and Odds Ratio Data for the Serotonin Transporter Markers.....	87
Table 3.3	Individual Study Results for Transmission of Serotonin Transporter Variants.....	91
Table 3.4	TDT Results for TPH2 Markers.....	95
Table 3.5	Haplotype Transmission Results for the TPH2 Gene.....	117
Table 4.1	DBH Activity Paired Difference Summary.....	121
Table 4.2	Paired Samples t-Test for Differences in DBH Activity.....	122

Table 4.3	Two-Sample t-Test for Differences in DBH Activity in Proband Sample and Child Control Sample.....	124
Table 4.4	Comparison of DBH Activity Between Individuals with Epilepsy and Autism, Autism And Child Control Samples.....	125
Table 4.5	Kruskal-Wallis test results for differing allelic variants and the associated DBH activity.....	127
Table 4.6	Regression analysis of effect of variants on overall DBH activity in parental samples.....	128
Table 4.7	Regression analysis of effect of variants on overall DBH activity in the proband sample.....	129
Table 4.8	TDT Results for DBH Variants.....	132
Table 4.9	Haplotype transmission results for the DBH variants.....	134
Table 5.1	Selection of SNPs within and surrounding the ITGA4 gene.....	161
Table 5.2	TDT results for individual ITGA4 SNPs.....	167
Table 5.3	Transmission of ITGA4 two-marker haplotypes.....	169
Table 5.4	PDT Results For AGRE Sample.....	180
Table 5.5	PDT Results For Vanderbilt Sample.....	181
Table 5.6	Transmission of ITGA4 two marker haplotypes in AGRE sample.....	183
Table 5.7	Transmission of ITGA4 two marker haplotypes in Vanderbilt sample.....	185
Table 5.8A	Comparison of 4 Marker ITGA4 Haplotypes Transmitted to Affected Samples.....	187
Table 5.8B	Comparison of 4 Marker ITGA4 Haplotypes Transmitted to Affected Samples.....	188
Table 5.9	Joint analysis of ITGA4 in the amalgamated sample.....	191
Table 5.10	Joint Analysis for Two-Marker Haplotypes in	

	The Amalgamated Sample.....	192
Table 5.11	Role of integrin proteins in brain.....	206
Table 6.1	Wnt signalling components as Wnt pathway targets.....	216
Table 6.2	Results of TDT testing of WNT2 allelic variants.....	230
Table 6.3	Results of haplotype testing of WNT2 variants.....	231
Table 6.4	TDT testing of Frzb Variants.....	234
Table 6.5	Transmission of Frzb Haplotypes.....	235
Table 6.6	Results of TDT testing of Dvl allelic variants.....	238
Table 6.7	Results of haplotype testing in Dvl gene.....	238
Table 6.8	Results of TDT testing in En2 gene.....	240
Table 6.9	Transmission of haplotypes in EN2 gene.....	241

Index of Figures

Figure 3. 1	Summary of the Serotonin Pathway and Links to autism.....	76
Figure 3. 2	Measures of LD (D') across the Serotonin Transporter Gene.....	84
Figure 3. 3	The Odds Ratio of each Serotonin Transporter Haplotype.....	88
Figure 3. 4	A Funnel Plot.....	92
Figure 3.5	Measures of LD (D') across the TPH2 Gene.....	94

Figure 4.1	The Dopamine / Norepinephrine Pathway.....	107
Figure 4.2	Dotplot showing the relationship between transformed DBH activity and age in the dataset including all probands.....	116
Figure 4.3	Dotplot showing the relationship between transformed DBH activity and age in individuals greater than 72 months of age.....	117
Figure 4.4	Measurement of the heritability of DBH activity.....	118
Figure 4.5	Spouse-spouse correlation of DBH activity.....	119
Figure 4.6	Boxplot of DBH activity in Mothers, Fathers and Children.....	120
Figure 4.7	Boxplot of DBH Activity in Proband Sample and Child Control Sample.....	123
Figure 4.8	Box plot of C1021T Allelic Variants and their associated DBH activity.....	126
Figure 4.9	LD measures and Haploblock Formation of DBH variants.....	131
Figure 5.1	Cytogenetic spread showing the presence of translocated chromosome 2q32.1-32.2 on Chromosome 9q31.1-31.3.....	146
Figure 5.2	Fluorescent in situ hybridization (FISH) showing the presence of a chromosome 2q probe (341A4) translocated to chromosome 9q.....	147
Figure 5.3	Location of initial Chromosome 2q screening markers and exons.....	150
Figure 5.4	The role of integrin proteins in linking the proteins in the extracellular matrix to actin fibres in the cell interior.....	154
Figure 5.5	Location of Exons and Markers in the ITGA4 Gene.....	162
Figure 5.6	LD Structure of ITGA4 in Parental Sample.....	164
Figure 5.7	LD Structure of ITGA4 in Control Sample.....	165
Figure 5.8	LD Structure of ITGA4 in HapMap Sample.....	165
Figure 5.9	dHPLC analysis of ITGA4 Exon 16 at 51°C.....	171

Figure 5.10	dHPLC analysis of ITGA4 Exon 16 at 53°C.....	172
Figure 5.11	dHPLC analysis of ITGA4 exon 16 at 56°C.....	173
Figure 5.12	dHPLC analysis of ITGA4 exon 17 at 52°C.....	174
Figure 5.13	Sequencing results for mutation detected by dHPLC.....	175
Figure 5.14	ITGA4 LD Structure in the AGRE Sample.....	177
Figure 5.15	ITGA4 LD Structure in the Vanderbilt Sample.....	178
Figure 5.16	Haplotype Tagging SNPs and Inter-Block Relationships.....	194
Figure 5.17	Haplotype Tagging SNPs and Inter-Block Relationships in AGRE sample.....	196
Figure 5.18	Haplotype Tagging SNPs and Inter-Block Relationships in Vanderbilt Sample.....	196
Figure 6.1	Canonical Wnt Signalling Pathway.....	214
Figure 6.2	Wnt canonical and non-canonical pathway.....	215
Figure 6.3	LD measurements within the WNT2 gene.....	229
Figure 6.4	Measures of LD and presence of haploblocks within the Frzb Gene.....	233
Figure 6.5	LD measurements within Dvl Gene.....	237
Figure 6.6	LD Measurements in EN2 gene.....	239

CHAPTER 1

INTRODUCTION

1.1 General Overview of Autism

1.1.1 Definition of Autism

Autism was first described by Leo Kanner in 1943 ¹⁰. He described the typical characteristics of 11 individuals with autism as being “autistic aloneness”, “obsessiveness”, “stereotypies” and “echolalia”. In the following year, Hans Asperger wrote a monograph in which he described individuals with autistic-like symptoms similar to those described by Kanner ¹¹. In his paper he described 400 individuals with deficiencies in speech, non-verbal communication and social interactions. He also noted the presence of repetitive behaviours, resistance to change, abnormalities of motor co-ordination, in addition to the presence of savant skills. The individuals described by Asperger were largely of normal intellectual functioning. Asperger’s Syndrome is used today to describe individuals with autistic like deficits associated with normal intelligence. It is included in the International Classification of Diseases, 10th Edition (ICD-10) ¹² and the Diagnostic and Statistical Manual, 4th Edition, (DSM-IV) ¹³ and as a separate diagnostic category from autism (see section 1.1.6).

Autism belongs to the category Pervasive Developmental Disorder (PDD) in ICD-10 and DSM-IV that includes classic autism and autism spectrum disorders (including Asperger disorder), childhood disintegrative disorder (involving behavioural, cognitive and language regression between the ages of 2 and 10 years of age following normal development), and Rett syndrome (a genetic disorder of postnatal brain development, caused by a single gene defect in the methyl-CpG binding protein, predominantly affecting girls).The definition of autism according to both ICD-10 and DSM-IV depends upon the presence of the following:

- (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.

The criteria for autism according to ICD-10 and DSM-IV are shown in Tables 1.1 and 1.2 respectively.

There is general agreement between ICD-10 and DSM-IV concerning the nature of the clinical abnormalities occurring in autism. Both agree that abnormalities should occur across the three domains of reciprocal social interaction, communication and behaviour. Minor differences do occur, e.g. in the social domain, ICD-10 includes impairments in the offering or seeking of comfort when distressed whereas this is not included in DSM-IV.

Classical autism is one of a number of subtypes in the spectrum based on number and description of impairments. The 5 DSM-IV Pervasive Developmental Disorder (PDD) subtypes are 1) classic autism, 2) Asperger disorder (involving language development at expected age and no mental retardation), 3) disintegrative disorder (involving behavioural, cognitive and language regression between the ages of 2 and 10 years of age following normal development), 4) PDD-not otherwise specified (individuals have autistic features but do not fit into the other categories) and 5) Rett disorder (a genetic disorder of postnatal brain development, caused by a single gene defect in the methyl-CpG binding protein, predominantly affecting girls).

Careful consideration must be given to the differential diagnosis for autism in the context of a genetic study in autism in order to avoid phenotypic heterogeneity. Both ICD-10 and DSM-IV include broader categories of autism spectrum disorders such as 'Atypical Autism', (ICD-10), Pervasive Developmental Disorder - not otherwise

specified (PDD-NOS) (DSM-IV) and Asperger Disorder. These diagnostic categories include some but not all of the criteria as defined for autism and may be thought to refer to individuals with milder subtypes. For the purposes of the investigations described in this work only individuals who meet the strict criteria for autism were included. In order to limit the extent of diagnostic heterogeneity the sample recruited were assessed using research diagnostic tools, the Autism Diagnostic Interview (Revised) (ADI-R) ¹⁴ and the Autism Diagnostic Observation Schedule- Generic (ADOS-G) ¹⁵. The inclusion and exclusion criteria for the sample are described in Chapter 2, section 2.1 (Materials and Methods).

<p>DIAGNOSTIC CRITERIA FOR AUTISM DISORDER (ICD-10) (WHO 1992) At least 8 of the 16 specified items must be fulfilled.</p> <p>A. Qualitative impairments in reciprocal social interaction, as manifested by at least three of the following five:</p> <ol style="list-style-type: none"> 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. <p>B. Qualitative impairments in communication:</p> <ol style="list-style-type: none"> 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. <p>C. Restricted, repetitive and stereotyped patterns of behaviour, interests and activities, as manifested by at least two of the following six:</p> <ol style="list-style-type: none"> 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. <p>D. Developmental abnormalities must have been present in the first three years for the diagnosis to be made.</p>
--

Table 1.1 ICD-10 DIAGNOSTIC CRITERIA

**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 preoccupation 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.

Table 1.2 DSM-IV DIAGNOSTIC CRITERIA FOR AUTISM.

1.1.2 Epidemiology

The prevalence of autism in the general population has been estimated to be 10/10,000¹⁶. This prevalence rate is the median rate of 32 autism prevalence studies (See Review by Eric Fombonne for more information¹⁶). These studies have spanned numerous countries (including Denmark, England, Japan, Sweden, France and the USA) and time periods (mid 1960's to present day) and therefore can be seen as an overall global average in the developed world^{17 18}. The first epidemiological survey for autism was undertaken in 1966. A survey of 78,000 children from the Middlesex region of the UK revealed that there were 32 diagnosed with autism, leading to a prevalence of 4.1 / 10,000¹⁷. In individual studies, the prevalence rate of autism varies widely from 0.7 per 10,000 in an American population¹⁹ to 72.6 per 10,000 in a small Swedish population²⁰. There are many reasons for the differences in prevalence estimate. They include sample size (prevalence is negatively correlated with sample size)¹⁶, year of study (prevalence shows increases with advancing year of publication) and method of diagnosis¹⁶.

There is much debate within the published literature with respect to the prevalence of autism. It is not yet clear whether the observed increases reflect a true rise in prevalence or if there has been better detection in recent years as a result of heightened awareness. Within the media there have been a number of reports suggesting an increasing prevalence of autism in recent years. The US Department of Developmental Services reported a 556% increase in prevalence from 1991 to 1997²¹. Here it is important to note the difference between prevalence (the number of cases of an illness or condition that exists at a particular time in a defined population) and incidence (the number of new cases of an illness diagnosed in a defined population over a specified period of time). In many cases, increased awareness of autism in the general population and in the medical profession may lead to an increase in numbers of individuals diagnosed with autism. Also in the past individuals with known genetic causes, e.g. Fragile X, or severe mental retardation, would have been excluded from prevalence estimates. Eric Fombonne argues

against an increase in the prevalence of autism stating “the available epidemiological evidence does not strongly support the hypothesis that the incidence of autism has increased.... Due to the inherent difficulties in studying a relatively rare disorder, most studies are inadequate to test the hypothesis of increasing prevalence”¹⁶. A study conducted in the state of California in 2002 based on information obtained from autism and intellectual disability registers found that rising autism diagnoses appeared to reflect a corresponding decrease in the rates of intellectual disability²².

In all epidemiological studies of autism to date, there has been an increased incidence of affected males compared to affected females. The male to female ratio varies from 1.33 to 16^{23 24}. The mean male/female ratio has been calculated to be 4.3¹⁶. This ratio correlates negatively with the level of mental retardation, i.e. the ratio of affected males to females decreases as the level of mental retardation increases. In a comparison of 12 studies, the ratio was 5.75 with individuals who were in the normal band of intellectual functioning. However, when the ratio was compared in individuals with moderate to severe mental retardation, it decreased to 1.9¹⁶.

1.1.3 Associated Medical Disorders

A number of medical conditions are associated with higher prevalence rates of autism than that observed in the general population. This observation led to the suggestion that autism had a biological basis. Several of these conditions are inherited; lending support to the hypothesis that autism has a genetic component to the aetiology. Most of these are rare disorders associated with intellectual disability. A few of the more common disorders are discussed briefly below, as these were disorders that were excluded from the sample described in this work.

1.1.3.1 Tuberos Sclerosis

The prevalence of autism/atypical autism and/or PDDs in individuals diagnosed with tuberous sclerosis has been estimated to be between 20 to 61% ²⁵. Tuberous sclerosis is a dominantly inherited disorder resulting from mutations in the hamartin gene (*TSC1*) located on chromosome 9q34 and the tuberin gene (*TSC2*) on chromosome 16p13 ^{26,27 28}. There is no homology between these two genes, but they are found to act in the same complex. Disruption / alterations to this complex leads to the presence of hamartomata (benign tumour like malformations resulting from faulty development in an organ) in multiple organs including the brain, which in turn result in an increased occurrence of epilepsy, learning difficulties and behavioural problems. The incidence of tuberous sclerosis in individuals with autism ranges from 0.4% to 2.99% ²⁵. The association between the two disorders has two important implications. Firstly, regions of the brain affected in individuals with tuberous sclerosis are also likely to be affected in individuals diagnosed with autism. Secondly, TS is a condition with a simple genetic architecture and it is possible that autism may be associated with abnormalities in genes involved in tuberous sclerosis. There has only been one report of an association with the *TSC2* gene and autism ²⁹. This result indicates that while *TSC1* and *TSC2* genes are important in the development of tuberous sclerosis, they are not responsible for the majority of cases in autism.

1.1.3.2 Phenylketonuria

Phenylketonuria (PKU) is a disease resulting from an inborn error in metabolism. Individuals with PKU are deficient in phenylalanine hydroxylase, but the development of autistic symptoms can be prevented by dietary means (OMIM reference: 261600). Widespread testing of neonates has significantly reduced the

problems associated with PKU. If left untreated, PKU leads to the accumulation of phenylpyruvic acid, which inhibits pyruvate decarboxylase in the brain ³⁰. The association with autistic symptoms in untreated PKU would suggest that some of these adversely affected regions might be involved in autism.

1.1.3.3 Fragile X

Fragile X is another Mendelian condition associated with autism. Fragile X is caused by the expansion of a CGG triplet repeat on chromosome X. Chromosome X has two fragile sites; fragile X site A (FRAXA) and fragile X site E (FRAXE). FRAXA has been mapped to Xq27.3, and occurs in the 5' untranslated region (UTR) of the fragile X mental retardation gene 1 (*FMR-1*). The number of repeats varies from individual to individuals. The normal repeat range, with stability of the expansion, is between 6 and 54 triplet repeats. A sequence of more than 200 repeats is unstable, and transcription of the gene is disrupted. A second fragile X site, FRAXE, has also been mapped to the *FMR-1* gene and occurs in the promoter region of the gene. In a fashion similar to the FRAXA site, there is a CGG repeat. Six to 25 repeats are considered stable, with greater than 200 being unstable ³¹.

The clinical observation of autism occurring in association with the Fragile X anomaly has been well established ³². 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 ³³. 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 ^{34 35}

1.1.3.4 Neurofibromatosis

Neurofibromatosis (NF) is an autosomal dominant disorder characterised by café-au-lait spots and fibromatous tumours of the skin. There are two forms of the disorder; Neurofibromatosis type I and Neurofibromatosis type II. Type I form of the disorder is caused by mutations in the neurofibrin-1 gene on chromosome 17q11.2, whereas type II form of the disorder is caused by mutations in the gene encoding neurofibrin-2 (also known as Merlin) on chromosome 22q12.2. Neuroimaging studies in patients diagnosed with NF type I revealed high signal intensity lesions in the basal ganglia, thalamus, brainstem, cerebellum and subcortical white matter ³⁶. It is hypothesised that these lesions may be a consequence of vacuolar change. Individuals who are diagnosed with NF type II characteristically present with tumours of the eighth cranial nerve, meningiomas of the brain and schwannomas of the dorsal root of the spinal cord. Intellectual disability occurs in 10-20% of cases with NF types I and II. Impulsivity and social difficulties are well-documented behavioural manifestations ³⁷. Several studies have reported an increased incidence of autism in association with NF1 ^{38 39}. The condition may go undiagnosed in childhood and should be excluded in the assessment of an individual with autistic symptoms.

1.1.4 Neuroanatomical Observations

One of Kanner's original observations was the occurrence of increased head size in 5 of his original 11 patients ¹⁰. Magnetic resonance imaging (MRI) studies have consistently found elevated brain volumes in individuals with autism (see Nicolson et al for a review on neurodevelopment and autism ⁴⁰). This finding gains increased support from post-mortem studies reporting increased cerebellum size in individuals with autism ⁴¹. Surprisingly, this is accompanied by decreases in the number of Purkinje cells in the cerebellum ⁴¹. Other neuropathological changes involve 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 ⁴².

Structural neuroimaging studies have shown increased average brain volume; decreased grey matter volumes in the limbic system ⁴³; reduced neuron numbers in the vermis of the cerebellum and gross structural changes in cerebellum ⁴⁴ and the parietal lobes ⁴⁵. Poorer neuronal integrity in prefrontal areas has been suggested ⁴⁶, as have concurrent abnormalities in the frontal cortex and cerebellum ⁴⁷. A longitudinal functional magnetic resonance imaging (fMRI) study suggests an accelerated period of growth during childhood compared with typically developing controls ⁴⁸. As medical imaging techniques improve, a better picture of the abnormalities involved in autism will occur.

1.2 Genetic Studies and Autism

1.2.1 Family Studies and Autism

Evidence for the role of genetics in the development of autism came initially from family studies. A large number of family studies have been undertaken in order to ascertain the rate of autism in first-degree relatives of affected individuals. The results of these family studies have all shown that the rate of autism is much higher in siblings of individuals who are diagnosed with autism, than in controls. The recurrence rates of autism however, varies between studies, ranging from 2% to 8.6% ^{49 50 51,52}. In a review of the literature, Smalley reported that the average recurrence risk, i.e. the likelihood that a given condition diagnosed in one or more

family members will recur in other family members of subsequent generations, was 3%⁵³. Bolton also reported a recurrence risk of 2.9%⁵⁴ for autism, however for traits associated with the broader phenotypes (communication/social impairments or stereotypic behaviours) there was a recurrence risk of between 12.4% and 20.4%. The increase in the presence of broad phenotype traits has also been reported by Pickles et al⁵⁵. This contrasted with the sharp decline in risk in 2nd degree relatives.

Although a 3% risk does not appear to be very high, this is almost 100 times that of the general population (if the strict definition of autism is taken into consideration, where the population rate is 3/10,000). If one considers the broad definition of autism, the recurrence rate is still 30 times that of an individual from the general population. These recurrence risks may actually be underestimated, due to the phenomenon of stoppage, whereby the first affected child tends to dissuade parents from having further children. It is also not possible to look at the transmission of the disorder from an affected individual to their child, as individuals with autism tend not to marry and have children. Autism risk alleles are not selected against in terms of evolution, as these are most likely to be common variants, which are carried by many individuals. The majority of individuals will not have a sufficient number of these risk variants, and therefore will not develop autism.

The heritability of autism (a statistical description of the proportion of phenotypic variance due to genetic variance) can be calculated from the recurrence risk and the monozygotic: dizygotic (MZ: DZ) concordance ratios (see Section 1.2.2) and is estimated to be greater than 90%²⁵. Autism has been shown to have the highest heritability of any neuropsychiatric disorder, including Attention Deficit Hyperactivity Disorder (heritability = 76%)⁵⁶, schizophrenia (heritability = 81%)⁵⁷ and bipolar affective disorder (heritability = 85%)⁵⁸. It is important to state that these figures do not mean that the environment has little influence on autism; it merely means that there was very little environmental variance in the populations studied.

1.2.2 Twin Studies

Twins are a naturally occurring genetic experiment that can assist in the investigation of the role of genes in a genetic disorder or trait. Twin studies work on the basis that those disorders or traits that are influenced by genes will show greater concordance in monozygotic (MZ) versus dizygotic (DZ) twins. In the case of autism, there have been four twin studies published⁵⁹⁻⁶². The first of these studies was published in the late 70's, and reported a concordance rate of 36% between MZ twins, compared to 0% in DZ twins⁵⁹. Although this was the first report of increased concordance in MZ twins compared to DZ twins, the power of the study was limited due to the small sample size (11 MZ pairs, 10 DZ pairs). Another study by Ritvo et al, also found an increased MZ: DZ concordance rate⁶⁰. However, in this study the MZ concordance rate was higher, being 96% compared to just 30% in DZ twins. It is important to note that there may be an inherent bias in this study, as Ritvo et al recruited his twins using an advertisement in a newsletter. This may have led to more affected twins responding than pairs of twins in which just one individual was affected. As a result, the heritability reported may be an overestimate. Furthermore Rivotto et al included DZ twins of different sexes, which is unlike all other studies that just include same sex twin pairs. Consequently, comparison of results is more difficult. Its similarity to other studies remains, in that a higher MZ concordance rate exists. A Scandinavian study found the MZ concordance rate to be significantly higher than the DZ concordance rate (>60% (MZ) vs. 0% (DZ))⁶¹. Similarly the final study included the initial 21 twins from the Folstein & Rutter study, with an additional 14 MZ twins and 10 DZ twins. Bailey et al reported a MZ concordance rate of 60% compared to 0% in DZ twins⁶². Despite varying concordance rates, all studies are consistent in that concordance rates are significantly higher in MZ twins compared to DZ twins. In all studies, with the exception of Ritvo's (which may be biased), the DZ concordance rate was 0%. This is less than the sibling recurrence risk of 3% (see section 1.2.1 for more info). This may have occurred due to the small sample sizes involved in each study and be a chance finding.

A number of studies looked at concordance of trait measures within the twins. Steffenberg et al included communication skill defects and social disorders, and found the MZ concordance rate rose to 92% in MZ and rose from 0% to 10% in DZ twins ⁶¹. Likewise Folstein and Rutter reported that if the criteria for concordance included cognitive impairment, the concordance rates increased to 82% (MZ) and 30% (DZ) ⁵⁹.

Increase of concordance in MZ twins is evidence for genetic influence, but the fact that concordance is not 100% shows that there is also an environmental effect as well, which may include being a twin itself (see section 1.2.4 for further information) ⁶³.

1.2.3 Mode of Inheritance

It is widely accepted that there is strong evidence indicating a genetic cause of autism (discussed in sections 1.2.1 and 1.2.2). A mode of inheritance is not yet clear. From the observed sibling recurrence risk (3%) it is obvious that autism is not a single gene disorder as one would expect a sibling recurrence risk of 50% in the case of a dominant disorder or 25% in the case of a simple recessive disorder. All evidence points to a multi-gene effect. A study by Jorde et al advocates a genetic mixed model, which specifies “continuous liability to disease as the sum of independent effects attributed to segregation of alleles at a major locus, transmission by polygenic inheritance, sharing by siblings of environmental factors and random factors specific to the individual” ⁶⁴. Autism occurs when an individual’s liability exceeds the threshold. There was no support for a single gene effect in this study nor in a study undertaken by Pickles et al ⁶⁵. In his analysis, the best fitting model was for three epistatic loci (limits of 2 to 10).

Another study, by Risch et al, also found evidence for multiple genes leading to the development of autism⁶⁶. Risch et al examined allele sharing in a sample of affected sib pairs (ASPs) and discordant sib pairs (DSPs). There were only slight increases in the level of allele sharing between these two groups (allele sharing = 51.6% in ASPs and 50.8% in DSPs). This did not support a small number of genes, e.g. 3 genes would have required >70% allele sharing in ASPs, but was most consistent with a multigenic inheritance pattern with equal to or greater than 15 susceptibility loci.

Although there remains to be agreement on the number of genes involved in the development of autism, it is clear that at least it is an oligogenic (i.e. 3 – 10 genes) disorder, if not polygenic. Epigenetic mechanisms may also prove to play a role in autism. They are mechanisms that influence gene expression but 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 genetic association between the general autism phenotype and markers in the region (see Sections 1.3.2 and 1.4.1.2). The role of epigenetic factors, if any, in relation to autism has yet to be investigated.

1.2.4 Environmental Factors

The observation that the MZ concordance rate is not 100% indicates that environmental factors are also likely to play a role in the development of the autistic aetiology, although the high MZ concordance rates suggest that this might be a limited role (see section 1.2.2). The most widely referenced environmental risk has been the Measles-Mumps-Rubella (MMR) vaccination. The publication of Prof. Andrew Wakefield's paper in 1998, which reported the onset of autism following the

MMR vaccination in 8 out of 12 individuals, led to widespread public concern ⁶⁷. Subsequent findings, including a population based retrospective study involving more than 1.5 million children in Denmark found no association between MMR and autism ⁶⁸. Likewise a study in the UK by Taylor et al, reported no significant increase in autism following the introduction of the MMR vaccine ⁶⁹

A more recently reported risk factor for autism is twinning. Greenberg, in 2001, observed that there were more twins in the AGRE dataset than would be expected in the general population ⁶³. In the general population the prevalence of MZ twins is 0.008 and DZ twins is 0.016, leading to an overall prevalence of 0.024 ⁶³. However in the AGRE population, the MZ prevalence was 0.142 and the DZ prevalence 0.075, leading to an overall prevalence of 0.226. These differences were deemed to be statistically significant (p value <0.000001). These findings were replicated in part by Betancur et al, who also found a significant increase in the numbers of MZ twins, but not DZ twins in their population ⁷⁰. However Hallmayer et al were unable to replicate this finding in their Australian population suffering autism ⁷¹. Two thirds of MZ twins, unlike DZ twins, share a chorionic membrane and hence placenta, therefore 1/3 of MZ twins and DZ twins may experience different prenatal influences ⁷². These monochoionic or monoamniotic factors may play a role in the development of autism.

Other agents, including thalidomide and valproic acid, have also been linked to autism ⁷³⁻⁷⁵ (see Section 1.3.3).

1.3 The Search for Autism Candidate Genes

There are approximately 30,000 to 40,000 genes in the human genome⁷⁶ and of these only a small number (see Section 1.2.3) are involved in the development of autism.

A number of methods and strategies have been developed to assist in identifying the most probable candidates. Testing all variants in all genes in samples of affected individuals would lead to the identification of susceptibility variants in candidate genes. At present, this remains impractical and uneconomical. Testing of all variants would ultimately lead to a high level of false positive results. Instead, strategies are required to whittle down the number of genes to limits that are realistic. These strategies include linkage analysis, cytogenetic studies, animal studies and medical observations. Finding genes responsible for autism may lead to the development of better therapies and medication to help ameliorate some of the phenotypic traits associated with the disorder.

1.3.1 Linkage Studies

Traditionally, when a disorder is known to be genetic, but nothing is known about the genes likely to be involved, then one method used to narrow down candidate regions is linkage analysis. This utilises a number of markers spread throughout the genome. It is not necessary to know if the markers (single nucleotide polymorphisms (SNPs), Variable Nucleotide Tandem Repeats (VNTRs) or microsatellites) are functional, just where they are located relative to each other. Microsatellites are the most frequently used markers in linkage analysis as the polymorphic information content (PIC) is higher in comparison to bi-allelic markers. Evidence for linkage arises from tracing the co-segregation of DNA variants with the trait under consideration along family lineages. Linkage analysis is based on departure from Mendel's Second Law, which states that during gamete formation the

segregation of alleles of one allelic pair is independent of segregation of the alleles of another allelic pair. This is only true when the allelic pairs are located on different chromosomes, or are located far apart on the same chromosome so that recombination will prevent alleles from being transmitted together. The further apart two markers are, the greater the chance of recombination.

Parametric linkage ideally uses large multi-generational families with affected and unaffected relatives. It also requires the *a priori* specification of the genetic parameters of a model for the disease in terms of allelic frequencies, penetrance (i.e. the probability that a functional variant results in phenotypic change), the mode of action (i.e. dominant or recessive) and the distance from the genetic marker under examination.

Non-parametric linkage methods, originally designed around small nuclear families, have been extended to include general pedigree structures. It is based on allele sharing in affected relative pairs (or allele non-sharing in discordant relative pairs). One advantage of non-parametric linkage methods is that it is not necessary to specify a model for the disease *a priori*, and generally, these methods are computationally simple. Also, since no model is specified, there is no chance of choosing the wrong model and as a result there is greater power. In addition, as there is no need to test several models, so multiple testing is reduced. However, they may lack power, compared to parametric methods under certain circumstances, i.e. where you have the correct model.

There have been a number of genome scans and linkage studies for autism. The first genome scan was undertaken by the IMGSAC consortium, which investigated linkage in 99 multiplex families with 354 markers⁷⁷. Evidence for linkage was found on 6 chromosomes (chromosomes 4, 7, 10, 16, 19 and 22), but they did not reach a genomewide significance threshold (maximum LOD scores (MLS) > 3.0 indicate highly significant linkage, whereas those with a score of 1.0 < MLS < 3.0 show weaker statistical evidence for linkage). The region around markers D7S520

and D7S684 (on 7q31) showed evidence of linkage in a subset of the sample (UK families only) (MLS = 3.55), pointing towards the potential problems of sample heterogeneity.

There have been an additional 13 genomewide linkage studies published ⁷⁷⁻⁸⁸. To date all chromosomes, with the exception of chromosomes, 14, 21, 22 and the Y chromosome, have shown at least nominally significant linkage to autism (MLS > 1). The most replicated chromosomal regions showing significant linkage to autism are chromosomes 2q (see Section 5.1), 7q, 16p and 17q.

Evidence of linkage to markers on chromosome 7 has been reported by IMGSAC ⁷⁷, Barrett et al ⁸⁹, IMGSAC ⁹⁰, Shao et al ⁸⁶, Alarcon et al ⁸², Auranen et al ⁸⁴, McCauly et al ⁸⁷, and Lamb et al ⁸⁸. The region 7q31-33, which showed evidence of linkage in the study by IMGSAC ⁹⁰, has had a physical map constructed, and there are 23 genes in this region alone ⁹¹. This region has also shown increased evidence of linkage when the analyses were confined to individuals showing delayed speech acquisition and language impairments. Barrett et al found that linkage (measured by MLS scores) in the 7q31-33 region rose from 1.4 to 2.2 in the impaired language sample ⁸⁹. In the sample showing no language impairment, the MLS score fell to 0.1 from 1.4. Another possible “speech acquisition” locus has been reported on chromosome 2q ^{80 92}.

The region around the serotonin transporter on chromosome 17q11.2, has also been identified in a number of genome scans (see section 1.3.1) ⁹³⁻⁹⁵, as have regions on chromosome 16 ^{77-79,87,88}. Considering there is a 4:1 male: female affected ratio, it is not surprising that a number of studies have reported nominal linkage to the X chromosome ^{78,79,83,85,86}.

Although on first inspection the number of chromosomal regions highlighted from genome scans seems wide and varied, and most “linked” regions do not seem to be reliably replicated, the number of loci identified supports the theory that the mode of

inheritance for autism is polygenic (see Section 1.2.3). There may also be a small number of population specific “minor” autism regions, e.g. the 3q25-27 region may be specific to the Finnish population in which it was identified⁸⁴.

1.3.2 Chromosomal Abnormalities

Chromosomal abnormalities are estimated to be present in approximately 3% of individuals with autism¹⁶. Chromosomal abnormalities have been reported on all chromosomes with the exception of chromosome 11 and chromosome 21 (see Table 1.3). These include autosomal aneuploidies, deletions, duplications, translocations, ring chromosomes, inversions and supernumerary marker chromosomes.

The most common chromosomal disorder in autism occurs as a trisomy of chromosome 15q11-q13⁹⁶⁻¹⁰⁵. This region undergoes imprinting and is associated with Prader-Willi syndrome (PWS) and Angelman syndrome (AS). PWS is characterised by muscular hypotonia, mental retardation, short stature, hypogonadotropic hypogonadism and small hands and feet (OMIM reference 176270). It is a result of the deletion of a gene/genes on the paternal chromosome 15q11-q13, or by maternal uniparental disomy of this region, as this region on the maternal chromosome is rendered inactive due to imprinting. These lead to complete loss of expression of genes in the region. AS is characterised by severe motor and intellectual retardation, ataxia, hypotonia, epilepsy, absence of speech and characteristic facial structure (OMIM reference 105830). 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 proteasome complex. It is expressed exclusively on the maternal chromosome in the brain¹⁰⁶. Deficiencies in the gene may arise through maternal interstitial deletions, paternal uniparental disomy or a single imprinting gene defect in *UBE3A*.

Due to the high frequency of chromosomal abnormalities in autism reported (see Table 1.3), there have been a number of genetic association studies undertaken. Two studies have reported associations in the GABA receptor B3 gene ¹⁰⁷ and the GABA receptor subunit G3 ¹⁰⁸. A number of studies have also reported evidence for linkage to markers in the 15q11-13 region ¹⁰⁹⁻¹¹². Linkage studies are somewhat hampered by the high level of recombination observed in this region ¹¹². Evidence for genetic association has also been found with variants in the *UBE3A* gene and individuals with autism ¹¹¹.

As mentioned above, chromosome 7q has been identified as showing evidence for linkage by several genomewide linkage analyses (see section 1.3.1). Vincent et al reported a case of autism with a translocation between chromosomes 7 and 13 (Ch7q31 t(7;13)(q31.2;q21)) (see Table 1.3). This led to the disruption of a tumour suppressor gene, suppressor of tumorigenicity 7 (*RAY1*) ¹¹³. Although this translocation may have affected the proband in question, the *RAY1* gene was shown to be unlikely to have a major influence in the development of autism, as no mutations were found in 27 unrelated individuals with autism.

Another gene, *AUTS2*, at 7q11.2, was identified at a breakpoint reported in a pair of monozygotic twins concordant for autism. These twins carried a translocation between chromosomes 7 and 20 (t(7;20) (q11.2; p11.2)) ¹¹⁴. The *AUTS2* gene is a large gene, of 1.2 Mb with 19 exons. Transcription of this gene produces a protein of 1295 amino acids, which does not correspond to any known protein. Exons and flanking intron sequences were screened for mutations in 65 nuclear multiplex autism families, and while a number of polymorphisms were found, there was no evidence for association of any of these markers in the wider autism sample. Linkage analysis of four dinucleotide repeat markers (two within and two flanking the gene) was also negative ¹¹⁴.

Other chromosomal abnormalities include those on chromosome 2q13 to 2q36 (see Casas et al for a review of 66 deletions in individuals with autism in this region ¹¹⁵)

and Fragile X abnormalities (see Table 1.3). While cytogenetic findings so far have not led to the identification of autism related genes, this method has proved successful in other diseases, including the identification of the dystrophin gene in Muscular Dystrophy ¹¹⁶, the *UBE3A* gene in Angelman Syndrome ¹¹⁷ and the SNRPN/ neccin candidate genes in Prader-Willi Syndrome ¹¹⁸.

Chromosome	Abnormality	Reference
1	T(1;7)(p22;q21)	Yan et al ¹¹⁹
	Uniparental disomy 1p22	Wassink et al ¹²⁰
2	Deletion of 2q37.3	Lukusa et al ¹²¹
	Review of 66 cytogenetic abnormalities between 2q36 and 2q37.3	Casas ¹¹⁵
3	Partial tetrasomy of chromosome 3	Oliveira et al ¹²²
	T(3;12)(p26.3;q23.3)	Fahsold et al ¹²³
4	Interstitial deletion on 4q	Ramanathan et al ¹²⁴
	46,XX, dupl.(4)(p12-p13)	Sabaratnam et al ¹²⁵
	(4;12)(q21.3;q15)	Nasr et al ¹²⁶
	T(4;6)(q23-24;p21)	Liu et al ¹²⁷
5	A <i>de novo</i> microdeletion on 5p	Harvard et al ¹²⁸
	T(5;18)(q33.1;q12.1)	Kroisel et al ¹²⁹
5	T(5;7)(q14;q32)	Tentler et al ¹³⁰
	Monosomy of (5pter-5p15)	Vostanis et al ¹³¹
6	Partial 6p trisomy	Burd et al ¹³²
	Del(6)(q25.1;q25.3)	Sukumar et al ¹³³
	T(4;6)(q23-24;p21)	Liu et al ¹²⁷
7	7q31 partial duplication	Sarajee et al ²⁹
	T(7;20)(q11.2;p11.2)	Sultana et al ¹¹⁴
	T(5;7)(q14;q32)	Tentler et al ¹³⁰
	Dupl(7)(p11.2-p14.1)	Wolpert et al ¹³⁴
	T(1;7)(p22;q21)	Yan et al ¹¹⁹
	T(7;13)(q31.3;q21)	Vincent et al ¹¹³
8	R(8)p(10p23.1)	Demori et al ¹³⁵
	T(2;8)(q35;q21.1)	Borg et al ¹³⁶
	Translocation breakpoint in releasing peptide Receptor	Ishikawa-Brush et al ¹³⁷
	T(X;8)(p22.13;q22.1)	Bolton et al ¹³⁸
9	Two cases of pericentric inversion of chromosome 9	Liu et al ¹²⁷
10	del(10p14-pter)	Verri et al ¹³⁹
12	T(3;12)(p26.3;q23.3)	Fahsold et al ¹²³
	(4;12)(q21.3;q15)	Nasr et al ¹²⁶

Table 1.3A A LIST OF CHROMOSOMAL ABNORMALITIES REPORTED IN INDIVIDUALS WITH AUTISM. T = translocation, Del = deletion, R = ring chromosome (Analysis of Abstracts undertaken).

Chromosome	Abnormality	Reference
13	Del13(q12-13)	Smith et al ¹⁴⁰
	Del13(q14-q22)	Steele et al ¹⁴¹
14	Del 14(q32.3)	Merritt et al ¹⁴²
15	See Bolton et al for review on chromosomal abnormalities of q11-q13 ¹⁴³	
16	Dup (16)(p11.2-p12.2)	Finelli et al ¹⁴⁴
	16q22 fragile site	Kerbeshian et al ¹⁴⁵
	Partial trisomy 16p	Hebebrand et al ¹⁴⁶
17	Dup17(p11.2p12)	Moog et al ¹⁴⁷
	T(17;19)(p13.3;p11)	Anneren et al ¹⁴⁸
	Del17(p11.2p11.2)	Vostanis et al ¹³¹
18	T(5;18)(q33.1;q12.1)	Kroisel et al ¹²⁹
19	T(17;19)(p13.3;p11)	Anneren et al ¹⁴⁸
20	Del20(p12.2-pter)	Sauter et al ¹⁴⁹
	t(20:22)(q13.3;q11.2)	Carratala et al ¹⁵⁰
	Del20(p11.2p11.23)	Michaelis et al ¹⁵¹
22	R(22)	Nair-Miranda et al ¹⁵²
	Del22(q13.3)	Goizet et al ¹⁵³
	R(22)(p11.31-q13.31)	MacLean et al ¹⁵⁴
	T(20:22)(q13.3;q11.2)	Carratala et al ¹⁵⁰
X	Monosomy X	Donnelly et al ¹⁵⁵
	Xp deletion	Thomas et al ¹⁵⁶
	T(X;8)(p22.13;q22.1)	Bolton et al ¹³⁸

Table 1.3B A LIST OF CHROMOSOMAL ABNORMALITIES REPORTED IN INDIVIDUALS WITH AUTISM. T = translocation, Del = deletion, R = ring chromosome (Analysis of Abstracts undertaken).

1.3.3 Animal Models

The use of animal models can lead to insights into the molecular basis of a disorder. This can be achieved by highlighting the role of individual genes, molecular pathways, potential mutagens or regions of the brain involved in the disorder.

A number of animal models have been useful in helping to pinpoint regions of the brain that may be involved in autism. This is important, as the low prevalence of the disease and the infrequency of autopsies, have made the study of brain structure associated with autism difficult. Induced lesions in the amygdala have produced autistic behaviours in monkeys¹⁵⁷. Bilateral temporal lesions were have led to abnormal social interactions, changes in body and facial expression and the development of abnormal stereotypic behaviours¹⁵⁸. Surprisingly these abnormalities only lasted until adulthood, whereupon the lesioned monkeys appeared no different to control monkeys¹⁵⁸. Based on this work, it has been suggested that the focus of gene searches for autism should centre on genes expressed at birth in the temporal region of the brain. In another study, by Wolterink et al, administration of ibotenic acid (on day 7 or day 21) led to the development of lesions in the lateral and central amygdala and/or the ventral hippocampus in rats¹⁵⁹. Both sets of rats (i.e. those in the 7 and 21 day administration) exhibited enduring behavioural disturbance. However, those rats that received ibotenic acid earlier showed more extreme disturbances. This may have arisen due to the effect of ibotenic acid on the developing brain. Monkeys with induced lesions in the amygdala have also shown abnormal fear responses, exhibiting increased fear in novel social situations despite normal social development, and decreased fear in normal fear inducing situations (see Amaral for a review¹⁶⁰). Amaral hypothesises that “the amygdala is not necessary for species typical social behaviour or for gaining social knowledge during development, but it is a critical component of a system that evaluates the environment for potential dangers.... It typically inhibits social interaction with novel conspecifics while they are evaluated as potential adversaries”¹⁶⁰. More recently face and emotion recognition deficits have been reported and attributed to

anomalies in amygdala development”^{161,162}. 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 deficits¹⁶³. The role of the amygdala is to mediate inborn and acquired emotional responses, and therefore it is a very interesting region for study in autism.

In an autopsy on a brain from a 21-year-old female patient with autism by Rodier et al, a significant decrease in the number of motor neurons in the facial nerve nucleus was reported (from 9,000 neurons in a control brain to ~400 in the autistic brain)¹⁶⁴. Shortening of the middle pontine regions was also noted. Rodier commented on the similarity of these abnormalities to those described in *HoxA* knockout mice. Exposure of rats to valproic acid (a powerful chemical teratogen) on day 11.5 resulted in a reduction of the formation of motor nuclei V and XII, but not of the facial nerve VII¹⁶⁵. Functional abnormalities of the cranial nerves are observed with higher frequency in autism⁷³. Exposure to this teratogen on day 12.5 resulted in a decrease of the posterior lobe of the vermis. Between the days 11.5 and 12.5, the neural tube closes. Andres interpreted these results by stating that “early exposure, at the period of neural tube closure, can lead to late effects, which are relatively limited and specific”¹⁶⁶. Other possible autism related teratogens include thalidomide. 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 dopamine¹⁶⁷. Thalidomide exposure in humans has also been reported to lead to an incidence of autism in 5 out of 15 cases^{73,166}. Valproic acid administration to rat foetuses on the 12.5 day of gestation resulted in the birth of rats who exhibited:

- 1) Decreased number of social behaviours and increased latency to social behaviour
- 2) Locomotor and repetitive/ stereotypic-like hyperactivity combined with lower exploratory activity

- 3) Lower sensitivity to painful stimuli with a higher sensitivity to non-painful stimuli
- 4) Diminished acoustic prepulse inhibition.

These behavioural aspects show similarities to autistic behaviour.

Biological chemicals, e.g. oxytocin and vasopressin, have also shown links to autism¹⁶⁸. Both of the aforementioned biochemical molecules have shown to be involved in communication and ritual and social behaviour traits that are considered to be aberrant in individuals diagnosed with autism, and therefore are strong candidates for dysregulation in autism^{166,168}. In the case of oxytocin, knockout mice for this gene exhibit profound deficits in social recognition and processing¹⁶⁸. The administration of oxytocin to infant rats also leads to decreases in social interactions. These rats show a decrease in the number of isolation calls, fail to develop social memory and exhibit stereotypic behaviour¹⁶⁹. Rats lacking vasopressin and mice lacking the Vasopressin V1a receptor also show deficits in social processing and recognition¹⁷⁰. There are no reported mutations in the oxytocin gene, although observations of decreased levels of plasma oxytocin in individuals with autism have been published^{171,172}.

Other mouse models have been described that may be relevant for autism, including *TPH2* (see section 3.1.5), *Wnt3*, *Dvl* and *En2* (see Section 6.1 for more information).

1.3.4 Clinical Observations

Clinical observations and trials have helped identify or strengthen the case for candidate genes. The administration of selective serotonin reuptake inhibitors (SSRIs), which prevent serotonin from being reabsorbed back into the presynaptic nerve, have shown to be of clinical benefit in some patients (see Moore et al for a review⁷⁵). SSRIs may lead to improvements in mood and temperament, ability to relate to others, language impairment and decreases in compulsive behaviours.

These observations have strengthened the case for the involvement of the serotonin pathway in autism (see Section 3.1.2 for further information). Likewise, dopamine agonists, e.g. haloperidol and risperidone, have shown to lead to some improvements in irritability and severe behavioural problems associated with autism, and this has given support to the role of dopamine and/or the dopamine pathway in autism ^{173 174}.

Within any sample of individuals with autism, there will be variability in the phenotypic presentation of autism, e.g. some individuals will have more repetitive behaviours than another individual in the same sample. Subdivision of samples is becoming more frequently used to try to approximate a single biological disturbance. With respect to phenotypic measures, the clinical data generated by the ADI-R ¹⁴ and ADOS-G ¹⁷⁵ 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 ¹⁷⁶. A study by Shao et al found increased linkage (from 1.5 in the general sample to 4.71 in the sub-phenotyped sample) at the *GABRB3* locus in the chromosome 15q11-13 region when the sample was reanalysed only including individuals with a high score on the “insistence on sameness” category ¹⁷⁷.

1.4 Testing Potential Autism Susceptibility Genes

Having identified potential candidate genes, a number of strategies exist in order to test if the gene is associated with a disorder. They include association testing, linkage disequilibrium (LD) mapping and haplotype testing, mutation screening and more recently expression profiling.

1.4.1 Linkage Disequilibrium Mapping, Association Testing and Haplotype Analysis

1.4.1.1 The Theory

Whereas linkage studies can now quite easily take a genome-wide approach, association studies are limited to candidate genes or regions. In the case of complex polygenic disorders, the power of a linkage study can be relatively weak, as each functional variant may only play a minor, but statistically significant, role in the development of the disease. Association studies have a greater power to detect variants of small magnitude in the same sample¹⁷⁸. An increased marker density is needed in the case of association studies in comparison to linkage studies due to the decreased PIC (polymorphic information content) gained from SNPs (the majority of variants tested in association studies), in addition to the fact that association can only be detected over a shorter range compared to linkage analysis. The basic aim of an association study is to test if a given marker allele occurs at a significantly higher frequency in a sample of cases (affected individuals) in comparison to controls (unaffected individuals). Associations can be sought with functional markers, as in the Serotonin transporter promoter insertion/deletion (see Section 3.1.4), or in variants that are in linkage disequilibrium (LD) with functional markers.

Linkage disequilibrium (LD) is defined as the non-random association of alleles at adjacent loci ¹⁷⁹. It occurs when an allele at one locus is found to occur more often than expected by chance with another allele at another locus. LD arises following the creation of a new mutation on a chromosome that carries a particular set of alleles in the loci surrounding it. Over time the degree of LD diminishes, due to increased recombination events. While the extent of LD is expected to decrease over time and distance between markers, not all closely linked markers are in LD, and some markers with a large inter-marker distance may show high LD. There have been reports of high LD between markers that are over 100kb apart ¹⁸⁰, whereas marker that are less than 2kb apart may show no LD ¹⁸¹. There are a number of factors that can influence the extent of LD across a region. These include allele frequency, genetic drift, population growth, admixture or migration, population structure, natural selection, variable recombination rates, variable mutation rates and gene conversion events (see Ardlie et al for a review on these factors ¹⁷⁹).

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 A (with alleles A and a) and B (with alleles B and b), the frequency of a two-marker haplotype is P_{AB} . Assuming random assortment, $P_{AB} = (P_A \times P_B)$. Thus D can be calculated using the following equation

$$D = (P_{AB} - (P_A \times P_B))$$

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 it's maximum possible value, given allele frequencies at two loci. It is the most widely used measure of LD. 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 ¹⁸².

r^2 is another measure for quantifying and comparing LD in association mapping ¹⁸³ ¹⁸⁴. 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:

$$r^2 = \frac{D^2}{P_{AB} + P_{Ab} + P_{aB} + P_{ab}}$$

$r^2=1$ if the two loci have not been separated by recombination and have the same allele frequency. An advantage of the measure is that individual values of r^2 are more easily interpreted. Also for two neighbouring loci where one is a disease causing allele and the other a nearby marker that is in $LD = r^2$, to have the same power to detect the association between the disease and the neighbouring marker, a sample size should be increased by a factor of $(1/r^2)$ in order to detect association between the marker and disease when compared with the sample size required to detect association with the disease locus itself ^{183,185}. Furthermore, r^2 takes account of differences in allele frequencies between two markers. Finally, the calculation is less inflated in small samples.

There are two forms of association testing, the case-control test and a family based association test. The original and most widely used study design for testing association is the case control design. This method consists of testing for significant allele and genotype differences between 2 samples, the cases (i.e. the affected individuals) and the controls. The control sample is matched as close as possible for age, gender, ethnicity, and socio-economic background. A statistical difference between the groups, tested using a basic χ^2 test, implies an association between the tested allele of the variant and the phenotype/disorder. Despite its ease however, this

method is liable to identify ‘false-positive’ associations, due to population stratification ¹⁸⁶. This is because allele frequencies differ between populations irrespective of disease status, based upon a population’s unique genetic and social history. Population stratification occurs when the case and control samples are not correctly matched, and this may not be easy to spot in advance of a study.

Another method has been designed to avoid this potential problem. This design, known as the family based or trio study design, is based on families where the sample consists of at least one affected child and one or both biological parents. It is similar to the case-control test in that it tests for differences between sample and control allele frequencies. This approach tests for the deviation from the expected transmission of alleles from parents to their affected offspring (i.e. the null hypothesis or Mendel’s first law). If an allele is significantly over-transmitted to probands then this marker becomes a potential susceptibility allele. The statistics used to test for significant transmission are the Transmission Disequilibrium Test (TDT) (see Section 2.3.1.2) ¹⁸⁷ and the Haplotype Based Haplotype Relative Risk (HHRR) ¹⁸⁸. The TDT test is a test for association in the presence of linkage.

There are a number of disadvantages to this method of association testing. Firstly, it is not always possible to collect parental samples, as is the case in late onset disorders, e.g. Alzheimer’s disease. It necessitates extra genotyping, as DNA extraction and genotyping have to be performed in the proband and both parents, versus a case individual and control individual. Finally, there is a loss of power associated with using the TDT test, as TDT calculation can only be performed in heterozygous parents and thus only a proportion of the sample can be utilised with a consequent loss of power. The HHRR test uses all transmission observations but it is rarely more informative.

It was once thought that LD was relatively consistent across the genome. Recent studies have shown this not to be the case. There are regions of high LD and low recombination interspersed with regions of low LD and high recombination (known

as ‘recombination hotspots’¹⁸⁹). This phenomenon has results in the formation of what have become known as haplotype ‘blocks’. The HapMap project (<http://www.hapmap.org/>) defines haplotype blocks as regions in the genome between 10 and 100kb that have between 2 to 4 haplotypes defining > 90% of all chromosomes. The aim of the HapMap project is to “compare the genetic sequences of different individuals to identify chromosomal regions where genetic variants are shared”. Awareness of the underlying haplotype structure may reduce the quantity of genotyping necessary over a region of interest, as one or a small number of variants might substantially represent the total genetic information contained within each block.

Testing the transmission of haplotypes to affected individuals is also useful. Regions between markers often ultimately contain the functional marker of interest. Transmission of this 2-marker haplotype may be more significant than transmission of either of the single markers alone. Transmission of haplotypes may also help identify regions of genes that may contain these functional markers, leading to more efficient mutation screens.

1.4.1.2 Potential Candidate Genes Highlighted by Association Studies

Association testing has led to the identification of a number of possible genes for autism (see Table 1.4 for a sample of these genes). Genes with functions ranging from neurotransmission processes, e.g. the serotonin transporter gene¹⁹⁰, to brain development proteins, e.g. EN2¹⁹¹, to mitochondrial genes, e.g. the aspartate/glutamate carrier *SLC25A12*¹⁹², have all shown association with autism.

A number of gamma-aminobutyric acid (GABA) receptors (GABRs) have shown association with autism (see Table 1.4). GABA is one of the major inhibitory

neurotransmitters of the brain. There are 18 known GABA receptors composed of 21 GABA subunits¹⁹³, a number of which lie on chromosome 15q11-13, a region that is the most frequently observed chromosomal abnormality in individuals with autism (see Sutcliffe & Nurmi for a review¹⁹⁴). The GABA related genes are of interest, due to the inhibitory action of GABA, and the increased levels of glutamic acid and decreased levels of glutamine in samples of autistic individuals^{195,196}. A more recent study by Dhossche et al has used a gas chromatography / mass spectroscopy technique to measure GABA levels in a sample of patients diagnosed with autism¹⁹⁷. This has shown an increased plasma GABA level in a sample of patients with autism in comparison to a sample of controls. Decreased binding of radio-labelled GABA receptor ligands has also shown to be significantly reduced in the hippocampal region of the brain in individuals with autism¹⁹⁸. A number of association studies of GABA related genes have been undertaken, and positive associations have been reported, including the *GABRA5*¹¹⁰, *GABRB3*^{199 110} and *GABRG3*¹⁰⁸ genes. One of the largest investigations of the GABA receptor system and its relationship to autism was undertaken by Ma et al, who analysed the transmission of alleles of 70 SNPS across 14 genes²⁰⁰. Allelic association was found with the *GABRA2* and the *GABRR2* genes, while genotype association was found with the *GABRA2* and the *GABRA4* genes. Gene-gene interactions between *GABRA4* and *GABRB1* were also reported. Ma et al concluded that *GABRA4* is involved in the development of the autistic aetiology, and the inheritance of certain *GABRB1* variants increases this risk further still²⁰⁰.

Another gene that has shown association with autism is the *Reelin* gene, located on chromosome 7q22²⁰¹ (see Table 1.4). It is a large gene encoding mRNA of approximately 12kb²⁰². The gene product is a large extracellular glycoprotein secreted by the most superficial layers of the brain. The expression in postnatal human brain is high in the cerebellum, a region shown to be enlarged in autism (see Nicolson et al for a review⁴⁰). 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*)²⁰³.

In relation to autism, evidence for linkage around marker D7S477 that maps to 7q22 has been reported ⁸¹. 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 ²⁰⁴. 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 ^{205,206}. However Zhang et al have proposed that there is an increase in transmission of the variant in autistic individuals with delayed phrase speech. There have also been reports of decreased blood levels of reelin ²⁰⁷ in addition to reduced mRNA in the frontal cerebellum region of post-mortem brain from a 21 year-old female diagnosed with autism in comparison to a normal age matched control ²⁰⁸.

Other genes with tentative evidence for association with autism, although the strength of replications often remains complex, include the serotonin transporter (see Section 3.1.4), *WNT2* (See section 6.1.3), and the *HOX* genes ²⁰⁹.

Gene	Reference	Origin of Sample
<i>5-HTT</i>	Conroy et al ²¹⁰	Irish
<i>5-HTT</i>	Cook et al ¹⁹⁰	American
<i>5-HTT</i>	Kim et al ³	American
<i>5-HTT</i>	Klauck et al ⁴	Data Unavailable
<i>5-HTT</i>	Mulder et al ²¹¹	Dutch
<i>5-HTT</i>	Tordjman et al ²¹²	French
<i>5-HTT</i>	Yirmiya et al ²¹³	Israeli
<i>ABAT</i>	Barnby et al ²¹⁴	European
<i>ADA</i>	Bottini et al ²¹⁵	Italian
<i>ADA</i>	Persico et al ²¹⁶	Italian
<i>APOE2</i>	Persico et al ²¹⁷	Italian
<i>EN2</i>	Gharani et al ¹⁹¹	AGRE
<i>EN2</i>	Petit et al ²¹⁸	French
<i>FOXP2</i>	Gong et al ²¹⁹	Chinese
<i>GABRA2</i>	Ma et al ²⁰⁰	American / AGRE
<i>GABRA4</i>	Ma et al ²⁰⁰	American / AGRE
<i>GABRA5</i>	McCauley et al ¹¹⁰	American / AGRE
<i>GABRB3</i>	Buxbaum et al ¹⁹⁹	American / AGRE
<i>GABRB3</i>	McCauley et al ¹¹⁰	American / AGRE
<i>GABRG3</i>	Menold et al ¹⁰⁸	Data Unavailable
<i>GABRR2</i>	Ma et al ²⁰⁰	American / AGRE
<i>GluR6</i>	Shuang et al ²²⁰	Chinese
<i>GluR6</i>	Jamain et al ²²¹	European
<i>GRIN2A</i>	Barnby et al ²¹⁴	European
<i>HOPA</i>	Beyer et al ²²²	German / Austrian
<i>HOXA1</i>	Conciatori et al ²²³	Italian / AGRE
<i>HRAS</i>	Comings et al ²²⁴	American
<i>HRAS</i>	Herault et al ²²⁵	French
<i>INPP1</i>	Serajee et al ²⁹	AGRE
<i>LAMB1</i>	Hutcheson et al ²²⁶	American
<i>MAO-A</i>	Cohen et al ²²⁷	American
<i>NF1</i>	Marui et al ²²⁸	Japanese
<i>NLGN1</i>	Ylisaukko-Oja et al ²²⁹	Finnish
<i>NLGN3</i>	Ylisaukko-Oja et al ²²⁹	Finnish
<i>NLGN4</i>	Ylisaukko-Oja et al	Finnish

Table 1.4A SAMPLE OF GENES SHOWING POSITIVE ASSOCIATIONS WITH AUTISM. These genes were chosen following PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) searches using the terms “genetic associations AND autism” and “associated genes AND autism”, (Abstracts only).

Gene	Reference	Ethnicity of Sample
<i>OXTR</i>	Wu et al ²³⁰	Chinese
<i>PIK3CG</i>	Serajee et al ²⁹	AGRE
<i>PON1</i>	D'Amelio et al ²³¹	American / Italian
<i>PRKCB1</i>	Philippi et al ²³²	Data Unavailable
<i>RELN</i>	Persico et al ²³³	Italian
<i>RELN</i>	Skaar et al ²³⁴	American / Italian
<i>SLC25A12</i>	Ramoz et al ¹⁹²	American / AGRE
<i>TDO2</i>	Nabi et al ²³⁵	AGRE
<i>TPH2</i>	Coon et al ⁷	American / Italian
<i>TSC2</i>	Serajee et al ²⁹	AGRE
<i>UBE2H</i>	Vourc'h et al ²³⁶	Data Unavailable

Table 1.4B SAMPLE OF GENES SHOWING POSITIVE ASSOCIATIONS WITH AUTISM.

These genes were chosen following PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) searches using the terms “genetic associations AND autism” and “associated genes AND autism”, (Abstracts only).

1.4.2 Mutation Screening

1.4.2.1 The Theory

Mutation screening, either by denaturing high performance liquid chromatography (dHPLC) (see Section 2.2.5.2), single strand conformational polymorphism (SSCP) analysis (see Section 2.2.5.1 for further information) or direct sequencing, is one method used in the search for novel or rare mutations. Screening for large deletions and insertions can be easily undertaken by a basic PCR reaction and searching for product size differences, however screening for single base pair changes (SNPs) or small deletions/insertions is more difficult. It is these minor changes that appear to predispose individuals to the majority of complex diseases.

The majority of mutation screens are undertaken in exonic regions of the candidate genes, and often include intron/exon boundaries, as changes in the DNA sequence in these regions are more likely to have an adverse effect on the structure of the resulting protein, through amino acid changes, or adverse splicing effects. Direct sequencing of these regions is perhaps the most straightforward method for identifying novel mutations.

Another method of mutation screening is the SSCP method. This is based on the principle that short DNA sequences that differ in sequence will migrate differently in a non-denaturing polyacrylamide gel ²³⁷. A disadvantage of SSCP analysis is that only small DNA sequences can be analysed (<200bp) with a detection efficiency of 90%. Sequences up to 400bp in size have efficiency between 70% and 80% ²³⁷. Varying experimental conditions, e.g. temperature and glycerol concentration, or using RNA molecules, can increase sensitivity. Another disadvantage is the possible missing of mutations and the fact that any sample showing a variation will then have

to be sequenced so the exact nature and location of the mutation is known. That said however, it is a quick, cheap and non-technical method of searching for mutations.

Another mutation detection technique is dHPLC²³⁸. This is used to detect single base pair substitutions, and small insertions and deletions. Heteroduplexes, occurring in the presence of a variant, are retained less than homoduplexes, occurring when there are no variants present, on a unique DNA separation matrix, due the decreased stability of the DNA fragment. It has been estimated that the sensitivity and specificity of dHPLC is consistently higher than 96%²³⁹, thus making it a highly efficient and cheap method of detecting mutations in a large number of individuals. Sequencing of samples showing the presence of mutations is still required, but the number of samples necessitating sequencing is dramatically reduced.

1.4.2.2 Potential Candidate Genes Highlighted by Mutation Screens

A number of mutation studies have been undertaken following the identification of a possible candidate gene for autism (see Table 1.5). One such gene is the *FOXP2* gene (also known as *SPCHI*). This gene was first identified as being responsible for a rare, severe speech and language disorder which was found to segregate in a dominant manner in a three-generational family (known as the KE family). Following the highlighting of 7q31 as a candidate region in a genome scan in the KE family, the Forkhead box P2 gene (*FOXP2*), a transcription factor, was screened for mutations²⁴⁰. A single base pair change in exon 14, in the forkhead domain of this gene, co-segregated with the severe language defects. Due to its relationship to speech and language defects and its presence in a region on chromosome 7 showing widespread evidence for linkage (see section 1.3.1), this gene became a potential gene for autism. A translocation in an individual with autism was also reported to cause disruption to this gene²⁴¹. Wassink et al sequenced exon 14 of this gene and performed SSCP analysis in a sample of 75 ASPs diagnosed with autism²⁴². They

identified 2 non-conservative changes, both involving the deletion of glutamine residues from a polyglutamine stretch occurring in 2 families with autism, which were not present in 160 control samples. No mutations were found to occur in exon 14, and there was one silent mutation in exon 5. There was a trend towards significant transmission distortion of a tetra-nucleotide repeat in intron 1 in individuals presenting with language abnormalities ($p = 0.06$). Newbury et al also screened the coding sequence of 43 ASPs who presented with autism²⁴³. Eleven new variants were detected, although all were intronic, and no mutations were found in the exons 12 – 14 (region giving rise to forkhead domain). A CAG-CAG insertion was also discovered in the exon 16 / intron 16 border. A number of other mutation screening experiments in samples with autism have also been reported^{219,244-246}. A small number of mutations have been discovered, including silent mutations²⁴⁵, mutations in exon/intron junctions²⁴⁵ and intronic SNPs²⁴⁴.

Following the identification of mutations, association studies are often carried out with the mutation in question, in addition to other more common SNPs throughout the gene. There have been conflicting results for association studies using common SNPs in this gene. Gauthier et al reported no association²⁴⁵, as did O' Brien et al in the overall sample²⁴⁴, but an association between marker D7S3052 was reported in individuals with severe language deficits. Marui et al also reported an association between a SNP in intron 7 in male probands, but not female probands²⁴⁶. Gong et al found over-transmission of a 2-marker haplotype in a sample of Japanese probands with autism²¹⁹. Overall, the evidence does not support a major role for *FOXP2* in the development of autism. It may however be important in a small subset of individuals with severe language impairments.

The neuroligin (*NLGN*) family of genes has also been investigated by mutational analysis (see Table 1.5). There are five *NLGN* genes, located on chromosomes 3q26 (*NLGN1*), 17p13 (*NLGN2*), Xq13 (*NLGN3*), Xp22.3 (*NLGN4*) and Yq11.2 (*NLGN4Y*)²⁴⁷. One of these genes, *NLGN3* is located within a region of chromosome Xp22.3, which was reported in a *de novo* deletion in a patient with

autism¹⁵⁶. Two linkage studies have also highlighted Xq13 as a region showing increased, but not significant, allele sharing between affected sib-pairs^{84,86}. Mutation screening of the *NLGN3*, *NLGN4* and *NLGN4Y* was undertaken in a Finnish sample of 36 ASPs and 122 trios diagnosed with either autism or Asperger Syndrome reported the presence of two mutations that segregated with autism²⁴⁷. One mutation introduced a stop codon at position 396 in the *NLGN4* gene, while the second resulted in the conversion of arginine 451 to cysteine 451 (a highly conserved amino acid). *In vitro* experimentation has shown that these two mutations lead to the intracellular retention of both proteins, which in turn leads to reduced formation of presynaptic elements in cultured hippocampal cells²⁴⁸. There have been a number of reports of attempted replication. One study of 148 patients diagnosed with autism (Portuguese and American) found 4 missense mutations in the *NLGN4* gene, all in unrelated individuals²⁴⁹. However, one of these mutations was also shown to segregate to 2 unaffected siblings of the proband with autism, thus making its role in the development of autism unclear. There has also been three studies that have failed to find any mutations in the *NLGN* genes²⁴⁹⁻²⁵¹. An association study using common SNPs in the *NLGN1*, *NLGN3*, *NLGN4* and *NLGN4Y* genes with autism has been undertaken in 100 Finnish families²²⁹. Modest associations were reported for *NLGN1* (p value = 0.002), *NLGN3* (p value = 0.014) and *NLGN4* (p value = 0.031). Further studies need to be undertaken to test common SNPs in these genes in other samples. The functional relevance of the mutations found in the two Finnish families point to the importance of this family of genes to the development of autism. It remains to be seen what the role of these genes is to larger samples diagnosed with autism.

Mutation screening has also led to the examination of a number of other genes and their role in the development of autism (see Table 1.5). They include the *RAY1* (also known as ‘suppressor of tumorigenicity 7’)^{113,252} and *WNT2* genes (see Section 6.1.3).

Gene	Author	Origin of Sample
<i>ABAT</i>	Barnby et al ²¹⁴	European
<i>cAMP-GEFII</i>	Bacchelli et al ²⁵³	European
<i>CENTG2</i>	Wassink et al ²⁵⁴	American
<i>FMR-1</i>	Shinahara et al ²⁵⁵	Japanese
<i>FMR-1</i>	Vincent et al ²⁵⁶	AGRE
<i>GRIN2A</i>	Barnby et al ²¹⁴	European
<i>HOXA1, HOXB1</i>	Ingram et al ²⁰⁹	American
<i>LAMB1</i>	Bonora et al ²⁵⁷	European
<i>MECP2</i>	Lam et al ²⁵⁸	Chinese
<i>MECP2</i>	Li et al ²⁵⁹	Japanese
<i>MECP2</i>	Shibayama et al ²⁶⁰	American
<i>NBEA</i>	Castermans et al ²⁶¹	Belgian
<i>NLGN3, NLGN4</i>	Jamain et al ²⁴⁷	Swedish
<i>NLGN4</i>	Yan et al ²⁴⁹	Portuguese/American
<i>NLGN4</i>	Laumonier et al ²⁶²	French
<i>SLC25A12</i>	Ramoz et al ¹⁹²	American/AGRE
<i>UBE2H</i>	Vourc'h et al ²³⁶	Data Unavailable
<i>WNT2</i>	Wassink et al ²⁶³	American

Table 1.5 SAMPLE OF GENES THAT HAVE REVEALED POTENTIAL MUTATIONS ASSOCIATED WITH AUTISM. These genes were chosen following PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) searches using the terms “mutation screen AND autism”, (Abstracts only).

1.4.3 Expression Profiling

Expression profiling is the *in vitro* investigation of the expression or activity of large numbers of genes using microarrays. It is a form of whole genome analysis. In a single experiment over a thousand potential candidate genes can be narrowed down to focus on genes whose expression differs between affected and unaffected individuals. Entire pathways can now be screened for differences in expression in comparison to controls. It is a rapidly growing research technique, that has widely been used in cancer research ²⁶⁴. There has only been one report of expression profiling in autism. A study of the expression profile of genes involved in inflammatory responses in individuals with autism, was undertaken by Vargas et al ²⁶⁵. Seventy-nine genes, including cytokines, chemokines and growth and differentiation factors, were selected for analysis in 11 post-mortem brain samples or CSF samples from 6 individuals with autism in comparison to 9 post-mortem control brains and the CSF from 9 control individuals. A marked increase in neuroglial responses characterised by activation of the microglia and astroglia in the brain tissue of the patients with autism was described. These responses were most marked in the cerebellum. This could be of relevance to autism as microglia and astroglia interactions are essential for neuronal activity and synapse function ²⁶⁶, neuron-glia interactions ²⁶⁷ and cortical modelling, organisation and remodelling during brain development ²⁶⁸.

It is likely that as the technology required for expression profiling and the cost of such work decreases, this method of investigation will play an increasing role in

autism research, although access to the appropriate tissue samples will pose ethical and practical problems.

1.5 Aims of thesis

- 1) To construct hypothesis driven support for the testing of candidate genes. Such hypotheses will be based on published and unpublished information, e.g. linkage results, mouse models, the presence of chromosomal abnormalities, brain expression and biochemical links to autism.
- 2) Following the identification of a possible autism susceptibility gene, the gene will be tested for association as follows:
 - a) All genes will be tested for allelic association using transmission disequilibrium testing (see Section X).
 - b) All genes will be tested for haplotypic association using either the TRANSMIT test (see Section X) or the UNPHASED test (see Section Y).
 - c) The structure of LD will be examined in all cases.
 - d) In the case of replication studies, i.e. the serotonin transporter gene and the tryptophan hydroxylase 2 genes (see Chapter 3), power tests will be undertaken to examine to potential of missing true (or false) positive results.
 - e) In the case of the Dopamine beta hydroxylase gene (DBH) (Chapter 4), DBH enzyme activity will be measured in fathers, mothers and probands. Comparisons will then be made, in addition to analysing the contribution of each variant to overall DBH activity.
 - f) The usefulness of haplotype tagging SNPs will be examined using the Integrin alpha 4 gene (ITGA4) data (see Chapter 5).
 - g) Replication in additional samples will be sought to test for genetic associations in other samples where possible, i.e. the ITGA4 gene (see Chapter 5).
 - h) Mutation screening will be undertaken in order to replicate previous findings, i.e. in the WNT2 gene (see Chapter 6), and to identify potential mutations in exons close to associated variants, i.e. in the case of the ITGA4 associations (see Chapter 5).

CHAPTER TWO

MATERIALS AND METHODS

2.1 Sample Description

The sample described in this thesis was collected with ethical approval from the Eastern Regional Health Authority, Child and Adolescent Psychiatry Ethics Committee, Ireland. Affected autism trios (proband and both parents) were recruited through schools, advocacy groups and professionals. The sample described here was collected in two separate phases, and recruitment is ongoing.

Phase I included 101 affected individuals (77 male, 24 female), with a male to female ratio of 3.2:1. Parental samples were also collected where possible. 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). There were 2 incomplete trios, in which DNA was not available from the fathers in the family. The mean age of the affected individuals at the time of diagnosis was 123.7 months (range = 56 – 408 months, standard deviation = 102 months). This phase of the sample was used in the investigations of the serotonin transporter and dopamine- β hydroxylase (DBH) and autism (see sections 3.2.2 and 4.2).

Phase II included the original 101 samples and an additional 83 families. The male to female ratio was 4.11: 1 resulting from the addition of an additional 71 males and extra 12 females. The average age was 90.3 months (range = 46 – 408 months, standard deviation = 27.8 months). There were 148 males and 36 females in total.

Subjects were assessed by the ADI-R¹⁴, ADOS-G¹⁷⁵, neurological examination, skin examination with Wood's lamp and karyotyping/Fragile X testing. All subjects met the ADI-R criteria for autism and the ADOS-G criteria for autism/autism spectrum

disorder. Individuals were excluded if they had a known medical cause of autism (e.g. tuberous sclerosis, extreme prematurity, congenital rubella), IQ of 35 (or mental age of 18 months) or chromosomal abnormalities/fragile X.

2.2 Biological Techniques

2.2.1 DNA Extraction

2.2.1.1 DNA Extraction from Blood

Approximately 6ml of blood was collected from affected individuals, and their parents, in the study who consented to give blood. The blood samples were stored in a -70°C freezer until extraction. Before extraction, samples were thawed on ice.

The blood was added to 6.5ml of sterile water and 12.5ml of Lysis buffer (see Appendix I) in a 50ml falcon tube. The sample was placed on ice for 30 minutes, with occasional inversion. Following centrifugation at 3500rpm for 15 minutes, the supernatant was poured off, leaving $\sim 4\text{ml}$ of supernatant in the falcon tube. Lysis buffer was added to the sample to a volume of 25ml. The sample was re-incubated on ice for 10 minutes with occasional inversion before once again being spun at 3500rpm for 15 minutes. The supernatant was decanted off and the pellet resuspended in 2ml Suspension buffer (see Appendix I), before being transferred to a 15ml falcon tube, which contained containing $150\mu\text{l}$ 10% (w/v) SDS and $60\mu\text{l}$ Proteinase K (10mg/ml). The sample was incubated overnight at 37°C .

An equal volume of buffer saturated phenol was added to the samples, which were thoroughly mixed, and centrifuged at 3800 rpm for 10 minutes. The aqueous layer was removed and extracted twice more using 2ml of a 1:1 phenol chloroform

solution and pure chloroform respectively. Addition of 50 μ l 3M Sodium Acetate (pH 5) and 6ml ice cold 100% ethanol causes the precipitation of DNA, which was removed using sterile Pasteur pipettes to labelled eppendorf tubes. The DNA pellet was washed three times in 70% ethanol, air dried over three days and then resuspended in 250 μ l TE Buffer before quantification.

2.2.1.2 DNA Extraction of DNA from Cheek Cells

Cheek cell samples were taken from individuals, who were unable or unwilling to have blood taken for the study, Cheek cell samples were collected using sterile cytology brushes (CytoSoft, Medical Packaging Corporation).

The cheek swabs were stored in a fridge at 4 $^{\circ}$ C for a maximum of two weeks prior to extraction. The cheek swab brush heads were cut from the handles and placed in an eppendorf containing 500 μ l of 50mM NaOH and vortexed for 60 seconds. This was followed by incubation at 95 $^{\circ}$ C for 10 minutes, which in turn was followed by the addition of 50 μ l of 1M Tris (pH 8). The samples were then vortexed for 30 seconds. 200 μ l of Instagene Matrix (Biorad Labs, 2000 Alfred Nobel Dr., CA, USA) was added and the samples incubated for a further 30 minutes at 56 $^{\circ}$ C. The samples were mixed by vortexing for 10 seconds and heated for 8 minutes at 100 $^{\circ}$ C. The brush heads were removed from the tubes and discarded. The tubes were centrifuged at 6000rpm for 3 minutes. The supernatant, which contained the DNA, was transferred to a labelled tube and stored at 4 $^{\circ}$ C when not in use.

2.2.1.3 DNA Quantification

The stock DNA solutions were analysed by spectrophotometry at appropriate dilutions to determine concentration of DNA (260 nm) and RNA (280 nm). The DNA concentration (ng/μl) was calculated using the formula:

$$[\text{DNA}] = \text{OD}_{260} \times 50 \times \text{Dilution Factor}$$

Where OD_{260} is the absorbance of the sample at 260nm, and the dilution factor represents the dilution from stock of the test solution. The ratio of $\text{OD}_{260}:\text{OD}_{280}$ was used as a control for the presence of RNA.

2.2.2 Polymerase Chain Reaction (PCR)

Sixty nanogrammes of DNA was added to a PCR master mix containing 2.5μl of MgCl_2 PCR Buffer (X10), 4μl dNTP mix (200 μM dATP, dCTP, dGTP, dTTP), 20 pM of Forward and Reverse primers and 1U Taq Polymerase and made up to a final volume of 25μl (in the case of all reactions requiring RFLP digestion or mutation analysis). In the case of samples undergoing genotyping by SNaPshot method (Applied Biosystems, Foster City, CA, USA), the same reaction was prepared but all volumes were adjusted to a final volume of 10μl. All PCR reactions were performed on a MJ Research DNA Engine Thermal Cycler and consisted of an initial denaturing step of 4 min at 94°C typically followed by 30 cycles of 30 sec at 94°C, 30 sec at optimum annealing temperature and 30 sec at 72°C. This was followed in all cases by a final extension step of 10 min at 72°C. Details of variations to these conditions, primer sequences and of optimum annealing temperatures are provided in Appendix II.

2.2.3 Gel Electrophoresis and Visualisation

2.2.3.1 Agarose Gels

PCR products were visualised on 2% (w/v) agarose gels containing 0.02mg/ml ethidium bromide. The gels were run in TAE buffer (X1) at 80V. The gels were then viewed on an ultra violet illuminator and photographed using a Polaroid camera. PCR samples were mixed with 2µl loading dye prior to being loaded on to the gel along with appropriate size standards.

2.2.3.2 Non Denaturing Polyacrylamide Gels

Non-denaturing polyacrylamide gels were used to analyse some of the RFLPs. The 6% (w/v) gel was run at 80V after samples mixed with 1µl loading dye had been loaded on to the gel. The gels were stained with 0.02mg/ml ethidium bromide before being visualised on an ultra violet illuminator and photographed using a Polaroid camera.

2.2.3.3 Denaturing Polyacrylamide Gels for Use with the ABI PRISM 377 DNA Sequencer

These gels were used to visualise fluorescently labelled PCR fragments in conjunction with the ABI PRISM 377 DNA Sequencer. A 64-well 12% (w/v) gel (width = 2mm) was used for SNPs. The denaturing polyacrylamide gel was preheated to optimum temperature (51°C) on the sequencer before the samples mixed

with 1µl formamide loading dye were applied. The gel then underwent electrophoresis at 3000V (60 miliamperes; 200 Watts) using 1X TBE for sufficient time to allow the samples to run off the end of the gel (~ 1 hour 40 minutes), where their fluorescent label (6-FAM, HEX, NED or ROX) was read by laser and recorded on to an attached computer.

2.2.4 Genotyping

2.2.4.1 Insertion / Deletions (InDel) and Variable Tandem Nucleotide Repeats (VNTRs)

The Serotonin Transporter In/Del and VNTRs were amplified according to the PCR conditions according to the PCR conditions outlined in Appendix II. Samples were then loaded onto agarose gels and stained with ethidium bromide according to section 2.2.4.1.

2.2.4.2 Restriction Fragment Length Polymorphism (RFLP) Analysis

A number of markers (in the DBH and Serotonin Transporter genes) were analysed using the RFLP technique. This technique involves amplification of the region of the gene surrounding the marker of interest. Digestion of the PCR product using the appropriate restriction enzyme and buffer (supplied by New England Biolabs, Beverly, MA, USA) allows the determination of the allelic form of the marker. Details of the enzymes used and conditions of the digest can be found in appropriate Appendices. The appropriate enzymes in each case were either determined from the literature or obtained by electronically digesting the PCR product sequence using the

web based restriction analysis program WEBCUTTER 2.0
(www.firstmarket.com/cutter/cut2.html).

2.2.4.3 SNaPshot

The SNaPshot method of primer extension (Applied Biosystems, Foster City, CA, USA) employs the principle of extending an unlabelled oligonucleotide primer by 1 basepair in the presence of fluorescently labelled ddNTPs. Each of the 4 ddNTPs is tagged by a different fluorescent dye (carboxyfluorescein (ROX) = Red = dTTP, FAM™ = blue = dGTP, hexachlorofluorescein (HEX) = green = dATP and NED™ = yellow = dCTP). Hence, when the primer extension products are run on an ABI DNA Sequencer or ABI Genetic Analyser, the specific allele products can be differentiated from each other on the basis of which fluorescent dye they carry. The initial step of the SNaPshot primer extension reaction involves the treatment of the amplified PCR product with Shrimp Alkaline Phosphatase (SAP) (USB Corp., Cleveland, Ohio, USA) and *Exo I*, a 3'-5' Exonuclease (New England Biolabs, Beverly, MA, USA) to remove excess dNTPs and primers respectively (1U SAP, 1U *Exo I* and 0.5µl SAP dilution buffer per 5µl PCR product). Conditions for this treatment step are 37°C for 30 min followed by 80°C for 15 min. The 10µl primer extension reaction is then set up consisting of 1.5µl treated PCR product, 0.4µl extension primer [5pmol], 1µl SNaPshot buffer (see Appendix I 1.3.2), 1.5µl SNaPshot Ready Reaction Mix (Applied Biosystems, Foster City, CA, USA) and 5.6µl H₂O. The ABI Prism SNaPshot ddNTP Primer Extension Kit (Applied Biosystems, Foster City, CA, USA) provides the SNaPshot Ready Reaction Mix that contains AmpliTaq DNA polymerase and fluorescently labelled ddNTPs. The thermocycler conditions for the extension reaction consist of 50 cycles of 95°C for 5 sec, 43°C for 5 sec and 60°C for 5 sec preceded by a 2 min denaturing step. The extension reaction then undergoes another treatment step, containing only SAP at the same conditions given above (0.5 units SAP and 0.5µl SAP dilution buffer per 5µl DNA) before being analysed on an ABI 377 or 3100 DNA Sequencer using the ABI GENOTYPER (version 2.5) software package.

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

Genotyping of a number of markers (TPH2, an Integrin alpha 4 marker, DVL and EN2) was performed using the SNaPshot technique on an Applied Biosystems 3100 Genetic Analyzer. This is a multi-colour fluorescence-based DNA analysis system using capillary electrophoresis. PCR product was diluted prior to running on the ABI 3100. For each sample, 1µl of the pooled dilutions was mixed with 0.1µl of LIZ120 Size Standard (Applied Biosystems, Foster City, CA, USA) and 9.4µl of Hi-Di® Formamide Solution (Applied Biosystems, Foster City, CA, USA). Genotyping was performed using the ABI 3100 data collection software and Genemapper v.2.7. 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. After 3 failed genotyping attempts, samples were excluded from analysis.

2.2.4.5 TaqMan Allelic Discrimination

Allelic discrimination using the ABI 7900 HT was also performed. This technique is based on the design of two TaqMan probes (Applied Biosystems, Foster City, CA, USA), one specific for the wild-type allele, the other for the mutant allele. Both of these probes are fluorescently tagged (with FAM and VIC). These probes are degraded during the amplification process by the 5' → 3' Exonuclease activity of the polymerase. During this degradation process the fluorescent group becomes separated from the quencher group, thus leading to increased fluorescence. The binding efficiency of the wild-type probe to the mutant allele and vice-versa is very low due to mismatch within the TaqMan probe and target sequence. Consequently mismatched binding is much reduced. All markers genotyped using this technique were not included in the “assays-on-demand” product listing. Therefore all probes

were purchased as an “assay-by-design”. In brief, 3µl of DNA was added to 2.5µl of TaqMan Universal Master Mix (Applied Biosystems, Foster City, CA, USA) and 0.15µl of the specific probe. The samples were then run on the ABI 7900HT using the absolute quantification programme, which was modified to run for 10 min at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 min at 60°C. Samples were analysed using the SDS allele discrimination programme.

2.2.5 Mutation Detection

2.2.5.1 Single Strand Conformational Polymorphism (SSCP) Analysis

SSCP is a gel-based method for mutation detection. The process is based on the principle that homoduplexes will migrate differently to heteroduplexes in a non-denaturing polyacrylamide gel. The PCR products are heated to 95°C for 5 min and then cooled at a rate of 0.5°C per minute to room temperature to allow for the formation of heteroduplexes. 20µl of product is mixed with 2µl of loading gel, and run on a non-denaturing polyacrylamide gel (for 3 hours at room temperature or 7 hours at 4°C). The gels are then stained using a silver staining technique. The gels were fixed by incubation for 3 minutes using a silver staining solution 1 (SSS1) (see Appendix I, section 1.3 for solution composition), while being gently rocked throughout the process to ensure adequate and equal staining, before being discarded (for all silver staining solutions see Appendix I, section 1.3.8). Another volume of SSS1 was added, and the gel incubated for a further 3 minutes before being discarded. Silver staining solution 2 was added and incubated for 15 min. After this time, the gel was quickly rinsed with ddH₂O. Silver staining solution 3 was added to the gel, and incubated for ~ 3 minutes, until the DNA bands become visible. When the DNA bands are distinct enough, SSS3 was discarded, and the gel rinsed twice

with ddH₂O. The gel was fixed for 2 min with silver staining solution 4. Following this the gel was photographed.

2.2.5.2 Denaturing High Performance Liquid Chromatography (dHPLC) Analysis

DHPLC is a highly sensitive and specific method for mutation detection. The process uses ion-pair reverse-phase high-performance liquid chromatography under partially denaturing conditions to differentially retain double stranded heteroduplex and homoduplex molecules. The PCR products are heated to 95°C for 5 min and then cooled at a rate of 0.5°C per minute to a temperature of 40°C to allow for the formation of heteroduplexes. The eluted molecules are detected using a UV detector set at 260 nm. DHPLC was performed on a WAVE DNA Fragment Analysis System (Transgenomic Inc, Omaha, NE, USA) containing a DNASep column. The temperatures used for the analysis of each amplicon were determined using WAVEMAKER software (Transgenomic Inc, Omaha, NE, USA). PCR products were eluted from the column using a linear acetonitrile gradient in a 0.1M triethylamine acetate buffer (TEAA), pH 7 at a constant flow rate of 0.9 ml/min. The gradient was achieved by the mixing of Buffer A (0.1 M TEAA (pH 7), 0.1 mM Na₄EDTA) and Buffer B (25% acetonitrile in 0.1M TEAA (pH 7)). The combining of buffers to create the gradient was done automatically once the column temperature was given to the software. At the end of each analytical run the column was washed with a solution of 50% acetonitrile in 0.1M TEAA (pH 7). Results were automatically recorded in the form of chromatogram images, by the WAVEMAKER software. They were then visually analysed to determine difference between the traces indicative of the presence of a sequence variation. Individual samples with differing traces for the same amplified region were sequenced using both the forward and reverse primers of the original PCR to identify the sequence variation.

2.2.5.3 DNA Sequencing

Sequencing was performed in house, and is a 3 step process, following the initial amplification of the region of interest.

2.2.5.3.1 Post PCR Cleanup

A post PCR cleanup step was performed to remove the original PCR primers and unlabelled dNTPs to ensure they did not interfere with the subsequent sequencing reaction. The cleanup process was performed using the QIAquick PCR Purification Kit (Qiagen Ltd, Crawley, West Sussex, UK).

2.2.5.3.2 Sequencing Reaction

Sequencing of the amplified DNA was performed using the ABI BigDye Terminator (v3.0) Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA). The process is similar in nature to that of a PCR except fluorescently labelled ddNTPs are present in the reaction mix that randomly terminate the amplification process and the mix contains proof reading polymerase. The reaction mix contains 8 μ l Dye Terminator reaction mix, 3.8 μ l PCR product (post cleanup) and 3.2pmol primer. The mix was made up to a final volume of 20 μ l with sterile H₂O. The cycling conditions for the sequencing reaction were 25 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. All sequencing reactions were performed in the forward and reverse directions to confirm results.

2.2.5.3.3 Post Sequencing Reaction Cleanup

The post sequencing reaction cleanup was performed using the DyeEx 2.0 Spin Kit (Qiagen Ltd, Crawley, West Sussex, UK). This process removes the excess unincorporated fluorescently labelled ddNTPs, which would interfere with the analysis of the sequencing products.

2.2.5.3.4 Sequencing Analysis

Sequencing reactions were run on the ABI 3100 DNA Sequencer and the results were visualised and analysed using SeqScape (v.2) and Sequencing Analysis 5.1 (Applied Biosystems, Foster City, CA, USA).

2.2.6 Measurement of DBH Activity

DBH activity was measured by Dr. George Anderson, Yale Child Study Centre, New Haven, Connecticut, USA. Ten microlitres of plasma sample was added to 900 μ l of a reaction mix A and vortexed briefly to ensure adequate mixing in a glass test tube. Samples were then incubated for 1 hour at 37°C with gentle shaking before being placed on ice. One hundred microlitres of 3.4M HClO₄ was added to cause deproteinisation. Samples were spun at 8000rpm for 5 minutes and the supernatant was collected and stored at -70 °C until analysis. The amount of octopamine produced depends on the activity of DBH in the samples. The octopamine was measured following injection of 10-25 μ l of the supernatant on a two-column “backflush” high performance liquid chromatography (HPLC) system. This system comprises of a 5- and 15cm Microsorb 5-mm C18 columns with a connecting 10-port valve. Two pumps delivered the mobile phase (97% pH3.7 1.5% acetic acid with 50mg/L Na₂-EDTA; 3% methanol: 1ml/min). 1 minute after injection, the 10-

port valve was switched on to allow the substrate tyramine to be directed to waste. Octopamine was detected fluorometrically (excitation wavelength = 270nm, emission wavelength = 310nm). Enzyme activity was calculated and reported on an nmol/min/ml plasma basis. The sensitivity of this process allows the determination of activity < 1nmol/min/ml.

2.3 Statistical Tests

2.3.1 General Statistics

2.3.1.1 Hardy Weinberg Equilibrium

All markers were tested to ensure that they conformed to Hardy-Weinberg Equilibrium. The Hardy Weinberg principle assumes that allele frequencies should conform to the following mathematical equation:

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

Where $p^2 = P(AA)$, the probability of being an AA homozygote

$2pq = P(AB)$, the probability of being an AB heterozygote

$q^2 = P(BB)$, the probability of being a BB homozygote

Conformation to this principle can be tested using a χ^2 statistic according to the equation:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where O is the observed genotype frequency

E is the expected genotype frequency

A significant χ^2 value is indicative of deviation from the Hardy Weinberg principle. This may result from genotyping errors or from lack of conforming to the Hardy Weinberg assumptions, which are as follows: 1) an infinite population size, 2) discrete generations, 3) random mating, 4) no selection, 5) no migration, 6) no mutation, 7) equal initial genotype frequencies in the two sexes.

2.3.1.2 Transmission Disequilibrium Testing

The TDT is a test of association in the presence of linkage used in family based studies¹⁸⁷. 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. It is similar to a case-control χ^2 test in the sense that alleles not transmitted act as controls. This lack of a separate control population eliminates the problem of population stratification, which can lead to false positive results in case-control studies.

The TDT is calculated using the McNemar Statistic of

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

Where a = the number of high risk alleles transmitted from heterozygous parents to the affected offspring

b = the number of high-risk alleles not transmitted from heterozygous parents to the affected offspring.

The corresponding p-value of the statistic is calculated using the SymtWin.exe software package. A p value of < 0.05 indicates significant association between the marker being tested and autism.

An extension of the TDT test is the Pedigree Disequilibrium Test (PDT) test, which is used to test for significant deviations of allelic transmissions in general pedigree structures^{269,270}. The PDT gives two statistics, the χ^2_{avg} and χ^2_{sum} . The PDT-avg weights all families equally, whereas PDT-sum gives more weight to families of a larger size. Testing has shown that the PDT-sum test is more powerful than the PDT-avg test, when families of varying sizes are used. PDT-sum is equal to PDT-avg when each family contains the same number of phenotypically informative individuals.

2.3.1.3 Transmission of Haplotypes

2.3.1.3.1 TRANSMIT

TRANSMIT tests for association between genetic marker and disease by examining the transmission of alleles or genotypes from parents to affected offspring²⁷¹. 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/>).

2.3.1.3.2 UNPHASED

UNPHASED is another programme available freely under remote access via the HGMP website (as above)²⁷². It also tests for association between genetic marker and disease by examining the transmission of alleles or genotypes from parents to affected offspring. Unlike the test TRANSMIT, the UNPHASED programme tests the transmission of haplotypes versus the non-transmission of haplotypes, and as

such is more similar to the original TDT test. The programme is based on a likelihood ratio test of a log-linear model²⁷³. Output is given as a Likelihood Ratio Score (LRS) value, which is equivalent to the Extended TDT (ETDT) chi square of Sham and Curtis²⁷⁴ and resulting global p value for the set of haplotypes under consideration.

2.3.2 Other Statistics

2.3.2.1 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 compared with the probability that it is absent. It is measured according to the following formula

$$OR = \frac{a}{b}$$

a = Number of allele or haplotype transmissions by heterozygous parents to affected probands

b = Number of allele or haplotype non-transmissions by heterozygous parents to affected probands

2.3.2.2 Graphical overview of linkage disequilibrium (GOLD)

This is a software package that gives a graphical representation of linkage disequilibrium in genetic data²⁷⁵. D' values (see Section 1.4.1.1 for more details) 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).

2.3.2.3 Haplotype Block Structure Analysis using Haploview

Haploview is a software package still in development at the Whitehead Institute (Massachusetts Institute of Technology) <http://www-genome.wi.mit.edu/personal/jcbarret/haplo/>. Haploview currently allows users to examine block structures, generate haplotypes in these blocks, calculate D' measures of linkage disequilibrium and tag haplotypes with SNPs (htSNPs) that are most informative for the haplotype under consideration. The input files are in an Unphased format (see Appendix III, section 1.2). Output files are graphical in nature. LD measurements between markers are given in D' , and the colour of the block (between two SNPs) indicates the statistical strength of this measurement. If $D' = 1$ and the $LOD > 2$, then this block is bright red in colour. If the D' value = 1 but the LOD value is less than 2 then the block is blue in colour. All other D' values are represented by decreasing shades of red (where the LOD is > 2) or white (where the LOD is < 2). A dark line outlining the block in question indicates the presence of a haploblock.

2.3.2.4 Heritability

Heritability is defined as “the proportion of phenotypic variation in a population attributable to genetic factors”²⁷⁶. The heritability of a trait, e.g. DBH activity, can be calculated using a widely used method, which involves finding the slope of the regression line of the plots of offspring value for a trait graphed against the parental value. A regression slope of 1 indicates that variation in a trait is entirely familial, whereas a regression slope of 0 indicates that variation in a trait is not due to genetic factors. All other values (i.e. $0 < x < 1$) signify that both genetics and environment play

a role in determining the trait. Almost all known traits can be classified as having both influences. In the case of DBH activity, the average parental DBH activity was plotted against the probands DBH activity. The best-fit regression line was plotted and the slope of this line was calculated (using Excel).

2.3.2.5 T-Tests

T-tests are statistical tests that are used to determine if statistical differences occur between two groups of individuals. Independent t-tests are used when the two groups being tested are independent of each other, whereas paired t-tests are used when each individual in one group is related to another individual in the other group. In the presence of equal sample sizes the expression for calculating t is:

$$t = \frac{\bar{x}_{group1} - \bar{x}_{group2}}{(SD / \sqrt{n})}$$

$$SD^2 = \frac{\sum(x - \bar{x})^2}{(n - 1)}$$

Where \bar{x} = mean
x = sample DBH value
n = number of individuals in sample group
SD = standard deviation

These unpaired t-tests are valid only when:

- 1) The samples in the two groups are randomly selected from the larger populations.
- 2) The data is quantitative and forms a normal distribution.
- 3) The variance of the two groups are almost equal

- 4) Each set of observations in the data set is independent of all other observations in the data set.

If the samples are large enough, the two sample groups are almost equal and there are no outliers, then the t-test is still valid in the presence of moderate violations of these assumptions.

2.3.2.6 Kruskal-Wallis Testing

Kruskal-Wallis testing, measured by the H statistic, was performed to test if different alleles in a variant were associated with different DBH activity levels. It is an extension of a two-sample t-test, where more than two sample populations (in the case of SNP there will be three populations, 2 homozygous and 1 heterozygous) can be tested simultaneously, in which non-parametric methods must be used. It was necessary to use non-parametric method as the different allele groups had different variances and sample sizes, and therefore could not be tested using the normal one-way analysis of variance (ANOVA) test. A Kruskal-Wallis test, in comparison to multiple t-tests, reduces the probability of making a type I error, i.e. a false positive where one rejects the null hypothesis when the null hypothesis is true. In the test the null hypothesis is that $\text{mean}_{\text{grp1}} = \text{mean}_{\text{grp2}} = \text{mean}_{\text{grp3}}$. The alternative hypothesis is that either $(\text{mean}_{\text{grp1}} \neq (\text{mean}_{\text{grp2}} = \text{mean}_{\text{grp3}}))$ or $((\text{mean}_{\text{grp1}} = \text{mean}_{\text{grp2}}) \neq \text{mean}_{\text{grp3}})$ or $((\text{mean}_{\text{grp1}} = \text{mean}_{\text{grp3}}) \neq \text{mean}_{\text{grp2}})$ or $(\text{mean}_{\text{grp1}} \neq \text{mean}_{\text{grp2}} \neq \text{mean}_{\text{grp3}})$. Kruskal-Wallis testing was performed using SPSS (version 11).

2.3.2.7 Regression Analysis

Regression analysis was performed to determine the correlation of each marker to the square root of DBH activity, i.e. a marker – phenotype correlation. The method used is based on a strategy reported by Zabetian et al (2003)²⁷⁷. The role of the putative functional polymorphism C-1021T was assessed first by simple linear regression.

Markers were subsequently added in, one marker at a time, and multiple linear regression was used to assess its contribution on DBH activity. Any increase in r^2 values was taken as the contribution of the marker to overall DBH activity.

2.3.2.8 Selection of Tagging SNPs

When testing for associations there are two methods for finding disease-causing variants. The first method, the direct method, is to test all putative causal variants for correlation with the disease in question. The second method, the indirect method, is based on the idea that a set of sequence variants in the genome act as genetic markers that can be used to test association between particular regions of the genome. Using this method, it is not necessary for the markers tested to have any functional consequence. The local region surrounding markers showing association can then be screened for functional mutations, resulting in a more economic and time efficient method for finding disease-causing variants.

The aim of the HapMap project (at <http://www.hapmap.org/>) is to “determine the common patterns of DNA sequence variation in human genome, by characterising sequence variants, their frequencies and correlations between them in DNA samples from populations from parts of Africa, Asia and Europe...thus providing the tools that will allow the indirect association approach to be applied readily to any functional candidate gene in the genome, to any region suggested by family-based linkage analysis, or ultimately to the whole genome for scans for disease risk factors”²⁷⁸. Haplotypes arise because when a single mutation occurs, it is associated with the other alleles surrounding it. This specific set of alleles observed on a chromosomal region is called a haplotype. Within this haplotype, it is only necessary to genotype a small number of SNPs, “tagging SNPs”, in this region to provide enough information to predict the majority of other common SNPs in the same haplotype.

Tagging SNPs were also chosen in the control sample using Haploview. Haploview automatically chooses tagging SNPs on a block-by-block basis. It aims to maximise genotypic information while minimising the number of SNPs genotyped. Tagging SNPs are chosen based on completeness of genotyping and a haplotype tagging SNP (htSNP) set is chosen that defines all haplotypes above a certain frequency threshold (default = 1%). The resulting htSNPs are not necessarily chosen to be the most parsimonious for the sample in question, but instead to provide the most efficient and thorough testing when moving onto a larger sample, i.e. there is a greater chance of catching variation in the new, larger sample in comparison to the initial test sample.

CHAPTER 3

THE SEROTONERGIC SYSTEM AND AUTISM

3.1 Introduction

3.1.1 The Serotonin System

The production of serotonin, also known as 5-hydroxytryptamine (5-HT) results from the hydroxylation of tryptophan by the enzyme tryptophan hydroxylase (TPH) (see Figure 3.1). This step, which is the rate-limiting step in the synthesis of serotonin, leads to the formation of 5-hydroxytryptophan (5-HTP). The decarboxylation of this product results in the formation of serotonin. Serotonin can then be released from a synaptic vesicle across a nerve synapse, whereupon it stimulates a reaction in the post-synaptic nerve, before being degraded or taken back up into the presynaptic cell by the serotonin transporter (5-HTT). After re-uptake, serotonin can be recycled by storage in another synaptic vesicle or it can be broken down by the enzyme monoamine oxidase (MAO) leading to the formation of 5-hydroxyindoleacetic acid, which is excreted by the body with urine (see Figure 3.1 for a summary of the serotonin pathway).

The serotonin producing-nerve cells are found in 9 clusters in the brain that extend into the raphe nuclei of the midbrain, pons and medulla. The projections are also sent to the cerebral cortex by the dorsal and medial raphe nuclei ²⁷⁹. This wide projection of serotonergic neurons, especially in the limbic regions of the brain that are critical for emotional expression and social behaviour, leads to numerous potential functions for the neurotransmitter. Serotonin plays a key role in behavioural inhibition, appetite, aggression, mood, social affiliation, sleep and sensory gating ^{75,280-284}. This has led to the hypothesis that abnormalities in the serotonergic systems might contribute to autism.

3.1.2 Serotonin and Autism

It is not just the disruption of serotonin related behaviours in autism that supports the serotonergic hypothesis. One of the most widely replicated biological findings relating to autism is the occurrence of hyperserotonemia in individuals with autism. Hyperserotonemia is defined as serotonin levels that are greater than 2 standard deviations above the mean serotonin level²⁸⁵. A report by Schain and Freedman was the first to show an increase in endogenous serotonin levels in the blood of children with autism²⁸⁶. This finding has since been replicated many times. A recent study by Mulder et al reported significantly increased serotonin levels in individuals with autism and PDD compared to a control sample and a sample of individuals with mental retardation but without PDD²⁸⁷. A study by Coutinho et al described a non-significant increase in platelet serotonin levels in 105 individuals with autism compared to 52 controls². Approximately 11% of these affected children were also classified as hyperserotonemic. Most studies have reported increases of between 25% and 50% in platelet serotonin in individuals with autism^{288 289 290 291}. Variation in the magnitude of the increases may in part be explained by age²⁹², although this was not the case in the Coutinho study², and in part by the different ethnicity of the samples involved in the studies²⁹². The greatest increases compared to controls occur in children before the onset of puberty. Increased serum serotonin levels have also been reported²⁹³. Hyperserotonemia also occurs in the immediate family of these individuals with autism. An initial study by Cook et al found that the presence of hyperserotonemia in one individual with autism often co-occurred with hyperserotonemia in a close family member^{294,295}. Lenventhal et al calculated that if an individual suffered autism, then their first degree relatives were ~2.4 times more likely to be hyperserotonemic in comparison to controls. However, individuals with autism were still more hyperserotonemic in comparison to the other family members²⁹⁵.

Serotonin receptor 2A is one of 14 receptors for serotonin²⁹⁶. Decreases in the density of platelet 5-HT₂ receptor binding sites, labelled with the radio-ligand [³H]-

lysergic acid diethylamide, was reported in a sample of individuals with autism²⁹⁷. However, within the sample of individuals with hyperserotonemia who had both serotonin uptake and 5-HT₂ binding measured (n = 8), there were two subgroups. Firstly, there were 4 individuals with increased serotonin uptake, while 3 individuals had decreased 5-HT₂ binding. Data was not available for the final individual. Cook et al, therefore suggested that within any given sample of hyperserotonemic individuals with autism, there may be intergenic heterogeneity, with one subgroup having increased serotonin uptake, while another subgroup has decreased 5-HT₂ binding²⁹⁷.

Not all individuals diagnosed with autism exhibit elevated serotonin levels. Indeed there has been two reports of decreased serotonin levels in autism samples compared to control samples^{298 299}. Hyperserotonemia has also been associated with other medical and psychiatric conditions including schizophrenia, unipolar depression, ADHD and mental retardation³⁰⁰⁻³⁰². The association with mental retardation could be a possible confounding factor. Serotonin levels are most increased in individuals with severe mental retardation, followed by children with autism and mental retardation, and relatively normal in high functioning individuals without retardation³⁰³. In the first study, by Schain and Freedman, all the individuals with autism who's blood serotonin levels were measured suffered intellectual disability in addition to autism²⁸⁶. Hyperserotonemia may thus be a marker of intellectual disability in general rather than autism specifically. Ethnicity of samples may also be another confounding factor. A study by McBride et al reported that Caucasian children have lower serotonin levels compared to African / African-American or Hispanic children, regardless of age²⁹².

It should be noted that all of the above studies have been undertaken in the peripheral system. Whether the level of serotonin in the central nervous system is also increased in individuals with autism is still unknown. In order to address this particular question it is necessary to analyse the levels of 5-HIAA in cerebrospinal fluid (CSF), as it is the major metabolite of serotonin in the brain. Due to the

strength of the blood – brain barrier, the 5-HIAA levels from the CSF are free from contamination of peripheral serotonin and 5-HIAA. Eight studies (summarised in Lam et al ³⁰⁴) have undertaken this experiment ^{305 306 307 308 309 310 311 312}. In all but one of these studies there were no differences found in 5-HIAA levels in the sample with autism in comparison to the control samples. One study, by Cohen et al reported a significant decrease in the levels of 5-HIAA in CSF from 10 patients with autism compared to 10 age-matched non-autistic “psychotic” controls. It is important to point out that in all of these 5-HIAA studies, the number of individuals in each study is relatively small ($8 < n < 25$), and therefore the power may not be adequate to detect differences between samples.

Evidence for the potential role of serotonin in autism also comes from pharmacological observations. Fluvoxamine (a selective serotonin reuptake inhibitor, SSRI) is a drug that targets the serotonin transporter, preventing the reabsorption of serotonin back into the presynaptic nerve cell. It is widely used in reducing the restrictive and repetitive behavioural phenotypes associated with autism ³¹³. Irritable and stereotypical/compulsive behaviour can also be reduced in some individuals with autism with the administration of risperidone, which is an antagonist of the 5-HT_{2A} receptor ¹⁷³. Deterioration of autistic behaviours has also been reported following administration of the serotonin releaser, fenfluramine, and the 5-HT_{1B/D} receptor agonist, sumatriptan ^{314 315}.

In recent years positron emission tomography (PET) scans have been used to trace and measure serotonin synthesis in the brain. A study by Chugani et al measured the synthesis of serotonin in 8 children diagnosed with autism and 5 of their siblings ³¹⁶. Significantly decreased serotonin synthesis was observed in the thalamus and frontal cortex region, and increased serotonin synthesis in the contralateral dentate nucleus of the cerebellum regions of the children with autism. Also in 5 out of 7 of these children, decreased accumulation of serotonin in the left frontal cortex and thalamus, with an increased accumulation in the right dentate nucleus was observed. In another 2 cases, the mirror image result was reported, i.e. decreased accumulation in the right

frontal cortex and thalamus with increased accumulation in the left dentate nucleus. The reason for this mirror image result is not known. However, the observation of increased synthesis of serotonin in some parts of the brain with decreased synthesis elsewhere may help explain why there has only been one report of a significant difference between 5-HIAA levels in affected and control samples. Chugani et al looked at the relationship between serotonin accumulation and synthesis in autism and age in three samples; children with autism, their siblings, and children with epilepsy³¹⁷. In the children without autism, the level of serotonin synthesis was at its greatest between the ages of 2 and 5 years of age. Synthesis then continued to decrease between the ages of 5 and 14. In the sample of children with autism, this normal decline was not observed. Thus the development of the autism phenotype may be related to this dysregulation of serotonin synthesis.

Mouse knockout models of serotonin pathway genes have suggested possible links to autism. Deletion of the 5HT_{1A} receptor leads to increased anxiety behaviour, reported as between a three-fold (in males) or two-fold (in female) avoidance of the centre of the open field apparatus, in 5HT_{1A} knockout mice^{318 319}. These mice also had an increased response to stress, as elicited by increased mobility in the forced swim test, which may reflect the involvement of increased anxiety and emotional reactivity when mutant mice are exposed to inescapable stress. A common feature of autism is the inherent dislike and increased stress when an affected individual is placed in a novel situation. Interestingly, the mice can be rescued by expression of 5HT_{1A} in the forebrain only, indicating that 5HT_{1A} expression in the raphe nuclei is not responsible for the anxiety like behaviour. The expression of 5HT_{1A} is important during the early postnatal period as the rescuing of mice later in life does not reduce anxiety like behaviour. This may be because of the role that serotonin plays in brain development. During early development, the correct migration and differentiation of cells is critical for normal development. Serotonin has been shown to regulate mouse cranial neural crest migration and the differentiation of glutamate neurons in the developing cerebral cortex^{320 321}. With numerous links between the serotonin system and autism, association studies were the next logical test to evaluate whether

functional genetic variants in serotonergic genes have a role in the development of autism.

3.1.3 Genetic Studies of the Serotonin Related Genes And Autism

A number of studies have been undertaken to test the possible association of genes in the serotonin system with autism. Several genetic associations have been found between the serotonin transporter gene variants (*5-HTT*) and autism (see Section 3.1.4) and more recently with the tryptophan hydroxylase 2 gene (*TPH2*) (see Section 3.1.5). There have also been a number of studies looking at the role of the Monamine Oxidase A gene (*MAO-A*). Yirmiya et al examined the transmission of a 30bp repeat located 1.2kb upstream of the *MAO-A* gene. This marker occurs as a 3, 3^{1/2}, 4 or 5 repeat, and the longer repeats are associated with increased transcriptional efficiency³²². Although there was no evidence of over-transmission of any allele to affected probands, there was an association with reduced IQ scores and the 4 repeat allele. However, association between the 3 repeat allele and lower IQ was also significant in a different sample of young males diagnosed with autism²²⁷. The same marker was also studied in a Canadian multiplex sample³²³. This time however there was no difference in IQ associated with the child's *MAO-A* promoter genotype. However, a lower proband IQ was associated with a maternal 3 repeat allele (p value = 0.03 when analysing multiplex and simplex families, p value = 0.01 when analysing only multiplex families). These associations are interesting and warrant further investigation. Biochemical studies of MAO-A activity in autism have shown contradicting results. Filinger et al reported a lower platelet MAO-A activity in a small population of children suffering autism and pervasive developmental disorder³²⁴. Takahashi et al and Cohen et al found no differences between MAO-A activity in children diagnosed with autism compared to an age-matched control samples^{325 326}.

The role of the serotonin receptor 2A (HTR2A) in the development of autism has also been investigated. A study by Veenstra-VanderWeele analysed the transmission of 2 functional SNPs (one in the promoter which leads to decreased expression of the gene, and another in exon 3 which shows slower and damped Ca^{2+} response to serotonin) and a mutation in intron 1³²⁷. There were no reported transmission distortions when both parents were included in the analysis and when only the transmission of paternal alleles was considered, a trend of increased transmission of the A allele of the promoter variant. However following corrections for multiple testing and parent-of-origin testing, this result was considered to be non-significant. Another serotonin reporter, *HTR7*, has also been tested for association with autism, but likewise has shown no evidence for transmission distortion to affected probands³²⁸.

Tryptophan 2,3-dioxygenase (TDO2) is an enzyme that catalyses the oxidative catabolism of tryptophan to kynurenine³²⁹. It has shown genetic association with other childhood psychiatric disorders including Tourette Syndrome and Attention Deficit Hyperactivity Disorder^{330 331}. An association study in a AGRE sample consisting of 196 multiplex autism families has also shown an association with this gene²³⁵. An A to C transversion in the promoter region of the gene showed significant over-transmission of the C allele to affected probands, leading to an Odds Ratio (OR) of 3.86 (95% CI = 1.9 to 13.3) and a p value of 0.0006. There have been no replications to date.

3.1.4 The Serotonin Transporter Gene

The serotonin transporter (5HTT) is a major modulator of serotonergic neurotransmission. It directly controls the re-uptake of serotonin in presynaptic nerves. The serotonin transporter gene is located on chromosome 17q11.2 and contains 14 exons spread over ~35kb³³². Transcription of the gene produces a

membrane spanning protein of 630 amino acids of 70.3kDa (Genecard information: <http://genecards.bcgsc.bc.ca/>).

There have also been reports of linkage using markers close to this and within this gene. Two genomic screens have reported suggestive evidence with MLS. IMGSAC found a single point MLS of 3.6 using a VNTR in intron 2, and Yonan et al reported suggestive evidence of linkage using a marker (D17S1294) 142kb away from the gene with an MLS of 4.3 (p value = 0.0029)^{85,90}. Sex stratification using families with either male only ASPs or female containing ASPs, led to increased linkage in families with male only ASPs^{93,94}.

There are three extensively studied markers in this gene. They include:

- 1) A 44bp insertion / deletion in the promoter, and the basal activity of the long promoter allele has been shown to be three fold higher than that of the short promoter allele³³³
- 2) A variable tandem repeat (VNTR) in intron 2, containing 9, 10 or 12 repeats, in which the 12 repeat allele has been shown to drive higher expression in embryonic mouse rostral hindbrain³³⁴
- 3) A G to T transversion in a putative polyadenylation site in the 3' UTR of the gene.

Klauck et al and Yirmiya et al reported preferential transmission of the long promoter variant in autism^{4 213}. Tordjman et al reported the transmission of the short promoter allele in severely affected individuals, but transmission of the long promoter in the sample as a whole²¹². Cook et al, who reported the initial finding, found over-transmission of the short promoter allele¹⁹⁰. There have also been reports of absence of association^{5,335 1}.

A study by Coutinho et al, reported that increased serotonin levels were associated with the long serotonin promoter variant, and the 10 repeat VNTR². QTDT indicated a significant dominant variance effect of the long promoter allele on

serotonin levels, whereas the VNTR showed a significant additive genetic effect. No distortions in allelic transmissions were found with either variant however.

3.1.5 The Tryptophan Hydroxylase 2 (TPH2) Gene

The conversion of tryptophan to serotonin by tryptophan hydroxylase (TPH), in the presence of oxygen, tetrahydrobiopter (BH4) and ferrous iron, is the rate-limiting step of the serotonin pathway (see Figure 3.1). There are two forms of the TPH protein, resulting from the transcription of two separate TPH genes. The *TPH1* gene, located on 11p15.2, is expressed in the pineal gland, spleen, gut and thymus, whereas the *TPH2* gene on chromosome 12q21.1 is only expressed in the brain³³⁶. The TPH2 gene is 93.6kb in length, with 11 exons. Transcription of the gene produces a protein of 490 amino acids in length that is 56kDa (<http://bioinfo1.weizmann.ac.il/cgi-bin/genecards/carddisp?TPH2>). This protein shows a very high similarity (71%) to the protein produced by the *TPH1* gene. The discovery of the second *TPH* gene with brain expression may explain the lack of association found in earlier studies, which focused on *TPH1*²²⁵.

Studies in mice illustrated the importance of amino acids 106 and 116 in TPH2. Deletion of amino acids residues 1 to 106 has no effect on enzyme activity, however deletion of residues 1 to 116 abolishes enzyme activity³³⁷. In a mouse strain with a C1473G transversion leading to a Pro⁴⁴⁷ to Arg⁴⁴⁷ change, those mice homozygous for the G allele (arginine) showed a 50% reduction in serotonin synthesis in the frontal cortex and a 70% reduction in the striatum³³⁸. In humans, a mutation at Arg⁴⁴¹ (mutant amino acid = His⁴⁴¹), resulting from a G to A transition at position 1463, has also been described. In human PC12 cells, replacement of the arginine with histidine results in an 80% reduction in serotonin production³³⁹. This mutation is uncommon in the general population, with a frequency of ~1.4% (3 out of 219 individuals in a control population), but has shown to be associated in a sample of

patients suffering unipolar depression (rare allele found in 9 of 87 cases, i.e. 10.4%)
339 .

Two other SNPs in *TPH2* have been shown to be associated with autism ⁷. A T to G transversion in intron 1, rs4341581, is a rare SNP. In the control sample studied by Coon et al, the G allele only appeared once in 95 individuals, but 9 times in a sample of 88 individuals with autism ($\chi^2 = 9.58$, p value = 0.013). Another SNP, an A-T transversion (rs11179000), was also more frequent in the sample with autism, occurring at a frequency of 0.276 compared to a frequency of 0.202 in the control sample ($\chi^2 = 7.34$, p value = 0.024). It is unknown whether these SNPs are functional or whether they are in LD with a functional SNP elsewhere in the gene.

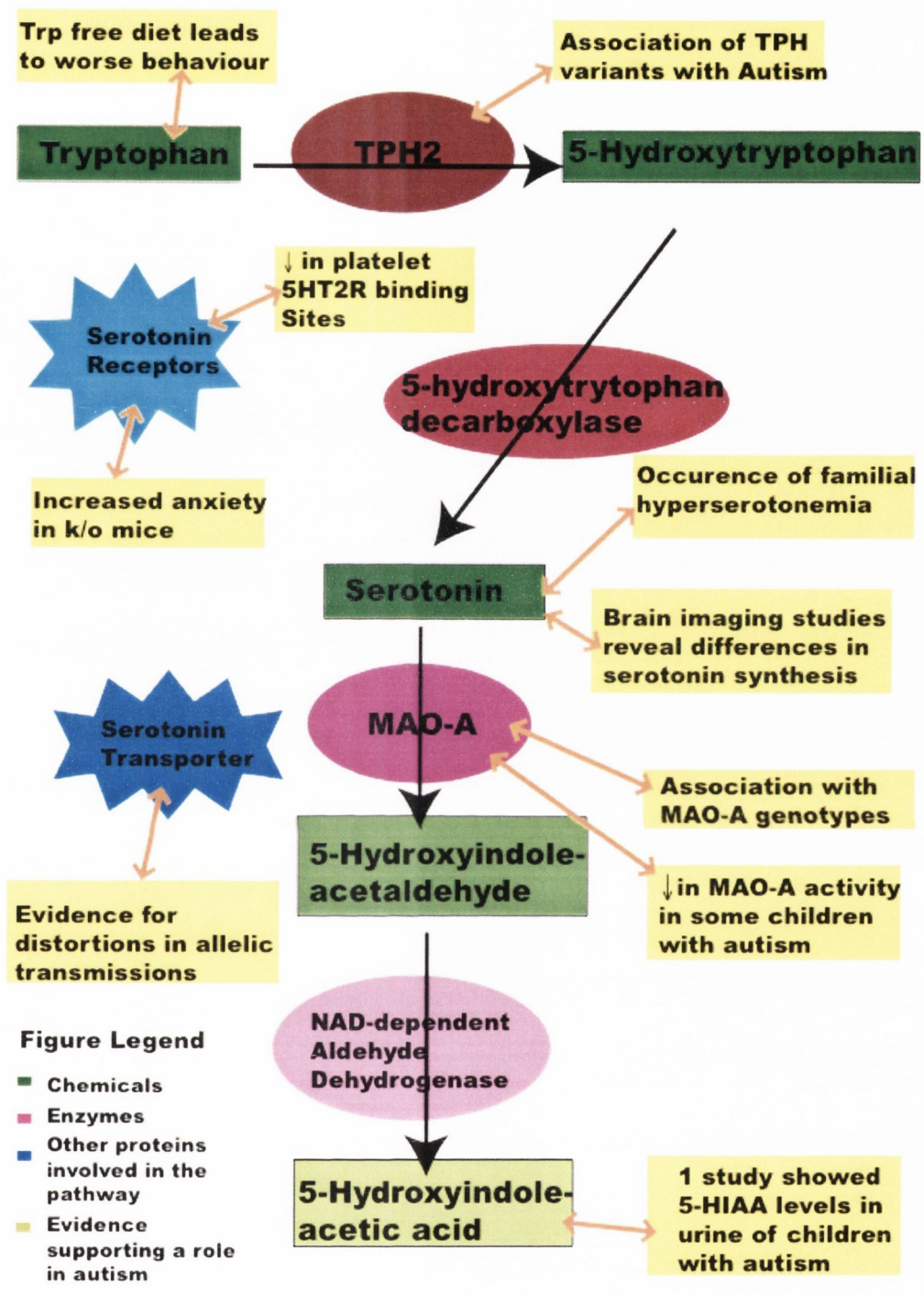


Figure 3.1 SUMMARY OF SEROTONIN PATHWAY AND LINKS TO AUTISM.

3.1.6 Aims

The aim of this study was to investigate the role of the serotonin transporter and the tryptophan hydroxylase 2 genes in the development of autism. Replication of previous reported findings is an important element in the study of complex diseases.

3.2 Materials and Methods

3.2.1 Samples

The serotonin transporter study was undertaken in the Phase I sample, whereas the TPH2 study was undertaken in the Phase II sample (see Section 2.1 for more details on these samples).

3.2.2 Amplification of Serotonin Transporter Variants

Five variants were screened in our sample. They included the serotonin in/del in the promoter region, a T to C transition in the promoter 1B (rs2020936) (referred to here as SNP10), the VNTR in intron 2, a G to A transition in intron 2 (rs2020942) (referred here to as SNP18), and the G to T transversion occurring in the putative polyadenylation signal in the 3'UTR. Details of amplification and genotyping methods can be found in the appendix (Appendix II, sections 1.1 and 1.2).

3.2.3 Amplification of TPH2 Variants

Four *TPH2* variants were investigated in our sample. They included the G to T transversion in intron 1, rs4341581, and an A-T transversion in intron 4, rs11179000, both of which showed association in an AGRE sample diagnosed with autism ⁷. Another two SNPs, a T to G transversion in intron 5 (rs1843809), and a G to T transversion 6.705kb from the first exon (rs4570625), were also investigated. These two SNPs showed association in an Irish sample suffering ADHD ³⁴⁰. Markers rs1843809 and rs11179000 were genotyped using an “assay-by-design” kit from Applied Biosystems (see Section 2.2.4.5 for details). The remaining two SNPs were

amplified and analysed using the SNaPshot primer extension technique (Applied Biosystems, Foster City, CA, USA), see Appendix II, sections 1.1 and 1.2, for more details.

3.2.4 Statistics

All variants were tested for Hardy-Weinberg equilibrium. LD was measured using the D' statistic by GOLD (in the case of the serotonin transporter) and Haploview (in the case of TPH2), see Section 2.3.2.2 and 2.3.2.3 for more information.

Transmission of alleles was tested using the TDT. In the case of the serotonin transporter, transmission of haplotypes was analysed using the TRANSMIT programme, whereas in the case of the TPH2 variants, transmission of haplotypes was tested using the UNPHASED programme (see Sections 2.3.1.3 for further details).

The power of the samples was calculated using the Genetic Power Calculator (available at <http://statgen.iop.kcl.ac.uk/gpc/dtdt.html>)³⁴¹. For ease of comparison it is assumed that the variants are in absolute LD with the disease causing variant, i.e. $LD = 1$. Power is calculated at a false positive rate of 5%. In the running of this programme, genotype relative risks (GRRs) for homozygotes for the high-risk allele and for heterozygotes, had to be estimated. They were estimated at three points. The first estimation was for the highest OR detected from the tests. Another two GRRs for high-risk homozygotes were estimated based on a multiplicative model, whereby high-risk homozygotes have a higher GRR than heterozygotes, and the overall OR is estimated to be that detected from the tests.

A joint-analysis of serotonin promoter markers was performed. All studies that were published since 1997¹⁹⁰ to 2004², with published allelic transmissions in

trio/multiplex families were included. Transmissions of the short and long promoter alleles were compiled and the sum of transmissions calculated. A TDT test was performed on this “summed” value. All individual studies were also plotted on a funnel plot to test for publication biases.

3.3 Results

3.3.1 The Serotonin Transporter Gene

Genotype frequencies did not deviate from Hardy-Weinberg equilibrium.

3.3.1.1 LD Measures

Linkage disequilibrium measurement demonstrates the presence of LD between markers across the gene, with D' values between 0.301 and 0.788 (see Figure 3.2). The promoter was not in strong LD with any marker, however LD was seen to increase with markers ranging from SNP 10 to the 3'UTR.

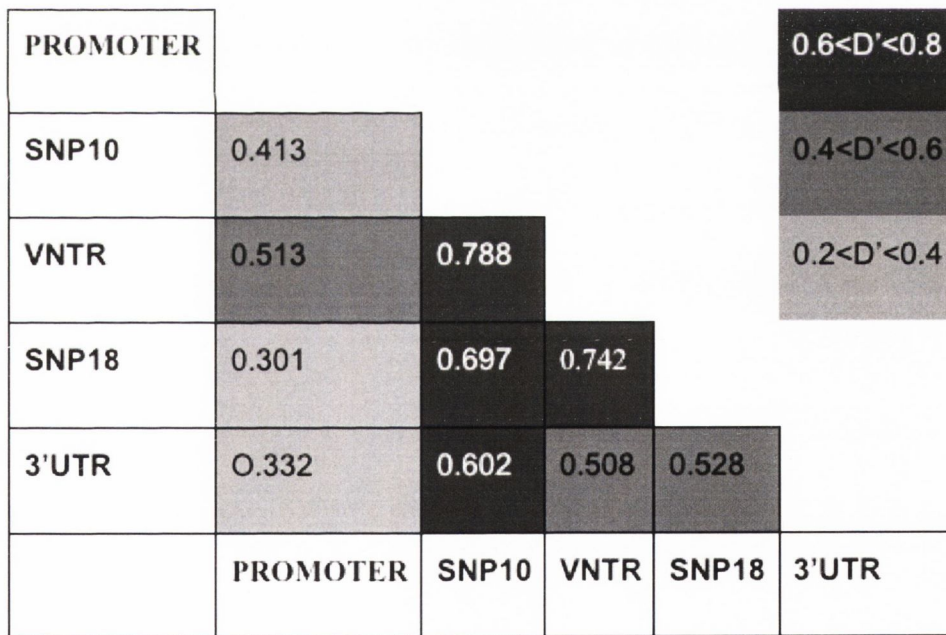


Figure 3.2 MEASURES OF LD (D') ACROSS THE SEROTONIN TRANSPORTER GENE.

D' was measured using GOLD.

3.3.1.2 TDT Testing

Due to the low frequency of 9 repeat alleles in the VNTR markers (frequency = 1.7%), the 9 repeat alleles were combined with 10 repeat alleles for testing. There was significant over-transmission of the short promoter allele to probands. With 55 transmissions of the short promoter allele and 34 non-transmissions, a χ^2 value of 4.5252, OR of 1.6176 and corresponding p value of 0.0334 was produced. No other significant deviations from expected equal allele transmissions were observed (see Table 3.1).

Marker	Allele	Transmitted	Not Transmitted	OR	χ^2	p value
Promoter Variant	Long	34	55	1.62	4.525	0.0334
	Short	55	34			
SNP10	C	54	42	1.29	1.261	0.2615
	T	42	54			
VNTR	12 repeat	45	37	1.22	1.000	0.3347
	9+10 repeat	37	45			
SNP18	G	46	37	1.24	0.771	0.3800
	A	37	46			
3'UTR	G	42	31	1.35	1.371	0.2416
	T	31	42			

Table 3.1 TDT RESULTS FOR SEROTONIN TRANSPORTER MARKERS. Significant transmissions are highlighted in yellow.

3.3.1.3 Transmission of Haplotypes

Examination of two marker haplotypes revealed increased transmission of several haplotypes. These include the haplotypes constructed from the short promoter allele and the C allele of SNP10 ($\chi^2 = 4.2277$, p value = 0.0398), and also the haplotype constructed from the C allele of SNP10 and the 12 repeat allele of the VNTR ($\chi^2 = 5.8912$, p value = 0.0152). In addition, when the promoter, SNP10 and VNTR markers were analysed together a significant preferential transmission of the short promoter – C allele (SNP10) – 12 repeat haplotype to autism cases was observed ($\chi^2 = 6.341$, p value = 0.0118). Transmission of a Prom-SNP10-VNTR-SNP18haplotype (short-C-12rp-G-G) produced the strongest OR (OR = 2.403, p value = 0.0085). Interestingly, almost all associated haplotypes, with the exception of haplotype (VNTR-SNP18-3'UTR, p value = 0.0355) contained the C allele of SNP10. ORs were also calculated for the haplotypes and ranged from 1.4512 (SNP18-3'UTR) to 2.24 (Promoter-SNP10-VNTR-SNP18), see Table 3.2 for further details.

It was found that ORs increased as neighbouring polymorphisms were included in the haplotype analysis. A good example of this is the 12 repeat VNTR, in which the OR of developing autism from inheriting this allele rises from ~1.35 to 1.70 when the C allele of SNP10 is included, and continues to rise until all polymorphisms, bar the G allele of the 3'UTR, are included, giving rise to the highest OR of 2.4. Interestingly, ORs decrease as one moves closer to the 3' end of the gene (see Table 3.1 and figure 3.3).

Marker	Alleles	Observed Transmission	Expected Transmission	Chi-square	Haplotype p-value	OR	Global p value
Prom-SNP10	S-C	40.05	32.91	4.227	0.0398	1.741	0.0784
SNP10-VNTR	C-12rp	56.68	47.93	5.891	0.0152	1.692	0.0776
VNTR-SNP18	12rp-G	94.08	86.601	3.0407	0.0812	1.4203	0.1964
SNP18-3'UTR	G-G	68.804	61.566	3.0933	0.0786	1.4512	0.3398
Promoter-SNP10-VNTR	S-C-12rp	30.951	23.819	6.341	0.0118	2.0436	0.0821
SNP10-VNTR-SNP18	C-12rp-G	50.531	41.129	7.6501	0.0057	1.8406	0.2323
VNTR-SNP18-3'UTR	12rp-G-G	68.143	59.578	4.4189	0.0355	1.6088	0.1901
Promoter-SNP10-VNTR-SNP18	S-C-12rp-G	29.163	21.869	7.6866	0.0085	2.4034	0.1121
SNP10-VNTR-SNP18-3'UTR	C-12rp-G-G	42.177	32.946	10.888	0.0052	2.062	0.565
Promoter-SNP10-VNTR-SNP18-3'UTR	S-C-12rp-G-G	24.981	18.717	5.878	0.0153	2.206	0.1766

Table 3.2 HAPLOTYPE TRANSMISSION AND ODDS RATIO DATA. Haplotype p-values represent the p-value for the transmission of indicated variants of the haplotype. The global p-value represents the overall significance for the haplotype using all possible variants. (S = Short promoter variant, 12rp = 12 repeat variant of intron 2 VNTR).

Odds Ratios

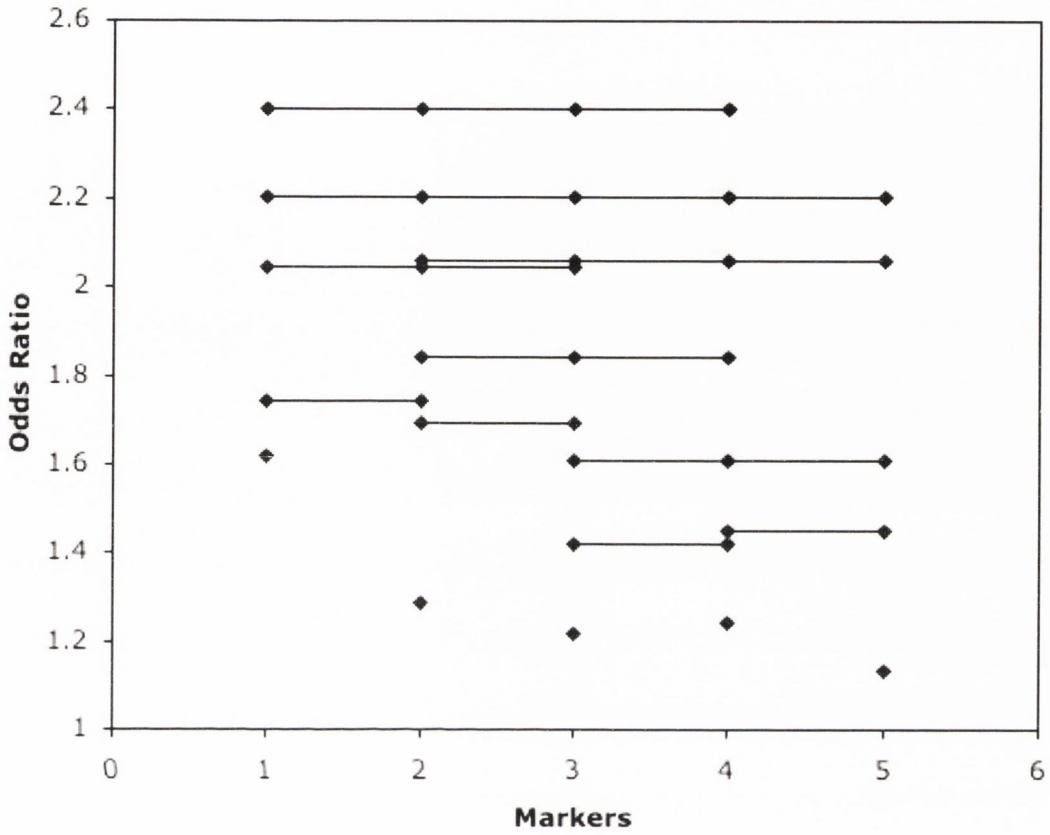


Figure 3.3 THE ODDS RATIO OF EACH HAPLOTYPE.

- | | | | |
|-----------|---------------------------|-----------|-------------------|
| Marker 1: | The short Promoter Allele | Marker 2: | C allele of SNP10 |
| Marker 3: | 12 repeat VNTR | Marker 4: | G allele of SNP18 |
| Marker 5: | G allele of 3'UTR | | |

3.3.1.4 Power Study

The power of this replication was calculated using the TDT power calculator (<http://statgen.iop.kcl.ac.uk/gpc/dtdt.html>). Given a sample size of 96 trios, a disease prevalence of 0.003, and a GRR for inheritance of two high-risk alleles of 2.4 (the

prom-SNP10-VNTR-SNP18 haplotype), and a GRR of 1.6 for heterozygous individuals, the power of this replication was calculated to be 66.93% to detect an association at p value < 0.05 . All other things being equal, and given a GRR of 4 for high-risk homozygotes and 2 for heterozygotes, the power of this test was 75%. Given a GRR of 3.2 for high-risk homozygotes and 1.6 for heterozygotes, resulted in power of 66%.

3.3.1.5 Joint-analysis of published serotonin transporter promoter variants.

Following a full literature review, and a PubMed search using the terms “autism”, “serotonin transporter” and “promoter”, 43 references were found. Of these, 11 contained TDT information regarding the transmission of long and short promoter alleles (see Table 3.3 for more details). ORs for the resulting transmissions (based on over-transmission of the short promoter allele) ranged from 1.65¹⁹⁰ to 0.375²¹², i.e. over-transmission of the long promoter allele. In the combined sample there were 730 transmissions of the short promoter allele versus 689 transmissions of the long promoter allele. This resulted in an OR of 1.059, and corresponding p value of 0.288.

Bias in the meta analysis was tested for using the funnel plot³⁴², which tests for publication and participation bias. The funnel plot is constructed by plotting the size of each sample against the odds ratio (OR) for the data. The theory behind the method is that the larger samples will have ORs closer to the combined OR, which, in the absence of heterogeneity between the samples (see 2.3.5.1), should also be the true OR of the combined sample. The smaller samples scatter themselves either side of the true OR value and when plotted they visually construct the shape of an inverted funnel. The presence of bias in the sample would result the plot deviating from the inverted funnel shape. For example if the study excluded all of the negative findings then one side of the funnel diagram would be missing. In the case of the

serotonin long/short promoter variant, the resulting funnel plot suggests that such biases are not apparent. However there are less positive findings in smaller samples than would be expected. This may be in part due to the small number of samples in the analysis.

Author	Transmission Short Allele	Transmission Long Allele	OR	p value	Origin of Sample	Number in Study
Tordjman ²¹²	24	64	0.375	0.046	French	71
Yirmiya ²¹³	11	25	0.44	0.027	Israeli	35
Klauck ⁴	43	68	0.632	0.032	German	65
Maestrini ⁵	72	76	0.947	0.805	Caucasian (European & American)	98
Mulder ²¹¹	54	51	1.059	0.77	Dutch	108
McCauley ⁶	221	180	1.228	0.01	American	137
Conroy (data in thesis) ²¹⁰	55	34	1.617	0.0334	Irish	86
Betancur ¹	71	55	1.291	<0.15	Caucasian (European & American)	93
Cook ¹⁹⁰	48	29	1.655	0.03	American	86
Coutinho ²	88	79	1.114	0.536	Portuguese	196
Kim ³	43	30	1.433	0.128	American	115
Total	730	689	1.059	0.288		1090

TABLE 3.3 INDIVIDUAL STUDY RESULTS FOR TRANSMISSION OF SEROTONIN TRANSPORTER VARIANT.

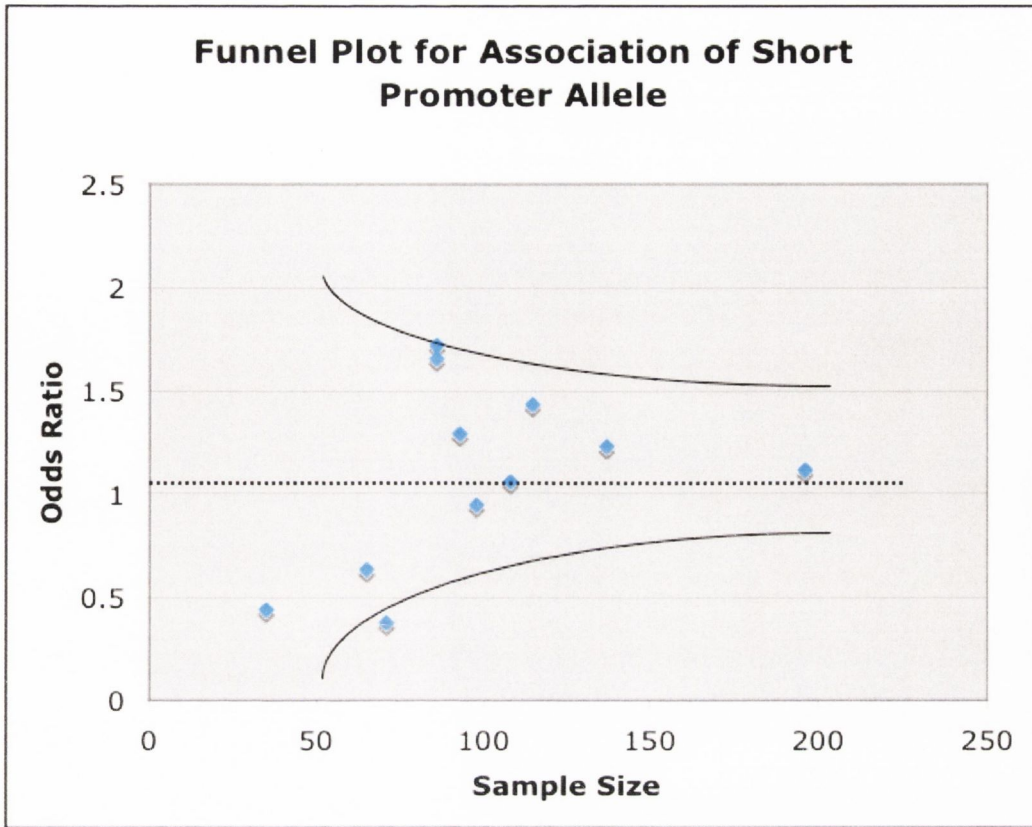


Figure 3.4 FUNNEL PLOT. The dotted line indicates the OR for the combined samples (OR = 1.059). Blue points indicate the OR of each individual sample (see Table 3.3 for further information). Funnel lines (in black) highlight the tendency of ORs of larger samples to be closer to the true mean OR (in dotted line) in comparison to ORs of smaller samples.

3.3.2 TPH2

3.3.2.1 LD Measures

LD measures (see Figure 3.5) show a region of moderate LD between markers rs4341581 (in intron 1) and rs1843809 (in intron 5). The SNPs in intron 1 and intron 4, rs4341581 and rs1117900 respectively, were in complete LD although the accompanying LOD score was not greater than 2.0 (blue colour). The marker 703bp 5' to the gene (rs4570625) was not in LD with any of the other markers (as can be seen in figure 3.5).

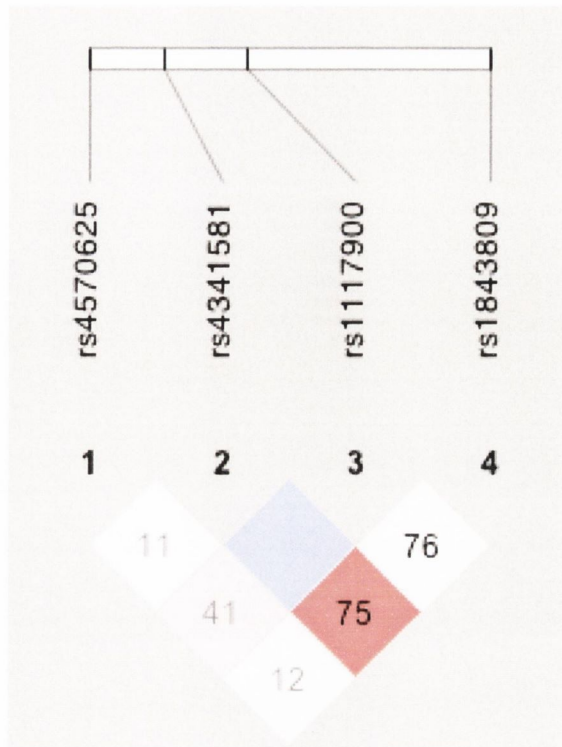


Figure 3.5 MEASURES OF LD (D') ACROSS THE *TPH2* GENE. LD was measured using the Haploview Programme. A blue block indicates complete LD (with LOD scores < 2), whereas red/pink and white blocks indicate D' values < 1.0, with LOD scores >2 and < 2 respectively.

3.3.2.2 TDT Testing

There was significant (at $p < 0.05$ level) over-transmission of the A allele to probands with autism (OR = 1.60, p value = 0.0406). There were no other significant deviations from expected allele transmission to affected offspring with the remaining three markers (see Table 3.4 for more details).

ID Number	Marker Name	Allele	Transmitted	Not Transmitted	OR	χ^2	p value
1	rs4570625	G	62	48	1.29	1.787	0.1813
2	rs431581	A	14	10	1.40	0.670	0.4131
3	rs1117900	A	48	30	1.60	4.192	0.0406
4	rs1843809	G	19	13	1.46	1.461	0.2874

Table 3.4 TDT RESULTS FOR TPH2 MARKERS. Significant transmissions are highlighted in yellow and bold.

3.3.2.3 Transmission of Haplotypes

Haplotype transmissions were tested using the UNPHASED programme (see Section 2.3.1.3.2). There were no significant haplotype transmissions to affected offspring (see Table 3.5). However there was a trend towards over-transmission of the A-T haplotype of markers rs431581 and rs1117900 (OR = 1.74, p value = 0.092).

Markers	Alleles	Transmitted	Not Transmitted	OR	LRS	p value
1-2	G-A	52	37	1.41	3.83	0.2803
2-3	A-T	40	23	1.74	4.767	0.0922
3-4	T-G	27	15	1.83	3.336	0.1886

Table 3.5 HAPLOTYPE TRANSMISSION RESULTS FOR THE TPH2 GENE. Transmission of haplotypes was analysed using the UNPHASED programme. Trends towards significance are highlighted in light yellow.

3.3.2.4 Power Study

The power of a study to detect a finding such as that of the association of the A allele of rs1117900, with a GRR for homozygotes of 1.6, and a GRR of 1.2 for heterozygotes in a sample of 184 trios, was found to be 36%. Given a GRR of 2.2 for high-risk homozygotes and 1.4 for heterozygotes, the power was calculated to be 56%. Finally given a GRR of 3 for high-risk homozygotes and a GRR of 1.4 for heterozygotes, the power was estimated to be 73%.

3.4 Discussion

3.4.1 Replication of Previous Results?

3.4.1.1 Serotonin Transporter

Five serotonin transporter polymorphisms were analysed for association with autistic disorder in our Irish sample. The short promoter allele was the only individual allele to reach significance (p value of 0.03554, see Table 3.1). Haplotypes constructed from alleles that were over-transmitted showed significant excess transmission for all two-marker combinations, with the exception of the VNTR-SNP18 haplotype (See Table 3.1). Moving to 3 and 4 marker haplotypes, it is apparent that each haplotype that contains SNP10 C allele is statistically significant. The haplotype with the highest OR is a haplotype containing the first 4 markers in the gene (Prom-SNP10-VNTR-SNP18 haplotype, C-12rp-G form) with OR of 2.4 and p value of 0.0085. It is interesting to note that as one moves towards the 5' end of the gene, the ORs calculated for the haplotypes increase, suggesting that the polymorphism(s) linked to autism are located at this end of the gene. Additionally, the observed pattern of associated haplotypes suggests that risk is associated with a combination of polymorphisms working together to alter the function or levels of the serotonin transporter produced.

Our findings contradict those of Klauck et al and Yirmiya et al, who reported association with the long promoter allele in samples of 52 German trios and 34 Israeli trios respectively^{4,213}. However, Cook et al found transmission of the short promoter allele in a sample of 86 trios, although this finding was not replicated in a further 81 new trios^{190 3}. Another study using 57 New England and 80 AGRE multiplex families found nominal association with the short allele of the serotonin promoter variant⁸⁷. More recently, Tordjman et al found association of the long

promoter variant in their total sample. However when the patient sample was divided into those mildly and severely affected based on the social and communication domains of the ADI, significant transmission of the short allele was found in those patients severely affected²¹². Finally, there are those studies that have reported no association between the promoter variants and autism^{5,335 1,343}. Due to the increasing number of reports between autism and the serotonin transporter, with the correspondingly different associations, the inclusion of the Irish sample does little to elucidate or simplify the confusion. Association with either the long or short promoter allele, or a lack of association is a replication of at least one previous report. A joint-analysis of the serotonin transporter long/short promoter variant in 11 family based reported studies was undertaken to estimate the OR of the short promoter in a larger sample. A funnel plot (see Figure 3.4) indicated that there was no bias in the publication of reports. In the combined sample of 1090 affected individuals, there was over-transmission of the short promoter allele overall. However the OR associated with this over-transmission was not significant (OR = 1.059, p value = 0.288).

Excess transmission of haplotypes containing the 12 repeat allele of the VNTR, but not the 12 repeat allele on its own, had previously been reported by Cook et al and by Kim et al^{3,190}. A study examining the transmission of the serotonin transporter allele and the VNTR in a sample of Dutch individuals diagnosed with PDD reported the increased presence of 12 repeat allele in the VNTR in individuals who scored highly in the “rigid-compulsive” factor in the ADI-R test²¹¹.

As of yet, there have been no reports of association with the 3’UTR variation and autism.

3.4.1.2 Replication of TPH2 Results?

Coon et al recently reported the association of two variants in the *TPH2* gene with an American / Italian sample with autism ⁷. This was an interesting association study based on the importance of the *TPH2* gene in controlling the synthesis of serotonin from tryptophan in brain tissue ³³⁶. Coon et al found an increased frequency of the G allele of rs4341581 (p value = 0.013) and the T allele of rs1117900 (p value = 0.024) in a sample of individuals with autism in a case-control study. A haplotype containing these two variants was also significantly more frequent in the sample with autism in comparison to a control sample (p value = 0.005). In the Irish sample, there was increased transmission of the A allele of rs1117900 (p value = 0.0406, see Table 3.4). This is not the original allele reported by Coon et al ⁷. Significant transmission of alleles of marker rs431581 was not observed in the Irish sample. Likewise, there was no increased transmission of the T allele of marker rs1843809, which showed increased transmission to a sample of Irish children diagnosed with ADHD ³⁴⁰.

3.4.2 Potential Reasons for Differing Results.

3.4.2.1 The Serotonin Transporter Gene

The studies included in the joint-analysis (see Table 3.3) highlight the wide range of results reported in the literature with the serotonin promoter long/short variation alone. An increasing number of meta-analyses have been published in an attempt to clarify some associations with differing results, e.g. testing the association of the CA(n) repeat in the dopamine D5 receptor in ADHD ³⁴⁴ and the association of the serotonin transporter short/long promoter variant in affective disorders ³⁴⁵. The joint-analysis result indicates that the overall effect of the serotonin transporter long/short promoter variant, plays only a minor role, if any, in the development of autism. That is not to say that the serotonin transporter does not play any role in the development.

It may play a role in a small subset of individuals. Phenotypic heterogeneity is a potential problem in any genetic study of a psychiatric disorder and autism is no exception. The sample described here underwent rigorous assessment using the ADI-R and ADOS-G instruments to define the phenotype. However heterogeneity may still exist within a sample meeting criteria for the narrow diagnosis of autism using these diagnostic tools. Sub-grouping the sample using particular aspects of the phenotype such as social abnormalities or restrictive and repetitive behaviours may be advantageous. This has been attempted in some studies, such as Tordjman et al and Mulder et al, with mixed results. In the case of Tordjman et al, categorising individuals based on their communication and social impairments resulted in the increased transmission of the short promoter allele in those individuals exhibiting severe impairments in social interaction (p value = 0.033), whereas those with mild/moderate impairments had preferential transmission of the long promoter allele (p value = 0.049)²¹². In the study undertaken by Mulder et al²¹¹, where the sample was also split according to severity in social and communication impairments, there were no significant differences in genotype frequencies were observed between those individuals severely affected and those only mildly/moderately affected in both domains. There was a trend towards an increased frequency of individuals homozygous for the short promoter allele in individuals with high scores in the compulsive obsessive domain. However, as the phenotypic data is often qualitative in nature and not quantitative, this is not a straightforward exercise. Cluster analysis has been proposed as an alternative to determine genetically relevant aspects of the phenotype, but, as for any study attempting to dissect phenotype, larger samples are required.

In addition to phenotypic heterogeneity there is the problem of genetic heterogeneity. There is a possibility that there are a number of haplotypes that confer susceptibility to autism. A study by Sutcliffe et al found three novel highly conserved coding mutations, all in the transmembrane domains, that segregated with autism⁹³. Two of these had functional significance. These mutations also showed a phenotypic correlation with rigid-compulsive behaviours. Sutcliffe suggests that the most

parsimonious model “involves multiple risk alleles (including the serotonin transporter promoter) that act in different families to collectively account for the observed linkage” in both his and other samples with autism⁹³. It is also possible that the contrasting long and short allele findings may occur because these variants are in LD with a causative variant but differ between populations. Also, another mutation or even the same mutation may have arisen on two separate occasions and are associated with different haplotypes. Another likely reason for the confounding results is that the underlying mechanisms of hyperserotonemia are heterogeneous³⁴⁶.

Finally, one must take into consideration that some or all of the above findings are false-positive results. The results of the joint-analysis would support this finding. Although the OR for individual studies ranges from 0.375²¹² to 1.655¹⁹⁰, the mean OR for association of the short promoter allele in the 11 samples included in the joint-analysis was only 1.06 (p value = 0.288). The funnel plot (see Figure 3.4) also highlights the tendency of the larger individual samples to conform to the overall mean. An example of this is the contrast in ORs between the largest and smallest studies. In the sample of Yirmiya et al, in which there were 35 families with an individual affected by autism, the OR of the short promoter allele was 0.44, i.e. there was over-transmission and association with the long promoter allele²¹³. However in the study by Coutinho et al, in which there were 196 affected individuals and their families, the OR of the short allele was just 1.14². The larger study has an OR closer to the mean of the combined sample.

The fact that there are so many associations with the transporter gene and autism, even though the associations are not the same in each population, would lead one to believe that the serotonin transporter does not play a major role in autism, although it may be partly responsible for the development of autism in a minority of cases. The role of the serotonergic system in autism is complex and as yet not fully understood.

3.4.2.2 The TPH2 Gene

The association with both alleles of the same marker in two different populations also occurred at the *TPH2* gene. Unlike the serotonin transporter however, there is only the original study that Coon et al published ⁷. This study by Coon et al was a case/control study in a sample of American/Italian individuals, whereas the study presented here was a family based study of Irish individuals. It is possible that population stratification may have resulted in a false positive finding. It is also possible that the finding in the Irish sample is a false positive. Equally possible is the occurrence of two false positive results, each with a different allele. Finally, in a fashion similar to that suggested for the serotonin transporter, there may be reasons why associations with different alleles occur. At this stage it is impossible to differentiate between the possibilities. Further replication studies in independent samples are required to determine the possible association of variants in the *TPH2* gene in other samples of individuals with autism.

3.4.3 Conclusion

There is a large and varied range of evidence supporting the role of serotonin in the development of the autistic aetiology. There is little doubt that all the findings (see the introduction) reported to date are indicative of a role for serotonin. Unfortunately the role of the serotonin transporter in this process is not clear-cut. It is still uncertain to what extent this gene is involved. In the Irish sample describe here there is over-transmission of the short promoter allele (See Section 3.3.1.1.2), which is a replication of a report by Cook et al ¹⁹⁰. However, other reports find no association ^{1,5,335} or association with the long promoter allele ^{4,213}, see Section 3.1.4. Joint-analysis of the serotonin transporter promoter variant revealed a non-significant role for the serotonin transporter gene promoter variant in the development of autism overall. This joint-analysis revealed an increased transmission of the short promoter

allele, but this was non-significant (OR = 1.06, p value = 0.288). Further investigations are required to determine if the serotonin transporter has a role in a small sub-group of individuals diagnosed with autism. The TPH2 gene has only recently been identified. This present study has, in a manner similar to the serotonin transporter gene, not replicated (with the original allele) the initial report by Coon et al of association of the T allele with autism ⁷. The serotonin pathway is a complicated one and it is possible that the effects of a small number of genes, each of minor effect, together contribute to the hyperserotonemia observed and replicated in many samples (see Section 3.1.2). It is also possible that not all variants are required for the development of autism, and that some of these may be population specific. The serotonin system and its role in the development of autism remains one of the interesting puzzles to be solved in the field of autism research.

CHAPTER 4

THE DOPAMINE- β -HYDROXYLASE GENE AND AUTISM

4.1 Introduction

4.1.1 Dopamine and Autism

Dopamine is a ubiquitous neurotransmitter in the CNS. Dopamine is synthesised in the body by the decarboxylation of DOPA by aromatic L-amino acid decarboxylase (see Figure 4.1). A role for dopamine in the aetiology of autism has been suggested by decreased dopaminergic activity reported in the prefrontal cortex of autistic children³⁴⁷, and by the therapeutic benefits observed from the oral administration of Risperidone and Haloperidol in the condition. Risperidone is an atypical antipsychotic medication that acts to block dopamine receptors. In a study comparing risperidone administration with placebo, treatment with risperidone led to significant decreases in the level of irritability in 69% of children with autism (n = 49) compared to 12% in the placebo group (n = 52)¹⁷³. Similarly haloperidol (a postsynaptic dopamine-receptor antagonist) has been reported to improve severe behavioural problems in individuals affected by autism, although long-term use is limited by side-effects³⁴⁸.

Dopamine- β -hydroxylase (DBH) is the enzyme, which is responsible for the conversion of dopamine to norepinephrine (NE), and thus it is important to maintaining the levels of both of these neurotransmitters (see Figure 4.1). Decreased DBH activity has been reported in children with autism^{349 350}, while increased DBH activity has been reported in another study³⁵¹. A further study found no difference between DBH activity in a sample of cases with autism in comparison to a sample with Down's syndrome³⁵². Although these studies provide an indication that DBH activity may be linked to autism, the power in each study was limited due to small sample size (n < 20 in all cases). A larger study of parent-child trios, n = 37 trios, with autism found lower serum DBH activity in mothers compared with controls.

Thirty-three percent of mothers in this study were found to be homozygous for a deletion polymorphism in the promoter region compared with 20% of controls. As DBH converts dopamine to norepinephrine, alterations in DBH activity could result in an imbalance between these neurotransmitters. Altered NE levels are also involved in other neuropsychiatric disorders including anxiety. Yohimbine, a drug which is believed to increase the synaptic availability of NE, induces a high level of panic-like anxiety in individuals suffering from panic disorder, whereas clonidine, an adrenergic agonist, reduces the rate of panic induced by the panicogenic probe sodium lactate^{353,354}. The NE system is one of the systems responsible for the response to stress. The neurons involved in NE transmission originate in the locus ceruleus, the medulla and pons. Upon stress activation they send signals to many parts of the brain including the amygdala, which acts to establish 'emotional valence' to stimuli and coordinate the correct response, and the hippocampus, both of which have been linked to autism (see Palmen et al for a complete review of neuropathological findings in autism)^{355 356}. Interestingly DNA variations in the DBH gene have been associated with other psychiatric conditions including attention deficit hyperactivity disorder (ADHD), cocaine induced paranoia, and major depression^{357 358 359}.

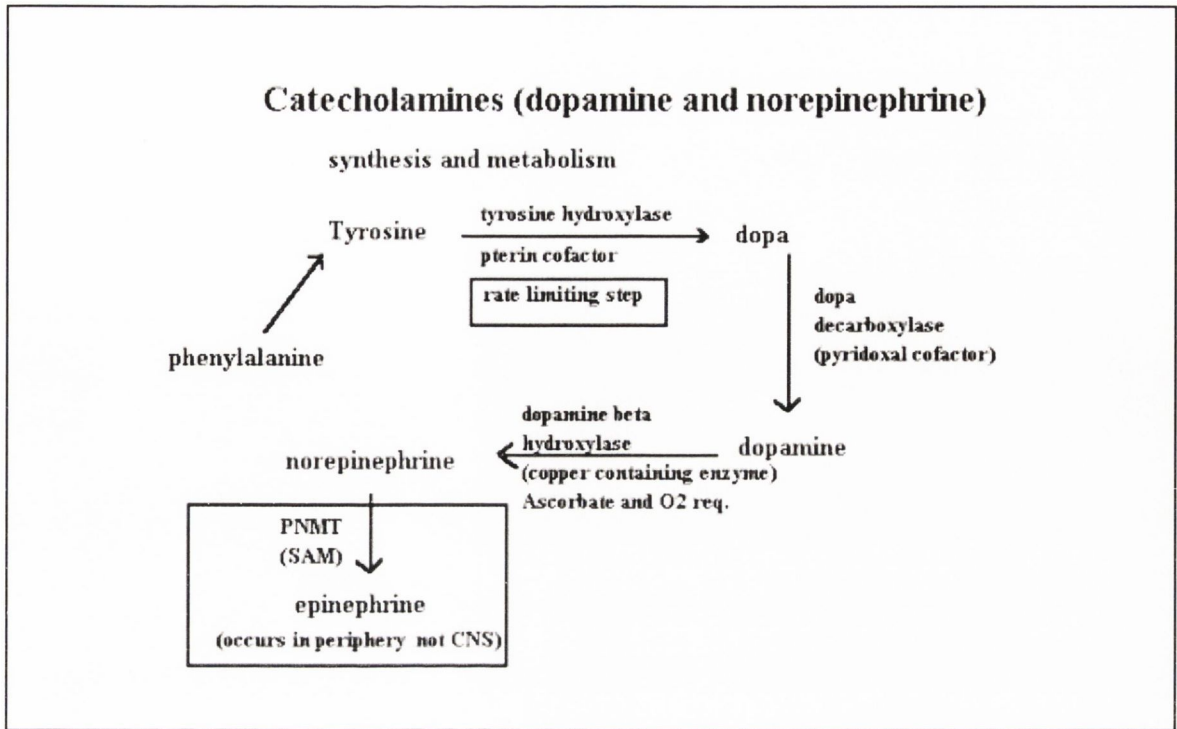


Figure 4.1 THE DOPAMINE / NOREPINEPHRINE PATHWAY.

Dopamine is converted to norepinephrine by dopamine- β -hydroxylase in the presence of ascorbate and oxygen. Although DBH is not the rate-limiting step of this pathway, its action and activity are important in controlling the levels of both dopamine and norepinephrine. This figure has been taken from <http://www.cellscience.com/CCA.htm>.

A number of other genes involved in the dopamine pathway have been investigated for an association with autism (see Figure 4.1). The Monoamine Oxidase genes (*MAO-A* and *MAO-B*) have been the most widely investigated, with studies at both DNA and protein levels being reported. A study by Takahashi et al in 1977, found no differences between MAO activity between individuals with autism (n = 20) and age-matched controls (n = 30)³²⁵. Filinger et al also reported no differences between autistic individuals (n = 6) compared to controls (n = 14). However, individuals with childhood-onset Pervasive Developmental Disorders (n = 6) presented significantly lower V_{max} (maximum catalytic speed of enzyme) and K_m (a measure of the specificity of the binding site of the enzyme for the substrate) levels compared to controls. A polymorphism in the promoter region of this gene has been linked to IQ in individuals with autism³⁶⁰. No association has been reported between autism and variants in the DOPA decarboxylase and catechol-O-methyltransferase (*COMT*) gene and COMT activity^{361 213 362}.

4.1.2 The Dopamine Beta Hydroxylase Gene

The gene for DBH is located on chromosome 9q34.2 and is 22,981bp long^{363 364} (<http://genecards.bcgsc.bc.ca/cgi-bin/carddisp?DBH>). It contains 12 exons, and encodes 2 types of mRNA due to alternative polyadenylation, denoted ‘type A’ and ‘type B’, with type A being more abundant than type B. Type A differs from type B due to the presence of an extra 300bp sequence (cDNA nucleotides 2394 – 2694). Type A is five times more abundant than type B³⁶⁵. This may be due to the presence of a “stabilising” sequence within this region. The 3’UTR region of a gene has been reported to be involved in both mRNA stability and efficiency^{366 367 368}. The gene produces a protein 603 amino acids long, which belongs to the copper type II, ascorbate-dependent monooxygenase family (information from gene cards). The gene is expressed restrictively in norepinephrine and epinephrine neurons, in addition to neurosecretory cells in the nervous system. Their expression within the brainstem is limited to a number of distinctive loci, the lateral tegmentum in the

medulla oblongata, the nucleus of the solitary tract, the locus coeruleus and the lateral tegmental area in the pons^{369 370,371}.

The transcription initiation site for the gene is located 52bp upstream of the initiation codon. The promoter region of the gene is located close to the initiation codon, with the TATA box at -29 to -24, and the CCAAT box at -159 to -155. A GC box also lies 14 to 19bp downstream of the transcription initiation site. A number of other sequence motifs including a cyclic AMP response element (CRE) and glucocorticoid response element lie upstream of the gene³⁶⁵.

DBH activity levels vary from individual to individual, with ~4% of individuals having a “very low” serum DBH activity, i.e. an activity of less than 50 U (where 1U of enzyme activity represents the production of 1 nmol of phenyl ethanolamine per hour per ml of serum incubated at 37°C)³⁷². This ‘low activity’ was seen to segregate in a recessive manner within families, and a strong sibling-sibling correlation has been reported³⁷³. These observations led to the hypothesis by Weinshilboum that a single low activity functional polymorphism (DBH^L) existed in the European population with a frequency of 20%, which accounted for this variation³⁷⁴. It has taken many years to determine the underlying genetic variation responsible for DBH activity. A single SNP in the promoter region of the DBH gene, C-1021T, has been shown to be responsible for 35 –52% of this variation. Individuals with a TT genotype (~ 4 – 7% of a Japanese and European population respectively) have the lowest DBH activity, and it is these individuals who represent the low activity subgroup. CC homozygotes have high DBH activity, while CT individuals have activity that is intermediate between the two homozygous groups. A number of observations support this C-1021T as being an important functional variant. Firstly the proposed DBH^L allele appears to ‘lower plasma-DBH activity by diminishing the levels of circulating protein, rather than by decreasing the activity of homospecific enzymes’. The location of the C-1021T polymorphism in the promoter region of the *DBH* gene may alter the levels of transcription of the gene and would thereby account for this variation. Also a reporter gene experiment highlighted the

–600 to –1100 region of the gene as being functionally important in DBH expression³⁷⁵. Other polymorphisms within the gene exist, including a small number of non-synonymous polymorphisms (e.g. 444A→G polymorphism, and the 1368G→A polymorphism). However these polymorphisms cannot account for the variation in DBH levels activity observed. Neither of these observations conclusively proves that C-1021T is the DBH^L polymorphism; and there remains a possibility that C-1021T is in LD with another functional polymorphism. However there are no other polymorphisms reported that appear to be of the same functional significance.

4.1.3 Dopamine-β-hydroxylase & Epilepsy

Epilepsy is defined as “two unprovoked seizures of any type”³⁷⁶. The prevalence of epilepsy in children with autism varies from 5% to 38.3% and is approximately 10 fold higher than children in the general population. Although the variation in these frequency estimates is high, the differences can in part be explained by the bimodal age distribution of epilepsy. The first major peak of epilepsy in autistic individuals occurs early in life, around the age of 5 years, with a second major peak occurring at the beginning of adolescence (~10 years). The studies with a low prevalence of epilepsy are often those studies whose sample is very young³⁷⁷ whereas those with a higher prevalence rate include adolescents and adults³⁷⁸. Epilepsy also seems to be more common in individuals where the level of brain injury is likely to be more severe, e.g. lower ranges of intellectual functioning³⁷⁹⁻³⁸¹

A role for DBH in epilepsy has been suggested through studies of animal models. A rat model for epilepsy (genetically epilepsy prone rats (GEPRs)) exhibit low levels of DBH, presynaptic NE content, NE turnover, NE uptake and tyrosine hydroxylase levels³⁸²⁻³⁸⁶. DBH knockout mice (who have no NE activity) exhibit increased sensitivity to a number of seizure-inducing stimuli, which included 2,2,2-trifluoroethyl ether, pentylenetetrazol, kainic acid and high decibel sounds. Decreasing NE (in wild type mice) increases seizure susceptibility, whereas the administration of NE

leads to a protective effect³⁸⁷. Although *Dbh* knockout mice showed a significantly higher rate of *in utero* mortality, those that did survive show no gross morphological abnormalities, and Thomas et al were unable to discern any physical differences between living knockout mice and their wild-type littermates³⁸⁸.

4.1.4 Dopamine- β -hydroxylase Animal Models

As stated in section 4.1.3, DBH knockout mice are susceptible to convulsant stimuli, and show higher rates of *in utero* mortality. A study by Thomas and Palmiter reported the presence of behavioural deficits in *Dbh* *-/-* maternal mice³⁸⁹. These *Dbh* *-/-* mothers showed impaired nesting and pup-retrieval. Pups were observed scattered around the nest, often unwashed and with placentas still attached. Despite the normal mammary gland function, pups were not nursing. Olfaction and problem solving abilities were normal. However some deficits in motor function and learning were observed. Maternal behaviour in the knockout mice was rescued with the administration of 4-dihydroxyphenylserine (DOPS), which can be converted to NE by aromatic L-amino acid decarboxylase, thus increasing NE content within the CNS, on the evening before birth, leading to the survival of 50% of pups. If the administration of DOPS was left until the morning after birth, rescuing did not take place, and the pups died. These results indicate that norepinephrine may be responsible for responding to environmental cues and determining what is relevant and important. The effects of administration of DOPS lasted into subsequent pregnancies³⁸⁹. The role of DBH in social behaviour, a characteristic affected in autism, implicates DBH in this developmental disorder.

4.1.5 The Aims

The aims of this study were three-fold.

- 1) To attempt to replicate Robinson et al ³⁹³ by examining the transmission of 7 markers spaced throughout the gene, looking for an association with autism.
- 2) To measure DBH activity in affected individuals and their parents.
- 3) To establish the contribution of each variant to the overall level of DBH activity.

4.2 Materials & Methods

4.2.1 Measuring DBH Activity

Plasma DBH activity was measured by Dr. George Anderson, of Yale Child Study Centre, using a sensitive high performance liquid chromatographic-fluorometric method (see Section 2.2.6). Individuals on medications, which may affect DBH activity i.e. beta-blockers, anti-psychotic medication, insulin and the oral contraceptive pill, were excluded from analysis³⁹⁰⁻³⁹².

4.2.2 Amplification of DBH Variants

Seven variants (6 SNPs, and an in/del variant) were chosen to be screened in our population. The selection of variants was based on a number of criteria:

- 1) Putative functionality (the C-1021T SNP, see section 4.1.2 for further details)
- 2) Non-synonymous changes (i.e. variants in exons 5 and 11)
- 3) Location within exons (exons 12, 11)
- 4) Use in the Robinson paper (In/del variant)³⁹³
- 5) Regions close to the potential promoter (C-2124T, C-1333T).

Fragments were amplified under varying conditions (see Appendix II, sections 1.1 and 1.2, for further information).

4.2.3 Statistics

Analyses were performed by assessing the transmission of individual polymorphisms using the TDT. Transmission of haplotypes was assessed using the UNPHASED program, and logistic regression based analysis was conducted for the genetic variants against DBH activity.

DBH activity was measured *in triplicate*, and the mean of these readings was taken for all subsequent tests.

Due to the observation that DBH activity increases over the first 3 years of life, it was necessary to ensure that any potential DBH activity differences between parents and their affected offspring were not due to this age effect³⁷³. The square root of children's DBH activity was plotted against their age (in months) twice. It was necessary to use the square root of DBH activity in order to normalise the data set. The first analysis included all probands in the study regardless of age. The second analysis included only those probands aged 6 years and over. In both cases a regression line was fitted to the data and the slopes of these lines noted.

Paired t-tests were used to make comparisons between father-child DBH activity and mother-child DBH activity. An independent t-test was used to compare differences between mother-father DBH activity and proband-child control DBH activity. The heritability of DBH activity was measured as described in Section 2.3.2.4. Kruskal-Wallis testing was undertaken to test if different allelic variants were associated with different DBH activity. Regression analysis was performed to determine the correlation of each marker to the square root of DBH activity, i.e. a marker – phenotype correlation (see Section 2.3.2.6).

4.3 Results

Plasma samples were collected from 54 full trios, 7 mother-proband duos and 3 father-proband duos. Individuals on medication ($n = 5$), which may affect DBH activity, e.g. beta-blockers, anti-psychotic medication, insulin and the oral contraceptive pill, were excluded from analyses.

4.3.1 Age effect and DBH Activity.

Probands DBH activity was plotted against age of proband in months. In the first test all probands were included in the analysis ($n = 59$, mean age = 140 months, standard deviation = 89 months, age range = 55, 408 months). In the second test only individuals who were greater than 72 months of age were included ($n = 45$, mean age = 166 months, standard deviation = 90 months, age range = 73, 408 months). A regression line was fitted to each dataset, and the correlation, i.e. slope, calculated. No significant correlation between the square root of DBH activity and age was found, even when all probands were included in the analysis. In the first dataset, the correlation was calculated to be 0.0025 ($p = 0.923$), see Fig 4.2. When probands less than 72 months were excluded the correlation was 0.00004 ($p = 0.992$), see Fig. 4.3. These results show that within this sample, there is no statistical change in DBH levels over time, therefore all probands, regardless of age were included in subsequent analysis in order to increase power.

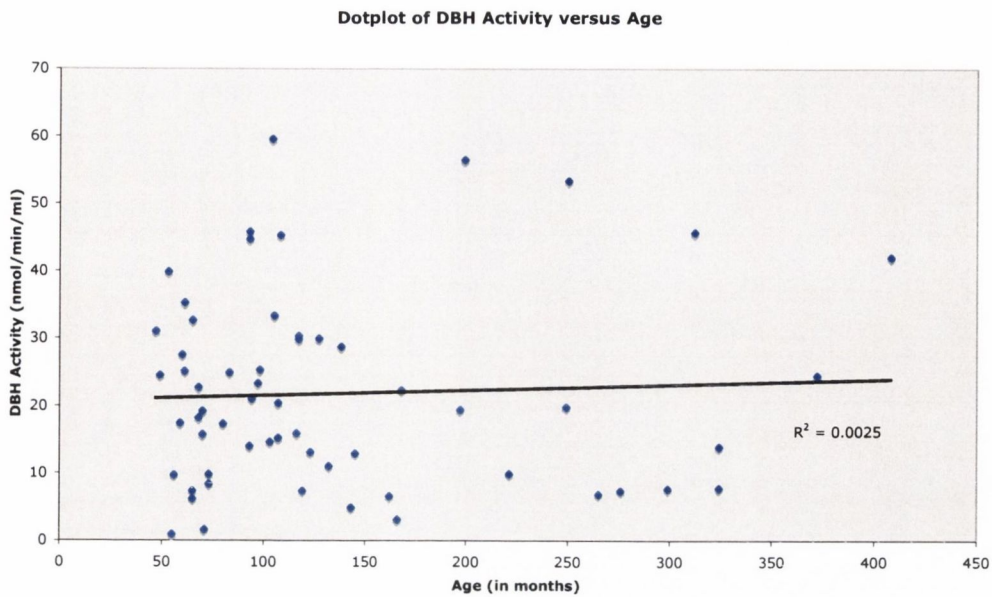


Figure 4.2 DOT PLOT SHOWING RELATIONSHIP BETWEEN TRANSFORMED DBH ACTIVITY AND AGE IN THE DATASET INCLUDING ALL PROBANDS. DBH activity was transformed by using the square root function, and age was measured in months. Regression analysis indicated that there was no increase in DBH Activity with age in this all-inclusive dataset. $R^2 = 0.0025$ with a corresponding p value of 0.923.

Dotplot of DBH Activity versus Age (proband > 72 months)

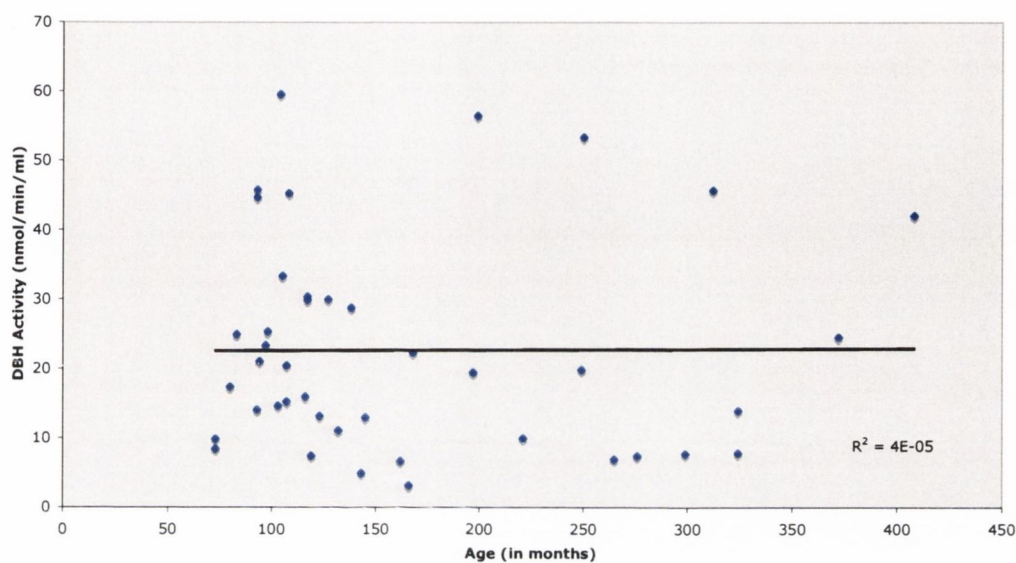


Figure 4.3 DOT PLOT SHOWING RELATIONSHIP BETWEEN TRANSFORMED DBH ACTIVITY AND AGE (IN MONTHS) IN INDIVIDUALS GREATER THAN 72 MONTHS OF AGE. Regression analysis indicated that there was no increase in DBH Activity with age in this dataset. $R^2 = 0.00004$ with a corresponding p value of 0.992.

4.3.2 Heritability of DBH Activity

Regression analysis of DBH activity was used to estimate the heritability of DBH activity (see Section 2.3.2.4). The mean parental DBH activity was plotted against proband DBH activity (see Figure 4.4). A regression line was fitted and the slope (i.e. heritability) was calculated. The heritability of DBH activity was 44.92%. A spouse-spouse DBH activity correlation analysis was undertaken (see Figure 4.5), to ensure that the heritability measured above a correct estimate. As expected, the correlation between paternal and maternal DBH activity was negligible ($r^2 = 0.0025$).

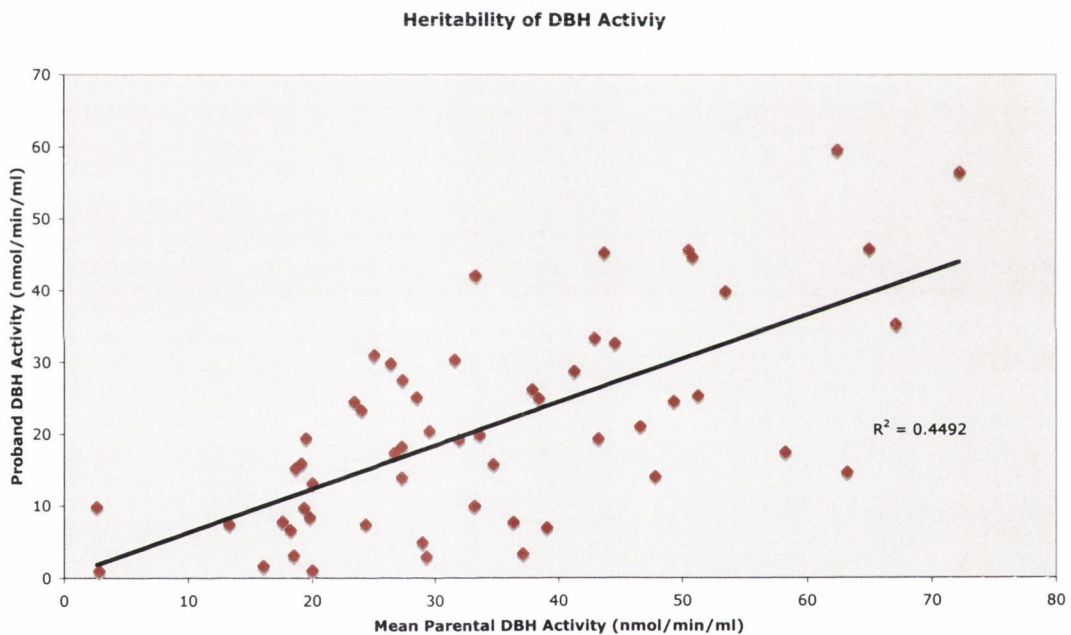


Figure 4.4 MEASUREMENT OF THE HERITABILITY OF DBH ACTIVITY.

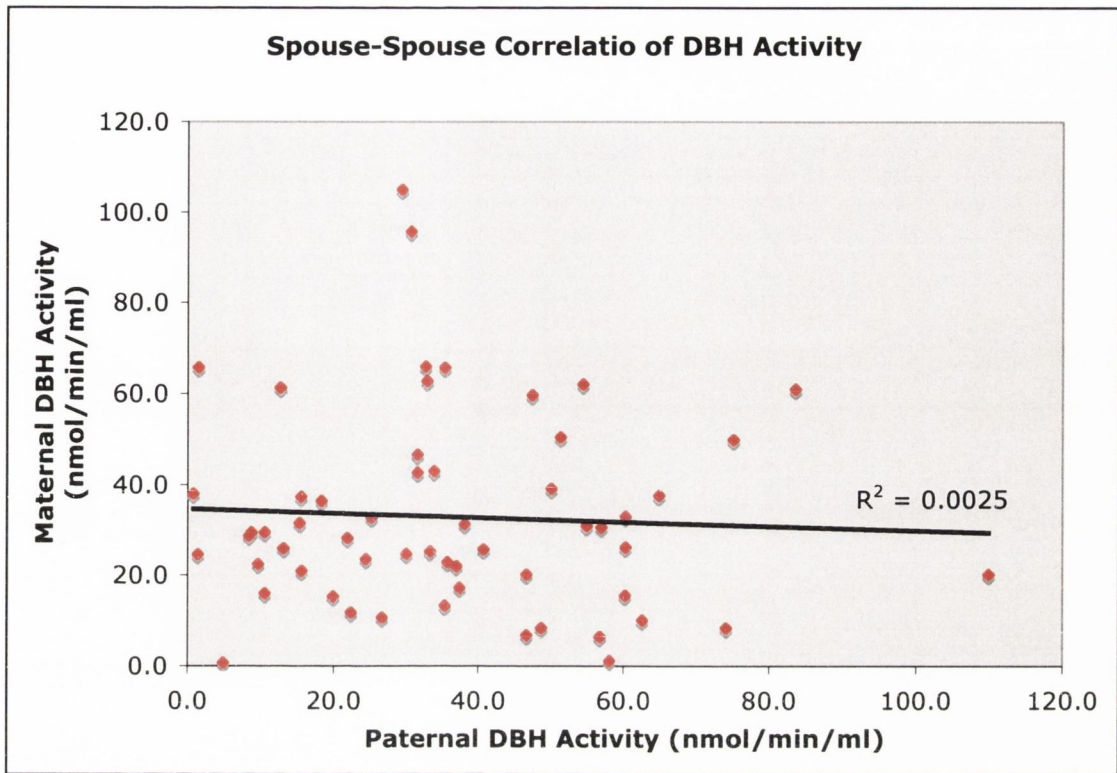


Figure 4.5 Spouse-Spouse Correlation of DBH Activity.

4.3.3 DBH Activity. Values and Comparisons.

The population was divided into mothers, fathers and affected probands. The parental midpoint (i.e. (Mothers DBH Activity + Fathers DBH Activity) +2)) was also calculated, as this was considered to be the best estimate of overall parental contribution to a probands DBH activity. The mean, variance and standard deviation of the DBH activity of each group were measured. The maternal DBH activity mean was calculated to be 34.811 nmol/min/ml and the mean paternal DBH activity was calculated to be 35.107 nmol/min/ml (see Table 4.4.2 and Figure 4.6). The difference between these parental DBH activities was compared in 54 parents. With a mean difference of ~ 0.18, the difference was deemed to be non-significant ($p = 0.97$, 95% Confidence Interval = -9.64, 9.28). This result was to be expected, as there have been

no reports of a sex difference in a normal control population with regard to DBH activity³⁹³. Power calculations (using http://www.dssresearch.com/toolkit/spcalc/power_a2.asp) indicates that the differences between children and their mothers, fathers and parental means are extremely powerful, ranging from 98.5% to > 99.9%.

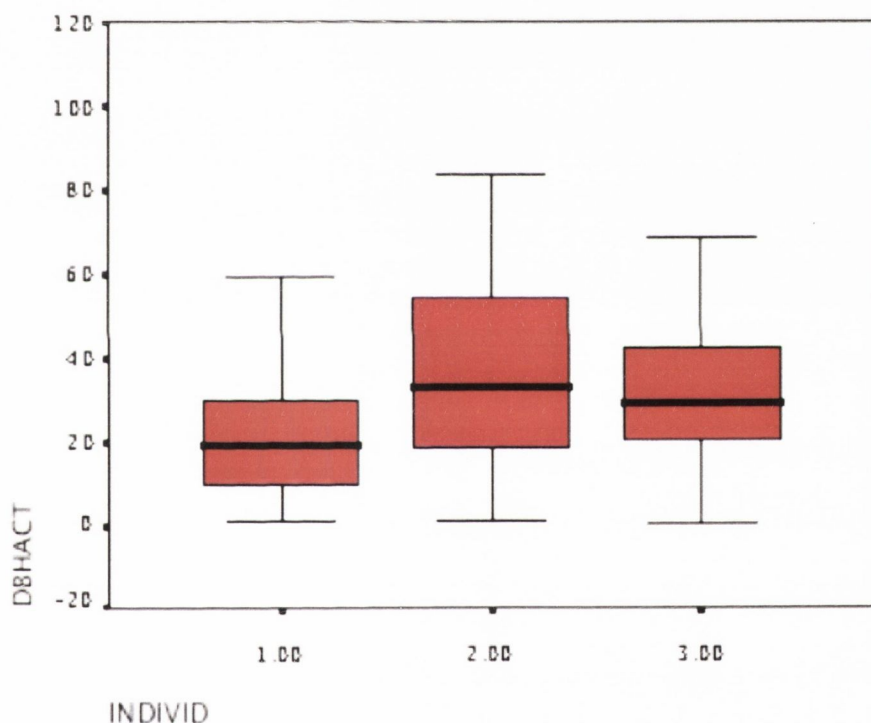


Figure 4.6 BOXPLOT OF DBH ACTIVITY IN MOTHERS, FATHERS AND CHILDREN. Individuals 1 represent children, whose mean DBH activity is 20.330 nmol/min/ml. Individuals 2 represent fathers, whose mean DBH activity is 35.107 nmol/min/ml. Individuals 3 represent mothers, whose mean DBH activity is 34.507 nmol/min/ml. The average child DBH activity is much lower than that of either parent.

Comparison Pair		Mean	N	Std Dev	Std Error Mean
M/F	Mother	34.212	51	24.00253	3.3610
	Father	34.392	51	21.43085	3.0009
M/C	Mother	34.811	56	23.486	3.1385
	Child	19.845	56	12.2942	1.6429
F/C	Father	35.107	54	21.2278	2.8887
	Child	20.330	54	11.9854	1.6310
PM/C	Parental Mean	34.302	51	15.3335	2.1471
	Child	19.447	51	11.6064	1.6252

Table 4.1 DBH ACTIVITY PAIRED DIFFERENCE SUMMARY. DBH summary statistics for pairs involved in paired t-tests. Parental mean (PM) is the average of mother (M) and father (F) DBH activity, which is then compared to the probands (C) activity. In all cases, the children's DBH activity is lower than that to which it is being compared, i.e. mothers, fathers and parental mean DBH Activity. The number of pairs varied between comparisons due to the presence of incomplete data for all members of the sample.

	Paired Differences					T	Sig (2-tail)	Power to Detect Difference
	Mean Difference	Std. Dev	Std. Error Mean	95% Confidence Interval of Difference				
				Lower	Upper			
M vs. F	0.1804	33.62	4.7078	-9.636	9.2755	0.038	0.97	5.40%
M vs. C	14.966	22.533	3.0111	8.932	21	4.97	6.89 x 10⁻⁶	98.50%
F vs. C	14.778	18.0178	2.4518	9.86	19.696	6.027	16.4 x 10⁻⁶	99.80%
PM vs. C	14.85	12.5582	1.7585	11.32	18.39	8.448	3.4 x 10⁻¹¹	> 99.9%

Table 4.2 PAIRED SAMPLES t-TEST FOR DIFFERENCES IN DBH ACTIVITY. Paired t-tests were undertaken to investigate if there were any statistical differences between the average DBH activity in mothers, fathers, children and the parental mean. M = mothers DBH activity, F = fathers DBH activity, C = child DBH activity and PM = parental mean DBH. All significant differences are highlighted in bold and bright yellow.

4.3.4 Comparison of Irish Proband DBH Activity with Previously Reported Children Control Sample.

Due to the lack of an Irish age-matched control sample, DBH activity from a normal child sample (between the ages of 6 and 12) was used from a publication from Weinshilboum et al ³⁷³. The mean of this sample (see Section 4.2.3 for more details) was calculated to be 34.429 nmol/min/ml, with a standard deviation of 20.595 nmol/min/ml (see Table 4.3). A t-test was undertaken to compare the mean of this sample to the proband sample described here. Individuals with autism were found to have a significantly lower DBH activity in comparison to the child control sample (T statistic = 5.3038, p value = 1.9737 x 10⁻⁸), see Table 4.3 and Figure 4.7. The power of this statistical test was found to be greater than 99.99% according to the power calculator available at http://www.dssresearch.com/toolkit/spcalc/power_a2.asp.

Boxplot of DBH Activity in Proband Sample and Child Control Sample

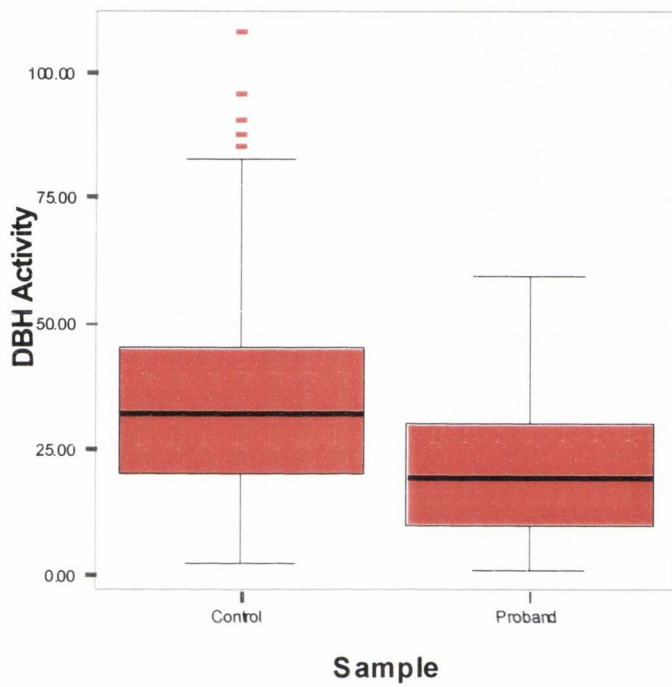


FIGURE 4.7 BOX-PLOT OF DBH ACTIVITY IN PROBAND SAMPLE AND CHILD CONTROL SAMPLE. A number of outliers were present in the Weinshilboum sample³⁷³. As no further details were available, all outliers were included in the test.

	Proband Sample	Control Sample
Mean	21.9	34.429
Std. Dev	14.058	20.595
Std. Error Mean	1.815	1.137
N	59	317
T statistic	5.3038	
p value	1.9737 x 10 ⁻⁸	

TABLE 4.3 TWO SAMPLE t-TEST FOR DIFFERENCES IN DBH ACTIVITY IN PROBAND SAMPLE AND A CHILD CONTROL SAMPLE

4.3.5 Comparison of DBH Activity in Individuals with Epilepsy

There were 5 individuals with autism in the sample that also had a diagnosis of epilepsy. The average DBH activity in these individuals were compared to the normal child control sample as described in Weinshilboum et al ³⁷³, and also to the remaining probands in the autism sample. No statistical differences in DBH activity were observed between those individuals with autism and epilepsy and those with autism or a control (see Table 4.4). This is not surprising however due to the small sample of individuals with epilepsy and autism (n = 5).

	Epilepsy	Autism	Control
Mean DBH Activity	35.7	20.6	34.429
Std. Dev	17.6	14.508	20.595
N	5	51	317
Epilepsy vs Autsim	t = 1.705	p value = 0.18	power = 58.6%
Epilepsy vs Control	t = 0.160	p value = 0.88	power = 6.9%

TABLE 4.4 COMPARISON OF DBH ACTIVITY BETWEEN INDIVIDUALS WITH EPILEPSY & AUTISM, AUTISM, AND A CHILD CONTROL SAMPLE. The statistical differences between DBH activities were calculated using t-tests, and power of each test was calculated using http://www.dssresearch.com/toolkit/spcalc/power_a2.asp.

4.3.6 Investigation of the Association of Allelic Variants with DBH Activity.

Kruskal-Wallis testing was used to evaluate if allelic variants were associated with different parental DBH activity levels. It was necessary to use a non-parametric form of testing to allow for differences in sample sizes and standard deviations. Statistically significant differences between allelic associated DBH levels were observed for a number of markers, especially those markers close to the C-1021T variant (see Table 4.5). The box plot revealed that the TT homozygotes (11) of the C-1021T variant have the lowest level of DBH activity, with CC individuals (22) having the highest level of DBH activity. Heterozygote individuals (12) have a DBH level intermediate between that of the TT and CC homozygotes (see Fig 4.8), Kruskal-Wallis testing indicated that these differences in DBH activity are statistically significant ($\chi^2 = 26.156$, p-value = 2.09×10^{-6}). Significant differences also existed for the in/del variant, C-2124T, and

exon 2 (see table 4.5). There was no significant difference between exon 5, exon 11 and exon 12 allelic genotypes.

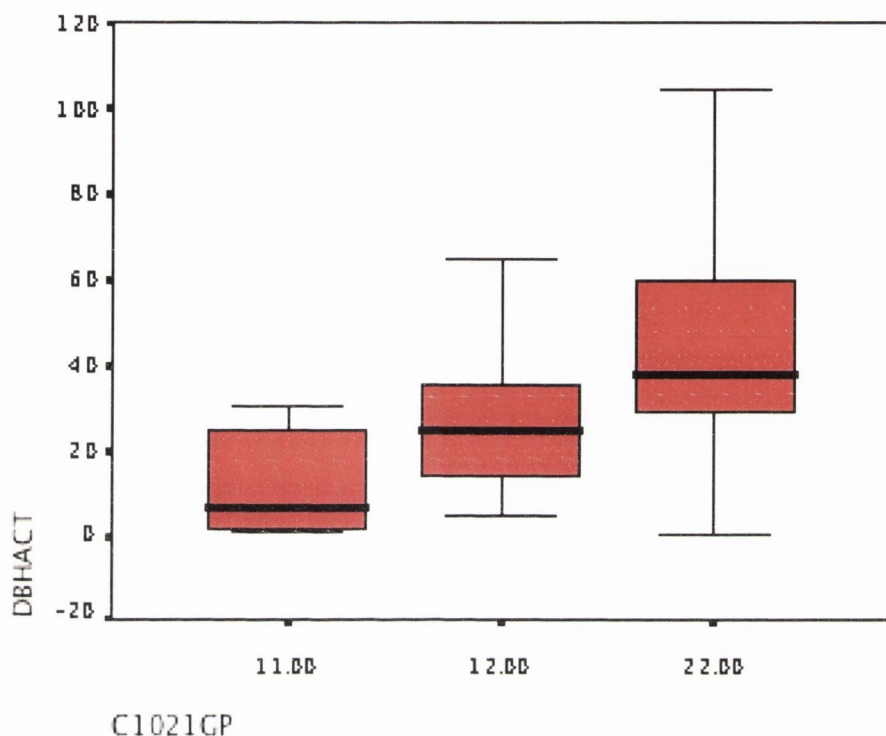


Figure 4.8 BOX PLOT OF C-1021T ALLELIC VARIANTS AND THEIR ASSOCIATED DBH ACTIVITY. TT homozygotes are represented by 11, and show the lowest DBH activity. CC homozygotes are represented by 22, and show the highest DBH activity, while heterozygous individuals (12) show a DBH activity level is intermediate of the level of either homozygote.

DBH Variant	Kruskal Wallis χ^2	d.f.	p-value
In/Del	9.501	2	0.009
C-2124T	8.766	2	0.012
C-1021T	26.156	2	2.09×10^{-6}
Exon 2	21.638	2	2×10^{-5}
Exon 5	0.339	2	0.844
Exon 11	0.671	1	0.431
Exon 12	1.591	2	0.451

Table 4.5 KRUSKAL-WALLIS TEST RESULTS FOR DIFFERING ALLELIC VARIANTS AND THE ASSOCIATED DBH ACTIVITY. Statistically different activities (highlighted in bold and bright yellow) were associated with allelic variants in the in/del, C-2124T, C-1021T and Exon 2 variants. As there were only 3 individuals with a TT genotype in the Exon 11 test, these three individuals were ignored (n must be > 5 to be included in the analysis). Therefore there were just two groups compared (CT vs. CC) and only 1 degree of freedom.

4.3.7 Regression Analysis of DBH Variants in Parents

Simple Regression analysis revealed that the *DBH* C-1021T variant is the most important variant analysed in this study. It was found to be responsible for over 40% of the variation in DBH activity ($r^2 = 0.408$, $p = 8.826 \times 10^{-14}$). The remaining variants were found to contribute to a lesser extent to DBH activity, with effects ranging from 6.2% (C-2124T) to 0% (in/del). The cumulative R^2 is the total variance explained by the presence of the markers, whereas the individual R^2 is the variance explained by the last marker added into the analysis.

Marker	Cumulative R ² (adjusted)	Individual R ² (adjusted)	F-value	p-value
C-1021T	0.408	0.408	39.891	8.823 x 10⁻¹⁴
In/del	0.402	0	19.965	2.350 x 10⁻¹²
C-2124T	0.470	0.062	14.877	1.138 x 10⁻¹¹
Exon 2	0.475	0.005	11.413	8.141 x 10⁻¹¹
Exon 5	0.522	0.047	10.492	5.909 x 10⁻¹¹
Exon 11	0.529	0.007	9.686	5.312 x 10⁻¹¹
Exon 12	0.538	0.009	7.021	6.248 x 10⁻⁰⁹

Table 4.6 REGRESSION ANALYSIS OF EFFECT OF VARIANTS ON OVERALL DBH ACTIVITY IN PARENTAL SAMPLES. The C-1021T marker is shown to be the strongest variant at describing / predicting DBH activity. It explains over 40% of DBH activity. Other variants, which predict DBH activity, albeit to a lesser extent, are the C-2124T marker (6.2%) and the Exon 5 marker (4.7%).

4.3.8 Regression Analysis of DBH Variants in Children.

Simple linear regression analysis was also performed to elucidate the role of the *DBH* variants in determining / explaining DBH activity in our affected children population.

Marker	Total R ² (adjusted)	Individual R ² (adjusted)	F-value	p-value
C-1021T	0.05	0.050	2.460	0.095
In/del	0.058	0.008	1.818	0.14
C-2124T	0.102	0.044	1.925	0.098
Exon 2	0.158	0.056	2.124	0.056
Exon 5	0.118	0.000	1.640	0.132
Exon 11	0.290	0.172	2.778	0.01
Exon 12	0.265	0.000	2.414	0.021

Table 4.7 REGRESSION ANALYSIS OF EFFECT OF VARIANTS ON OVERALL DBH ACTIVITY IN THE PROBAND SAMPLE. In our sample of individuals with autism, the C-1021T variant is shown to play a much weaker role in explaining DBH activity. DBH variants significantly explaining DBH variation are highlighted in bold and bright yellow. Those variants showing a non-significant trend towards significance (p value > 0.05) are highlighted in light yellow.

4.3.9 Linkage Disequilibrium Measurements

The level of linkage disequilibrium between the tested polymorphisms was measured using the D' statistic and the Haploview programme (see Sections 1.4.1.1 and Section 2.3.2.3). Strong LD was found between markers close to each other (see Figure 4.9). One haplotype block was shown to exist in the promoter region of the gene, including the insertion/deletion, the C-2124T and the C-1021T marker. The C-1021T and C-2124T markers are also in very strong LD with the Exon2 marker. However Exon2 is excluded from the block as the LD between Ex2 marker and

insertion/deletion is low (0.36). There is a break in LD after marker 5, which shows moderate LD with markers in the haploblock ($0.47 < D' < 0.60$) and is in moderately strong LD with Ex2 marker ($D' = 0.84$). However the exon 11 marker is not in LD with any of the markers in the haploblock, nor with the Exon 2 marker (see Figure 4.9). Due to the low minor allele frequency in both Ex5 and Ex11, the D' value is also low between these markers. Ex11 shows moderate LD with Ex12 however. This is once again not surprising due to the proximity of the two markers. These results are similar to those published by Zabetian et al (2003)²⁷⁷, who reported a haploblock stretching from the C-2124T marker to a marker (IVS4+601C→T) in intron 4. The haploblock for this sample does not stretch this far. This may be due to differences in linkage disequilibrium between populations. However, a study by Hawi et al³⁹⁴, which reported inter-marker LD in the *DBH* gene in a sample of Irish families with children with attention deficit hyperactivity disorder, showed that LD (in the parents of these affected children) between a marker in exon 2 and a marker in intron 5 was 0.845. Likewise, the level of LD between markers in the intron 5 and intron 9 (a distance of 9kb) was very weak ($D' = 0.172$)

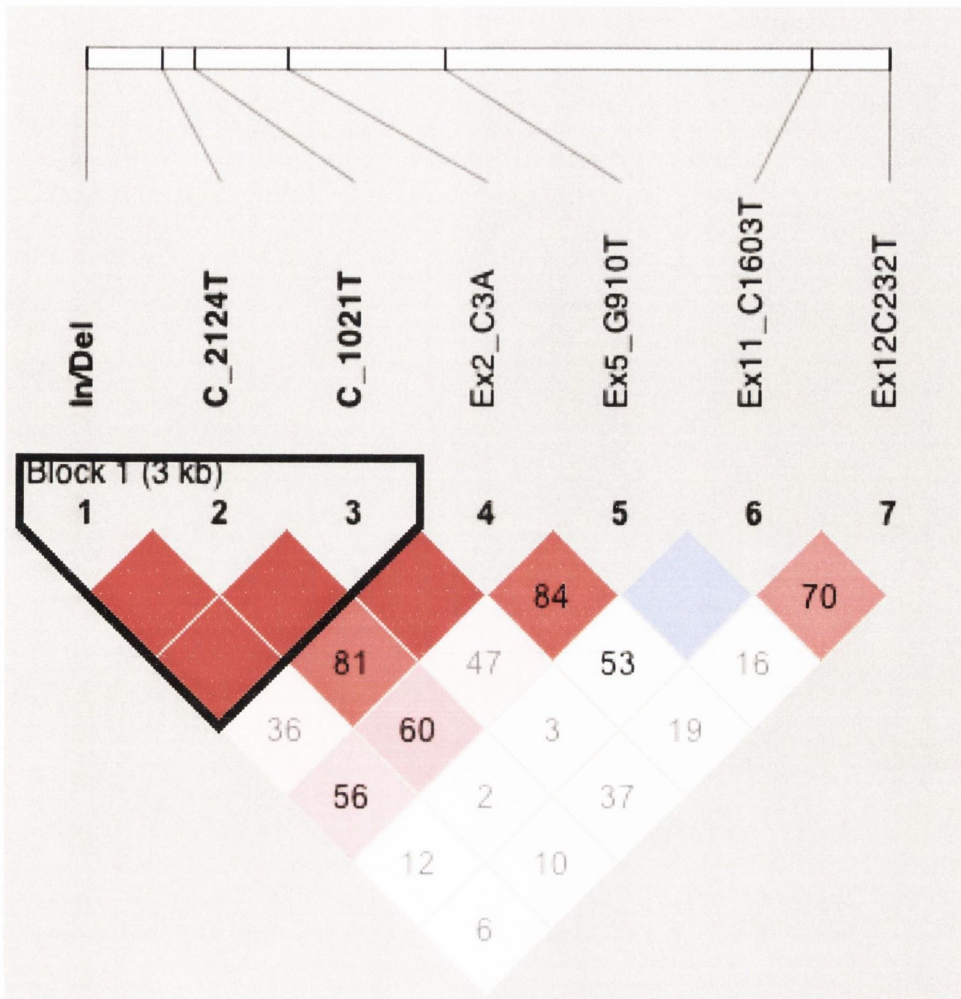


Figure-4.9 LD MEASURES AND HAPLOBLOCK FORMATION OF DBH VARIANTS. LD is measured using the D' statistic. A haplotype block was found to consist of the In/Del, C-2124T and C-1021T variants. The Exon 2 variant (Ex2C3A) is also in very strong LD with the C-2124T and the C-1021T variants. A break in LD is seen between the Exon 5 and Exon 11 variants.

4.3.10 TDT Testing

The transmission disequilibrium test (TDT) was undertaken with all variants using the Transmission Disequilibrium Test (see Section 2.3.1.2 for further details)³⁹⁵.

No significant preferential transmission of any allelic variant was observed (at $p < 0.05$) (see Table 4.4.8 for further information). Surprisingly the alleles associated with low DBH activity (i.e. the T allele of C-1021T) were not found to be over-transmitted to affected probands.

Marker	Allele 1	Allele 2	1T	1NT	2T	2NT	χ^2	p-value
In/Del	Insertion	Deletion	17	16	16	17	0.0304	0.8618
C-2124T	T	C	15	12	12	15	0.334	0.5633
C-1021T	C	T	9	8	8	9	0.05886	0.8083
Ex 2 (C3A)	A	C	20	28	28	20	1.34	0.2471
Ex 5 (G910T)	T	G	6	12	12	6	2.039	0.1533
Ex 11 (C1603T)	T	C	13	17	17	13	0.5349	0.4645
Ex 12 (C2320T)	T	C	17	23	23	17	0.9034	0.3419

Table 4.8 TDT RESULTS FOR DBH VARIANTS. 1T = number of transmissions of allele 1, 1NT = number of non-transmissions of allele 1, 2T = number of transmissions of allele 2, 2NT = number of non-transmissions of allele 2. A TDT χ^2 is calculated using the McNemar statistic (see Section 2.3.1.2), resulting in the p values as stated above. There is no significant over transmission of any one allelic variant to affected individuals suffering autism.

4.3.11 Transmission of Haplotypes

The construction of and analysis of transmission of haplotypes was undertaken using the UNPHASED programme (available from HGMP, see Section 2.3.2.3). Two and three-marker haplotypes were examined and the results are presented in Table 4.9. Although single haplotypes were observed to be over-transmitted, the global test did not reach the threshold of significance at $p < 0.05$. Once again, the haplotypes containing the alleles associated with low DBH activity were not found to be over-transmitted to individuals with autism.

Markers	Haplotype Over-transmitted	Hap T	Hap NT	Global LRS	Global p-value
In/Del – (C-214T)	Insertion – T	13	10	3.210	0.3603
(C-2124T) – (C1021T)	T – C	15	10	1.482	0.4766
C-1021T – Ex2	C – C	24	16	1.637	0.6511
Ex2 – Ex5	C – G	22	11	4.074	0.2536
Ex5 – Ex11	G – C	16	11	1.725	0.6313
Ex11 – Ex12	C – C	16	7	4.897	0.1795
In/Del – C2124T – C1021T	Insertion – T – C	13	9	3.9435	0.4137
C2124T – C1021T – Ex2	T – C – C	23	15	2.567	0.7663
C1021T – Ex2 – Ex5	C – C – G	21	10	7.328	0.2916
Ex2 – Ex5 – Ex11	C – G – C	23	11	5.709	0.4566
Ex5 – Ex11 – Ex12	G – C – C	19	10	6.792	0.4502

Table 4.9 HAPLOTYPE TRANSMISSION RESULTS FOR THE *DBH* VARIANTS. The haplotypes that showed any excess transmission are presented, along with the global χ^2 and the global p-value. No haplotypes were shown to be significantly over-transmitted to affected probands.

4.4 Discussion

4.4.1 DBH Activity with Age

As previous studies have indicated that there is a positive correlation of DBH activity over time, it was necessary to ensure that any differences in DBH activity were not as a result of this “age effect”. A previous study by Freedman et al reported an age increase in DBH activity in the first few years of life, especially around the ages of 2 and 3³⁹⁶. Another study by Weinshilboum et al in 1973, found that between the ages of 6 and 12, there were no increases in DBH activity in girls, and only very minor changes in boys³⁷³. In the Irish child population described here, likewise, there were no significant changes in DBH activity with age (see Section 4.3.1). Simple regression analysis was undertaken firstly in all child probands (n = 59), and secondly only in probands over the age of 6 years of age (n = 45). The age of probands in this population ranged from 47 months to 408 months. When all probands were included in the analysis, the effect of age on DBH activity was found to be minimal and non significant ($r^2 = 0.0025$, $p = 0.923$, see Figure 4.1). When only individuals greater than 6 years of age were included the effect of DBH activity on age was again found to be non-significant ($r^2 = 0.0004$, $p = 0.922$, see Figure 4.2). Although removing children less than 6 years (mean age of removed individuals = 62.88 months) did lead to a reduction in the “age effect”, the level of significance between both analyses remained almost identical. Therefore, it was decided to include all individuals in subsequent analysis. This was undertaken to increase the power of subsequent analysis. These findings confirm the results of Weinshilboum³⁷³, which indicate that there are no major changes in DBH activity over the age of 6 years.

4.4.2 Heritability of DBH Activity

Heritability of DBH activity in individuals diagnosed with autism was found to be ~45% (see Section 4.3.2). This was lower than previous reports regarding the heritability for DBH activity. There have been two studies that measured the heritability of DBH activity in a normal healthy control sample and a sample of affected individuals diagnosed with affective illnesses and their relatives. In the study by Oxensteirna et al, the heritability of DBH activity was measured as 98% in serum and 83% in cerebrospinal fluid in a normal healthy sample ⁸. In a sample of individuals diagnosed with affective illness, the heritability of plasma DBH activity was found to be greater than 90% ⁹. A study by Goldin et al found that a single locus (located on chromosome 9) accounted for 56% of the heritability ³⁹⁷. The C-1021T variant is now presumed, although not definitively proven, to be responsible for 35-52% of this variation. The finding of lower DBH heritability in individuals with autism is interesting, as it highlights a possible environmental effect in determining DBH activity.

4.4.3 Differences in DBH Activity

Unlike the Robinson study, this study did not show a lower maternal DBH activity ³⁹³. Within our sample, there was no difference in DBH activity levels between mothers and fathers of affected individuals (see Table 4.4.1). A similar study by Goldin et al, also found no differences in male and female DBH activity in a sample of relatives of individuals suffering major affective disorders ³⁹⁷. A limitation of this study is the lack of controls for the parent sample. In the Canadian Study by Robinson et al, DBH levels of both mothers and fathers were compared to sex matched controls. Mothers of autistic individuals were found to have significantly lower levels of DBH activity compared to the controls. Although our parental

sample has no sex-matched controls, their DBH activity levels are almost identical to the controls reported by Robinson et al (mean maternal DBH activity = 33.907 nmol/min/ml, mean paternal DBH activity = 35.107 nmol/min/ml, mean parental average DBH activity = 34.507 nmol/min/ml vs. mean Canadian control population DBH activity = 33.14 nmol/min/ml), and this would indicate to us that our parental DBH levels are comparable to those of controls. Another control sample had a mean DBH activity of 34.1 nmol/min/ml, indicating again that our parental sample is comparable to a normal control sample ³⁹⁸.

The children with autism were however found to have a significantly lower level of DBH activity compared to their parents and a previously reported normal child sample ³⁷³ (see Tables 4.4.2 and 4.4.3, and Figures 4.6 and 4.7). This result confirms the early results of the 1970's, in which DBH was reported to be lower in individuals with autism ^{349 350}. A major limitation of these previous studies was their very small sample sizes, which led to very low power.

A further limitation of this study is the lack of age-matched control sample from the normal Irish population. The question whether this decreased DBH activity is present only due to a normal "age effect" arises. Early studies report that this "age effect" takes place only during the early years of infancy ³⁷³. Regression analysis was undertaken in order to determine if there was any change in DBH activity due to the age of the children. Regression analysis with all probands indicated that age did not play a role in determining DBH activity ($r^2 = 0.0025$, $p = 0.923$). When children under the age of 6 were excluded from the analysis, once again the role of age in determining DBH activity was found to be non-significant ($r^2 = 0.0004$, $p = 0.922$). The results from the sample with autism described here have been compared to a previous study of DBH activity in a normal child population (aged between 6 and 12). This test indicates that DBH activity in this control sample is significantly lower than the activity in the sample of Irish probands with autism (see Table 4.3). The study by Weinshilboum reported an average DBH activity of 34.4nmol/min/ml in a sample of 317 control children aged between 6 and 12 years of age ³⁷³. The t-

test result indicate that this mean decrease of 12.53 nmol/min/ml was significant (T statistic = 5.308, p value = 1.97×10^{-8}) (see Table 4.3).

4.4.4 DBH and Epilepsy

Norepinephrine has been shown to play an important role in the control of seizures (see Section 4.1.3). Attempts were made to test if the affected individuals in the Irish sample who also had a diagnosis of epilepsy had lower DBH activity in comparison to the remaining non-epileptic sample with autism, and the normal child control sample described previously. Unfortunately, the number of individuals with epilepsy (n = 5) was too few, and the power of this test was not strong enough to provide conclusive evidence. Further tests in a larger sample would be required to test this hypothesis.

4.4.4 Transmission of Allelic Variants

Neither TDT of single allele nor haplotypes, using UNPHASED, revealed any significant distortions in the transmission of particular alleles to affected probands. This was a surprising result. Given that the children had significantly lower DBH activity than their parents (see Table 4.2), one would have expected that transmission of ‘lower activity alleles’ would be in excess of ‘higher activity alleles’. This was shown not to be the case. Similarly, there was no excess transmission of ‘low activity haplotypes’ to affected individuals. These results suggest that the lower levels of DBH in the children with autism are not due to the possession of low activity alleles. Regression analysis was undertaken to determine the role that each variant plays in determining the activity of DBH, in both parents and children.

4.4.5 Regression Analysis in Parents and Children

A previous study by Zabetian et al identified the C-1021T variant and reported that it accounted for between 35% and 52% of the variation of DBH activity³⁹⁹. The T allele predicts low DBH activity, with homozygotes having the lowest DBH activity. C homozygotes exhibit the high DBH activity, while heterozygotes show an intermediate DBH activity. Our data show a similar finding, with CC individuals having a higher activity than either CT or TT individuals, and TT individuals having the lowest activity of all (see Figure 4.8). Kruskal-Wallis analysis revealed that these differences were significant ($\chi^2 = 26.156$, $p = 2.09 \times 10^{-6}$). Likewise, differences between allelic variants were found to be significant in the In/Del, C-2124T and Exon 2 variants (see Table 4.5 for further details).

Regression analysis in parents indicated that the C-1021T allele accounted for ~41% of DBH activity variation. This was consistent with the findings of Zabetian et al³⁹⁹. Some of the remaining variants, which did not account for any variation in Zabetians's population, were shown to account for DBH activity levels ranging from 0.5% to 6.2% in our sample of parents (see Table 4.6 for more information). These differences may have arisen due to allele-specific effects in the Irish population or by chance.

Regression analysis in children showed that these variants played a significantly smaller role in determining DBH activity in children. Instead of the C-1021T explaining 41% of variation as in adults, in children it was shown only to contribute 5%, and this did not reach statistical significance. All variants, with the exception of exon 11, were shown to contribute less in children than they did in their parents (see Table 4.7). However, the analysis of only two variants, exon 11 and exon 12, reached the significance threshold. The exon 11 variant was shown to contribute 17% of the variation. This finding is interesting, as another yet unpublished study from Joseph Cubells, found that this same variant was responsible for a similar level of variation in a population of individuals suffering bipolar depression, where once

again the *DBH* C-1021T variation failed to account for the differences in DBH activity (personal communication). Once again however, TDT analysis did not reveal any preferential transmission of exon 11 alleles to affected children.

4.4.6 Possible Reasons for Biochemical/Genetic Discrepancy

There could be a number of reasons for the observed differences in enzyme activity without accompanying genetic evidence of allelic transmission distortion mentioned above. Firstly, there could be a mutation elsewhere in the gene, which is having a major negative effect on DBH activity within the proband sample. However it would be expected that the mutation causing such an effect would be present in either or both of the parents, where it should also exert a decreased effect on DBH activity. Although a control sample is not presently available to test if parental DBH activity is lower or higher than normal, comparison with other studies indicate that our parental DBH activity is very similar to other normal control samples reported^{393,398}. Also, one would expect that this 'mutation' would be in at least weak LD with one of the other markers spanning the gene, and thus might have been detected by TDT testing of the markers included in this study.

Another possibility is that there is a recessive mutation/variation in another gene in the children, which may have the downstream effect of reducing *DBH* transcription. That is to say, that epistasis is taking place. Potential candidates for this role include the *Arix* gene, a transcription factor for *DBH*, or another gene encoding proteins in the cAMP-dependent protein kinase (PKA) pathway that leads to changes in *Arix* phosphorylation and hence *DBH* transcription. A number of protein binding sites exist in very close proximity to the *DBH* transcription start site. One site, located at -58 to -40, matches 8 of the 9 consensus base-pairs for a general transcription factor called Sp1⁴⁰⁰. This transcription factor has been shown to be crucial in maintaining *DBH* promoter activity, by interacting with numerous other proteins⁴⁰¹. Another

domain, matches the AP2, another common transcription factor, consensus site. Another two domains in the promoter region have been identified as positive norepinephrine-specific *cis*-elements, which act to recruit two proteins, Phox2a (a.k.a. Arix) and Phox2b (a.k.a. NBPhox). It is necessary to have simultaneous occupation of both of these sites to initiate *DBH* transcription⁴⁰⁰. Phox2a activity is controlled via its phosphorylation status, which in turn is controlled by the PKA pathway. The DNA binding ability of Phox2a is greatly increased by dephosphorylation of the protein, in response to stimulation of the PKA pathway. All four domain-binding sites need to be activated for transcription of the *DBH* gene to take place. Future work may include association testing of variants within this gene.

Another possibility is an environmental stimulus or developmental switch, affecting DBH activity or transcription activity, which could be altered in the affected children. DBH activity increases significantly in the first three years of life³⁹⁶. The stimulus / developmental switch causing this rapid increase in DBH activity is yet unknown. It is possible that this 'switch' is not triggered in individuals with autism, and therefore their DBH activity remains at their infant levels. This theory would be supported by the finding of decreased heritability of the DBH activity in the sample of Irish individuals with autism in comparison to previous studies^{8,9}. All of the aforementioned hypotheses are worthy of further investigation.

The finding of decreased DBH levels in autistic probands is important. Noradrenaline has been identified as a modulator of seizures, with DBH *-/-* mice exhibiting increased susceptibility to seizure inducing stimuli^{382 384 387}. In humans, lower DBH activity may also lead to an increased risk of seizures, although this remains to be demonstrated. However it may help explain why individuals with autism are more likely than individuals from the general population to be diagnosed with epilepsy³⁷⁶. Furthermore, DBH knockout mice have shown deficits in maternal behaviour, e.g. nest building, pup retrieval and nursing etc³⁸⁹. They also show deficits in motor function, learning and memory⁴⁰². Thus we can speculate that

DBH may affect social and emotional interaction, which may contribute to the autistic phenotype.

4.4.7 Conclusions

A number of conclusions have been reached in this study. Firstly children with autism have lower levels of DBH activity compared to their parents and a published normal child sample. Secondly there is no increase in DBH activity from the ages of 47 months upwards. Thirdly, the study has replicated the finding that there are no sex differences between men and women in terms of DBH activity.

The reason for this lower DBH activity remains unclear however. There was no preferential transmission of low activity single alleles or haplotypes to affected individuals. In parents at least, the C-1021T marker was shown to account for 40.8% of DBH activity, which was comparable to Zabetian's finding (where in his population the C-1021T allele accounted for 35% to 52% of variation), with the remaining markers accounting for less variation. In children however, the same marker only explains 5% of the variation in DBH activity. This result indicates that there are other over-riding influences affecting DBH activity.

CHAPTER 5

INTEGRIN- α 4 AND AUTISM

5.1 Introduction

5.1.1 Chromosome 2q and Genome Scans

As discussed in the introduction (section 1.3.1), genomewide linkage studies are one method of highlighting chromosomal regions of interest that may contain genes important in elucidating the role of genetics in autism. Linkage to chromosome 2q has been reported in a number of studies. Findings by the International Molecular Genetics Consortium in Autism (IMGSAC) detected a Maximum LOD score (MLS) of 4.8 at marker D2S2188, located at 180.79cM⁸¹. A follow-up study resulting from the IMGSAC study also found evidence for linkage in the chromosome 2q using markers D2S2314 (at 188.9cM) and D2S2310 (at 191.8cM) resulting in a maximum multipoint MLS score of 2.54⁸⁸. Buxbaum et al also reported linkage in this region, between markers D2S335 and D2S364, located at 175.91cM and 186.21cM respectively on chromosome 2q⁴⁰³. Interestingly, restricting analysis to individuals with phrase speech delay further increased the significance of these findings. Another study, by Shao et al, described evidence for linkage in an independent sample telomeric to the two previously described studies, at the marker D2S116 at 198.65cM⁴⁰⁴. The statistical significance of this finding also increased with restriction of analysis to probands with phrase speech delay. The linkage ‘peaks’ resulting from such studies, while highlighting regions of interest may be broad and harbour a large number of genes. For the purposes of positional cloning there is large variation between linkage peaks relative to the physical location of disease genes⁴⁰⁵. Also, if the peaks of the linkage study are not very significant (MLOD < 3.0), they can also be displaced by up to 40cM⁴⁰⁶. The overlapping region of interest produced from the peaks observed in the above three studies is 22.74cM. This represents a challenge to identify the gene or genes associated with autism. A number of techniques are possible. Firstly, fine mapping and further linkage analysis of the region may refine the region, and hence reduce the number of candidate genes to be investigated by association testing. Also, reports of cytogenetic abnormalities in

patients presenting with the disorder close to or within these linkage peaks may also aid in narrowing the region of interest. The deletion in patient KM lies central to a number of peaks, which cluster around the proximal and distal ends of the deletion (see below). This could mean that there is a single gene in between these peaks, or that there are two genes, one close to the proximal end of the deletion, with another closer to the distal end of the deletion.

4.1.2 Cytogenetic Study in Irish Patient Suffering Autism

A previous study conducted in this laboratory in collaboration with the National Centre for Medical Genetics, Our Lady's Hospital, Crumlin, investigated a region on chromosome 2q that was defined by the presence of a cytogenetic abnormality in a patient with autism. This region occurred within the putative linkage region of chromosome 2q highlighted by the aforementioned genome scans⁴⁰⁷.

Patient KM was a fourteen-year old male with difficulties in communicating, reciprocal social interaction and behaviour. He met criteria for autism at age five years with the Autism Diagnostic Interview (Revised) (ADI-R) and assessment with the Autism Diagnostic Observation Schedule-Generic (ADOS-G) gave a diagnosis currently of autism spectrum disorder. Routine cytogenetic testing, conducted as an exclusion criterion for the study, revealed the presence of a *de novo* abnormality involving chromosome 2q32. Fluorescent In-Situ Hybridisation (FISH) analysis revealed the presence of a translocation of chromosome 2q32.3 DNA into chromosome 9q31 ((46, XY, ins(9;2)(q31.1;q31.2q31.3))) (see Figures 5.1 and 5.2).

There had been a number of previous reports of abnormalities involving chromosome 2q, including 2 studies reporting deletions 2q37 in association with autism⁴⁰⁸, 5 studies reporting deletions of 2q31.33^{409 410 411 412 413}. A duplication of the 2q31 region has also been reported. Neither the reports of duplications or deletions within the region commented on the presence of autism.⁴¹⁴ The KM case

appeared to be the first case of a chromosomal rearrangement in this region of chromosome 2q associated with autism. The section of chromosome 2q involved in this deletion translocation appears to be located within the linked regions reported in the aforementioned linkage studies.

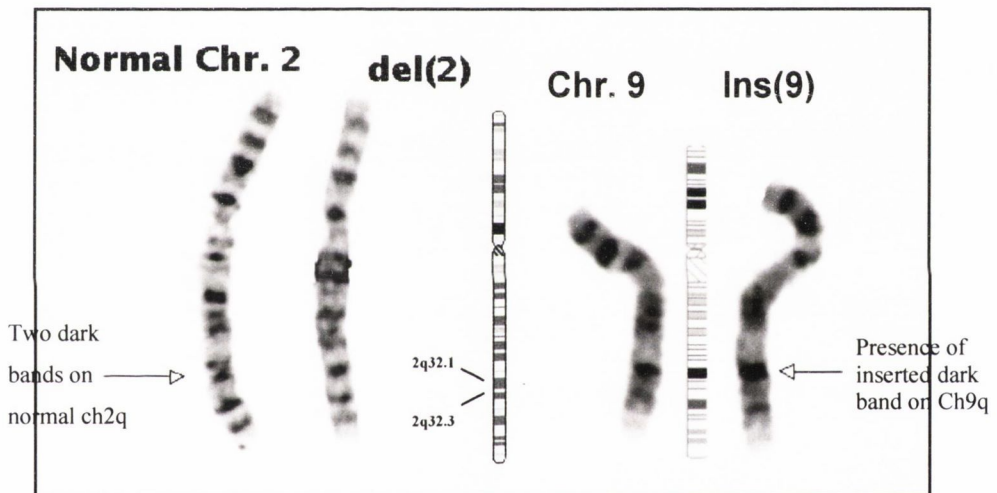


Figure 5.1 CYTOGENETIC SPREAD SHOWING THE PRESENCE OF TRANSLOCATED CHROMOSOME 2q32.1-32.2 ON CHROMOSOME 9q31.1-31.3 (Picture courtesy of Dr. Sean Ennis, National Centre for Medical Genetics).

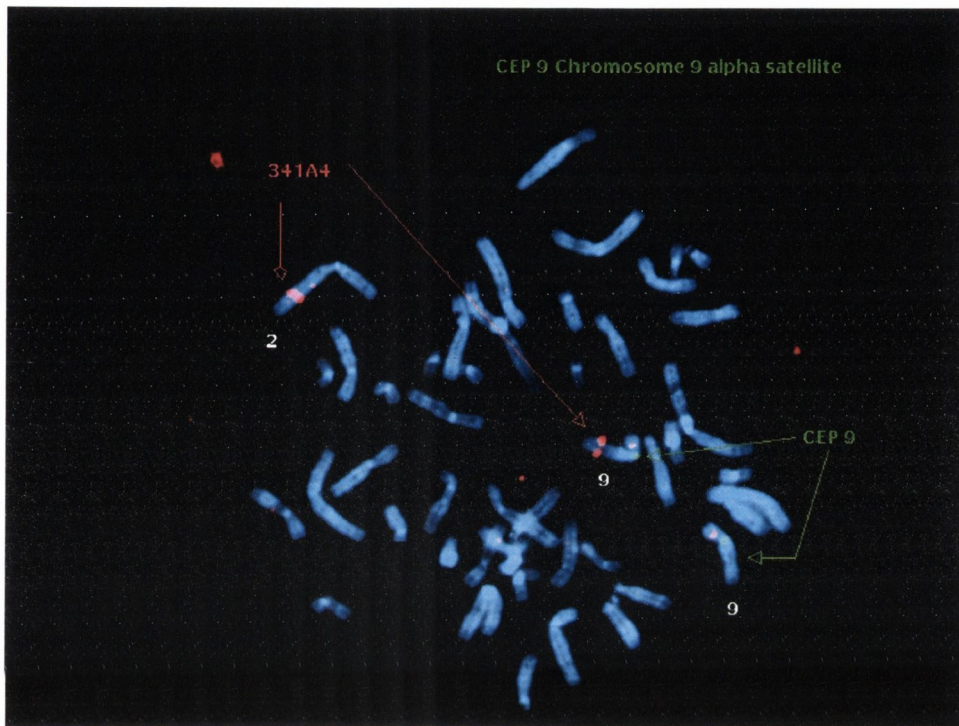


Figure 5.2 FLUORESCENT IN SITU HYBRIDIZATION (FISH) SHOWING THE PRESENCE OF A CHROMOSOME 2q PROBE (341A4) TRANSLOCATED TO CHROMOSOME 9q. The translocation of chromosome 2 DNA (identified by the red probe) to chromosome 9 (identified by the green probe) can be seen by the co-presence of both a red and a green probe on chromosome 9 (Picture courtesy of Dr. Sean Ennis, National Centre for Medical Genetics).

5.1.3 Molecular Studies Involving Chromosome 2q32

A number of previous studies have been undertaken to identify a candidate gene / candidate genes present in the region of 2q32 translocated to 9q31. A linkage disequilibrium screen was undertaken across the region defined by the cytogenetic abnormality. Fifteen microsatellite markers were genotyped across the region. Two markers, D2S2077 and D2S270, showed evidence for increased transmission in the autism sample. One marker, D2S2077, revealed evidence of significant association with autism ($\chi^2=5.8$, $p=0.013$, OR = 1.75). This bi-allelic marker is located close to the centromeric end of the mapped region.

The region around this marker was flagged for follow-up analysis, and was mapped with 18 SNPs with inter-marker distances between 30kb and 40kb (see Figure 5.3). No significant associations were detected with any of the individual SNPs using TDT. One SNP, SNP4 (rs155149), showed a non-significant trend towards increased transmission of the C allele to affected probands ($\chi^2=3.3$, $p=0.07$). Association testing using the haplotype based haplotype relative risk test (HHRR), revealed that transmission of the C allele of rs155149 was statistically significant ($\chi^2=5.16$, $p=0.023$, RR=1.48, $1.01 < RR < 2.16$). Transmissions for SNP 26, rs976384, and SNP 3, rs2035207, also showed a non-significant trend towards increased transmission in the autism sample (SNP 26: $\chi^2=2.7$, $p=0.1$, OR=1.47, SNP 3: $\chi^2=2.9$, $p=0.09$, OR=1.63).

A number of haplotypes showed association with autism significant at the $p < 0.05$ level. They included both bi-allelic microsatellites and SNPs. Many of the associations centred around those SNPs at the proximal end of the region mapped. A haplotype consisting of two microsatellites (D2S2310 – D2S2077) was found to be statistically associated with autism, producing a χ^2 of 24.98 and p value of 0.035. Testing the significance using a bootstrap method reduced this p value to 0.26. Other

haplotypes containing the D2S2077 microsatellite (148bp allele) were also significantly over-transmitted to affected probands. They included a haplotype with SNP 26 (bootstrap p value = 0.043), SNP3 – SNP4 (bootstrap p value = 0.01) and SNP4 – SNP6 (bootstrap p value = 0.02). Two and three marker haplotypes containing SNP4 and/or SNP6 were also shown to be over-transmitted.

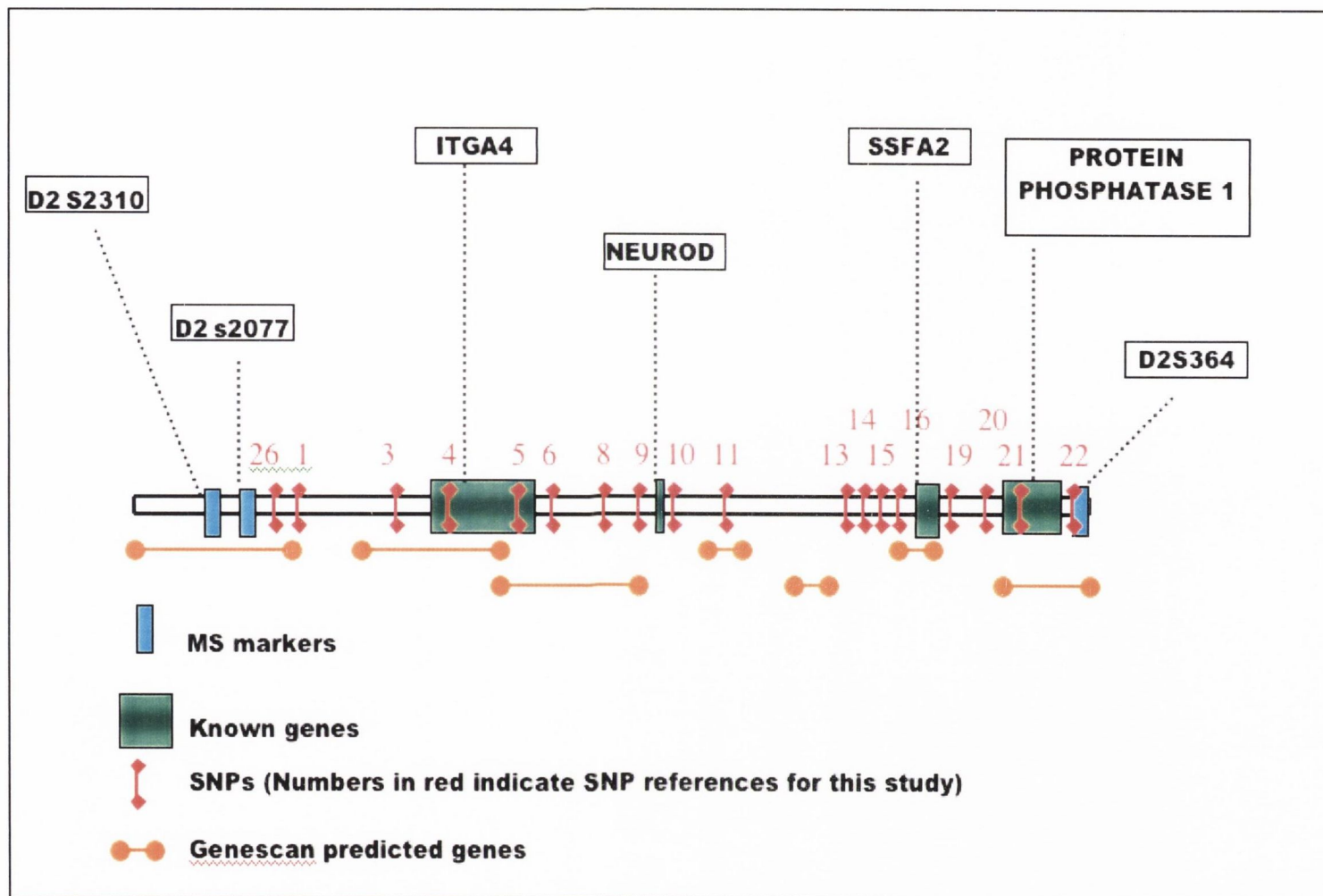


Figure 5.3 Location Of Initial Screening Markers And Exons.

As mentioned previously, the majority of markers used in association and mapping studies are not functional. However, those markers showing association, or trends towards association, may be in LD with functional markers / disease-causing variants, which remain as yet unknown. Two genes, Neurogenic Differentiation Factor 1 and Sperm Specific Antigen 2, which were in the region defined by the linkage disequilibrium mapping using SNPs, have already undergone mutation screening using dHPLC methods (unpublished results). No associations with autism were detected. Only two of the SNPs tested were in genes. Both SNPs 4 and 6 (rs155149 and rs921257) were located within intronic regions in integrin alpha 4 (*ITGA4*) SNPs, and were part of a haplotype that were significantly over-transmitted to autistic probands (see Figure 5.3).

5.1.4 The Integrin Alpha 4 (ITGA4) Gene

The *ITGA4* gene encodes one of 16 α and eight β integrin subunits identified to date⁴¹⁵. Integrins are a family of proteins that are membrane spanning, non-covalently bound $\alpha\beta$ heterodimers that act as cell-matrix, cell-cell adhesion receptors and / or cell signalling activators (see Figure 5.4). Cells interact with the extracellular matrix through integrins, binding collagen and fibronectin outside the cell to actin fibres in the cell interior (see Figure 5.4). The ability of cell-substrate interactions is crucial for cell motility. Increased interactions lead to reduced mobility, whereas decreased interactions allow cells to move to new locations. This process of cell migration is especially important during embryonic development as cells migrate to a final anatomical position in advance of final differentiation.

ITGA4 is located on chromosome 2q31.3, and spans 79.7kb. It contains 28 exons and produces a protein of 1038 amino acids, which has a molecular weight of 115.3kDa (from *ITGA4* Genecard database: <http://bioinfo1.weizmann.ac.il/cgi-bin/genecards/carddisp?ITGA4>). Integrins are best known as proteins playing an

important role in maintaining muscle cell structure and their role in immune responses. ITGA4 is a vitronectin receptor, which is involved in vasculogenesis and angiogenesis^{416 417}. It is also expressed in the CNS, where its distribution is limited almost entirely to the limbic telencephalon, especially the hippocampus, cortical amygdala, entorhinal and piriform cortex⁴¹⁵.

There is also increasing evidence that integrins play an important role in brain development and function. The process of neural crest migration occurs early in development, when neural crest cells leave the neural tube. Their interaction with the extracellular matrix controls the migration to their final destination⁴¹⁸. In chickens, ITGA4 mRNA expression is localised to migrating neural crest cells. The addition of blocking antibodies to the ITGA4 subunit causes a significant decrease in both the number of neural crest cells migrating (57–70 %) and the average distance travelled per cell⁴¹⁹. Other members of the integrin family have also been associated with brain development and function, ranging from impaired axonal regeneration (integrin $\alpha 7$ deficient mice)⁴²⁰, neuronal migration (Integrin $\alpha 3\beta 1$ in conjunction with Reelin)⁴²¹, the correct formation of radial glial scaffolding in the hippocampus (Integrin $\beta 1$ with Reelin and Disabled 1)⁴²² to regulation of functional synaptic plasticity processes (in *Volado*, a *Drosophila* integrin protein)⁴²³.

The ITGA4 dependent cell adhesion pathways are critical intervention points in a wide variety of inflammatory conditions. The integrin $\alpha 4\text{--}\beta 1$ complex is also expressed on most leukocytes at sites of inflammation in various organs⁴²⁴. Studies have found that T helper type 1 (Th1) cells expressing high levels of ITGA4 enter brain parenchyma at a much higher rate in comparison to Th1 cells expressing low levels of ITGA4⁴²⁵. The interaction of ITGA4 with vascular cell adhesion molecule-1 (VCAM-1) at the blood-brain barrier is critical for inflammation affecting the CNS⁴²⁶. Inhibition of ITGA4 action through the use of antibodies and anti-sense oligonucleotides have been proven to protect the brain against ischemic injury and ameliorate an animal model of CNS inflammation^{427 426}. Inflammation in the brain, in the form of encephalitis, has shown to be associated with autism^{428,429}. Also

systemic immunological aberrations, both autoimmune and functional (i.e. abnormalities and/or deficits in function of immune cell subsets), have been reported in autism (see a review by Ashwood et al for more information ⁴³⁰).

These links to brain development, synaptic functioning and the specific regional expression of *ITGA4*, in addition to the cytogenetic findings, and genome scans highlighting chromosome 2q31 as an interesting linkage peak, make *ITGA4* an interesting candidate gene for autism.

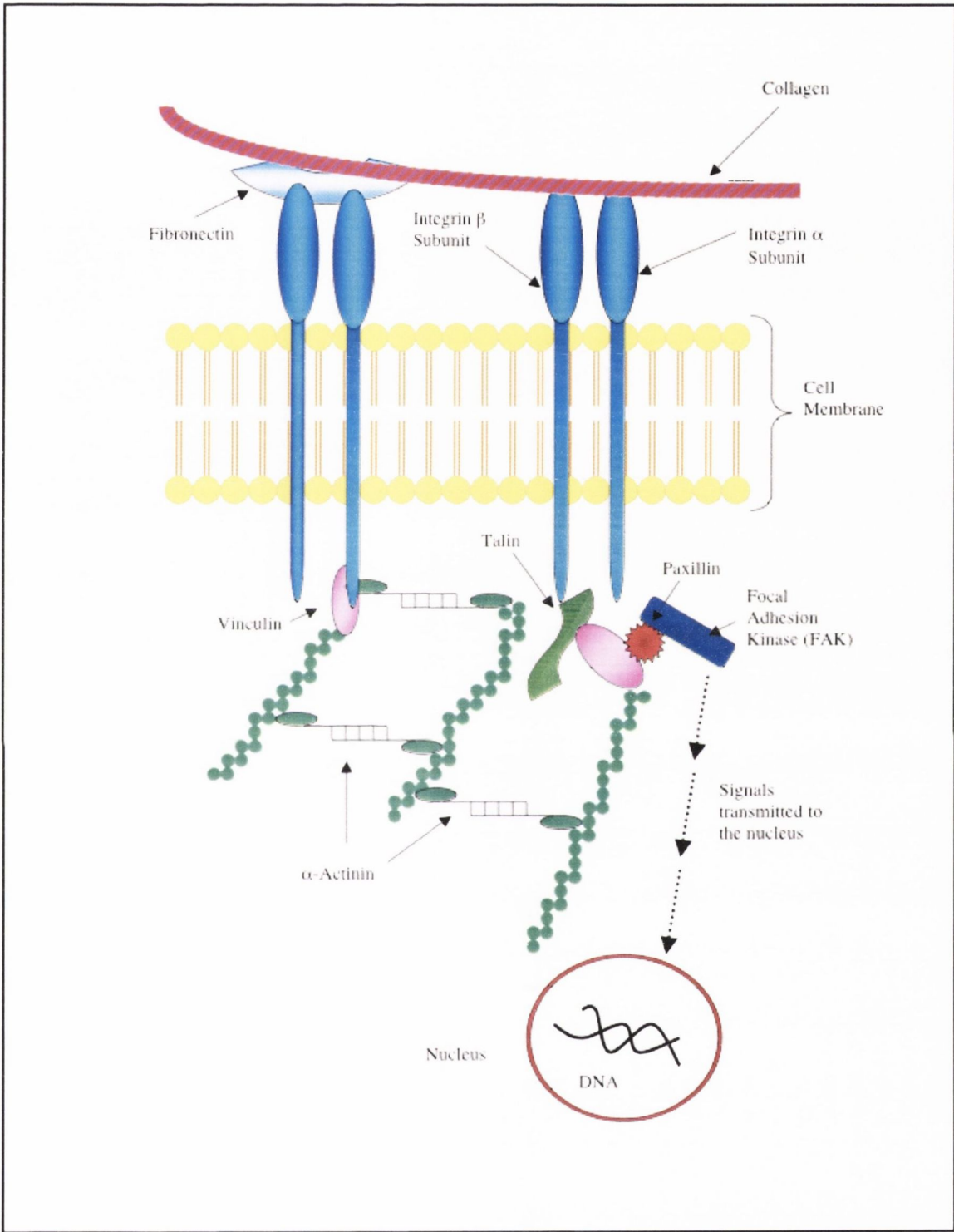


Figure 5.4 THE ROLE OF INTEGRIN PROTEINS IN LINKING THE PROTEINS IN THE EXTRACELLULAR MATRIX TO ACTIN FIBRES IN THE CELL INTERIOR. Adapted from G Karp (Cell and Molecular Biology⁴³¹).

5.1.5 The HapMap Project

When testing for associations there are two methods for finding disease-causing variants. The first method, the direct method, is to test all potentially casual variants for correlation with the disease in question. The second method, the indirect method, uses a subset of variants, each in linkage disequilibrium with its neighbour to screen for association under the assumption that any risk variants lying between the tested markers will themselves be in linkage disequilibrium with their tested neighbours. Using this method, it is not necessary for the markers tested to have any functional significance. The local region surrounding these markers that show association, can then be screened for functional mutations, resulting in a more economic and time efficient method for finding disease causing variants. Due to the size of the *ITGA4* gene and the potential number of SNPs to be typed to fully screen the gene, it was decided to test the usefulness of the HapMap in choosing “tagging SNPs” to find associations for autism.

The aim of the HapMap project (at <http://www.hapmap.org/>) is to “determine the common patterns of DNA sequence variation in human genome, by characterising sequence variants, their frequencies and correlations between them in DNA samples from populations from parts of Africa, Asia and Europe...thus providing the tools that will allow the indirect association approach to be applied readily to any functional candidate gene in the genome, to any region suggested by family-based linkage analysis, or ultimately to the whole genome for scans for disease risk factors”²⁷⁸. Haplotypes arise because when a single mutation occurs, it is associated with the other alleles surrounding it. This specific set of alleles observed on a chromosomal region is called a haplotype. Within this haplotype, it is only necessary to genotype a small number of SNPs, “tagging SNPs”, to provide enough information to predict the majority of other common SNPs in the same haplotype.

5.1.6 Aims

There were a number of aims regarding the testing of the role of the *ITGA4* with autism.

- 1) To fine map the *ITGA4* gene, and test its association with autism.
- 2) To undertake mutation detection in any exon close to markers showing association with autism.
- 3) To attempt replication of any significant findings in an independent sample kindly provided by collaborators from the Vanderbilt University, Tennessee, USA.
- 4) To test the validity of using “tagging” SNPs to find associations with autism, by testing the association of tagging SNP significance in comparison to significant results obtained in tests involving all markers.

5.2 Materials And Methods

5.2.1 Samples

The SNPs selected for genotyping (see Section 1.3.2) were initially genotyped in the Irish sample (see Section 2.1) and an Irish control sample. The genotyping in the control sample was undertaken in order to allow comparison of HapMap selected “tagging SNPs” the *ITGA4* SNPs (with the exception of rs12690517) in 94 control samples. These samples were collected from Irish blood donors. The sample was not specifically screened for psychiatric illness; however, donors were not taking regular prescribed medication as such individuals are excluded from blood donation in Ireland.

As the findings above are presented uncorrected for multiple tests, replication was sought in two independent samples kindly provided by colleagues at Vanderbilt University. The first sample was a sample originating from the Autism Genetic Resource Exchange (AGRE) (<http://www.agre.org/>). This sample was an American sample, and each family unit contained at least two family members with a diagnosis of autism or an Autism Spectrum Disorder. Children were diagnosed using the ADI-R, and had an age of onset of less than 3 years. The AGRE sample described here was comprised of 267 families, with affected sib-pairs. Further information regarding AGRE recruitment and sample description can be found at Liu et al ⁷⁹.

The second replication sample was a sample recruited locally in Tennessee, USA. The sample has been described in detail in McCauley et al ⁸⁷. Briefly it consisted of 102 families with at least one affected proband. All affected individuals were clinically assessed using the ADI or ADI-R. Most were also assessed using the ADOS test. All affected individuals were greater than four years of age, and

probands were excluded in the presence of an associated medical or neurological phenotype for autism, e.g. Fragile X syndrome.

5.2.2 Selection and Genotyping of SNPs

ITGA4 markers were selected from the publicly available genetic SNP database (dbSNP) at NCBI (<http://www.ncbi.nlm.nih.gov/SNP/>). SNPs were selected based on their location (intronic / exonic), known heterozygosity, inclusion in the previous chromosome 2 study (i.e. the SNP4 (rs155149) and SNP6 (rs921257) in the original mapping experiment, see Section 5.1.3, were included) and inter-marker spacing distance.

Details of the SNPs and surrounding sequences, with the exception of rs12690517, which was detected by mutation analysis following the SNP genotyping, were sent to a commercial genotyping company, K-Bioscience (<http://www.kbioscience.co.uk/>) where the Amplifluor™ allele detection (Serologicals Corp., Temecula, California) technique was used. One SNP, rs12690517, was genotyped using the SNaPshot method of allele discrimination (see Section 2.2.4.3 and Appendix II, sections 1.1 and 1.2, for further details).

5.2.3 Mutation Detection

Following preliminary findings, exons 16 and 17 were selected for mutation screening using the dHPLC method (see Sections 1.4.2 and 2.2.5.2 for details). Primers were designed to span the exons and include exon/intron boundaries. Details of primer sequences and amplification modifications can be found in Appendix II, sections 1.1 and 1.2.

5.2.4 Statistics

Allele and genotype frequencies were tested for Hardy-Weinberg equilibrium. Departure from expected allelic and haplotype transmissions were tested using the TDT and UNPHASED tests respectively (see Sections 2.3.1.2 and 2.3.1.3.2 respectively). LD was measured by the D' statistic in the control sample, the Irish parental sample, the AGRE sample, and the Vanderbilt sample using the Haploview programme (Section 2.3.2.3).

Tagging SNPs were also chosen in the control sample using Haploview Programme (see Section 2.3.2.9).

For the AGRE and Vanderbilt samples departure from expected frequencies of allelic transmission was measured using the PDT test.

5.3 Results

5.3.1 Selection of SNPs

Three hundred and eighty four SNPs were retrieved using the broad search term, integrin alpha 4, in the NCBI SNP search engine. Each SNP was assessed for genotyping based on location, heterozygosity, inter-marker distance and the population data sample size. After all markers with less than 20% heterozygosity were excluded, 27 SNPs remained, see Table 5.1. There were no exonic markers with known heterozygosity in the database at this stage. An average inter marker distance of <5kb was required. A minimum assay sample size of > 15 was set as a cut-off point for exclusion. There was one exception for this. One SNP, rs1551031, with an assay sample size of 12 was included, in order to map the 11.7kb gap between rs3770112 and rs2305581. Of these 27 SNPs, 18 were selected based on inter-marker distance for genotyping (see Table 5.1). Sequences for all chosen SNPs were sent to K-Bioscience for genotyping. Unfortunately, assay design failure occurred for SNP rs3770106, i.e the inability for a reliable genotyping assay due to the high CT sequence surrounding this SNP, and as a result this it was not genotyped. The average inter-marker distance was 4.4kb.

rs number	location	snp	Population validation number	heterozygosity	Location	Distance from previous marker	Selected	Distance from proximal marker
rs1449263	1.9 kb 5' to gene	a/g	15	0.498	32528717		*	
rs3770137	intron 2	g/c	42	0.365	32535331	6614	*	6614
rs3770136	Intron 2	a/g	48	0.355	32538831	3500	*	3500
rs155149	intron 2	a/c	65	0.477	32539021	190	*	190
rs155148	intron 2	c/t	67	0.272	32540711	1690	*	1690
rs1449260	intron 2	g/a	63	0.486	32542831	2120	*	2120
rs1449264	intron 2	c/t	60	0.383	32545890	3059		
rs3770126	intron 2	g/t	42	0.475	32546831	941	*	4000
rs199095	Intron 5	t/c	7	0.31	32552693	5862		
rs155096	Intron 6	a/g	17	0.632	32554771	2078	*	2078
rs2305590	Intron 7	a/g	48	0.234	32556071	1300		
rs155100	Intron 8	a/t	19	0.489	32559471	3400	*	4700
rs155107	intron 9	c/t	65	0.459	32561171	1700		
rs3770117	intron 10	g/a	48	0.139	32562331	1160	*	1160
rs3770116	intron 15	c/a	48	0.353	32574071	11740	*	11740
rs3770115	intron 15	c/t	48	0.284	32574131	60		
rs1349197	intron 15	c/g	60	0.477	32581531	7400	*	7460
rs3816521	Intron 15	a/t	48	0.477	32583743	2212		
rs3770112	intron 17	c/t	50	0.291	32586951	3208	*	3208
rs1551031	intron 17	a/g	12	0.485	32591011	4060	*	4060
rs2305581	Intron 20	a/g	48	0.371	32598681	7670	*	7670
rs3770108	Intron 20	g/a	48	0.488	32598961	280		
rs1375490	intron 21	g/t	10	0.451	32599861	900		
rs3770106	intron 21	t/c	48	0.364	32600741	880	*	2060
rs3770105	intron 22	a/g	52	0.292	32603261	2520	*	2520
rs921257	intron 23	c/t	60	0.306	32603961	700	*	700
rs2290517	2.4 kb 3' to gene	a/c	35	0.355	32612861	8900	*	8900

Table 5.1 SELECTION OF SNPs WITHIN AND SURROUNDING THE *ITGA4* GENE. All SNPs chosen for genotyping are indicated with a * in the “Selected” column. Although rs3770106 was selected to genotype, an assay design failure occurred, and therefore this marker was not genotyped. The average inter-marker distance is 4.4kb.

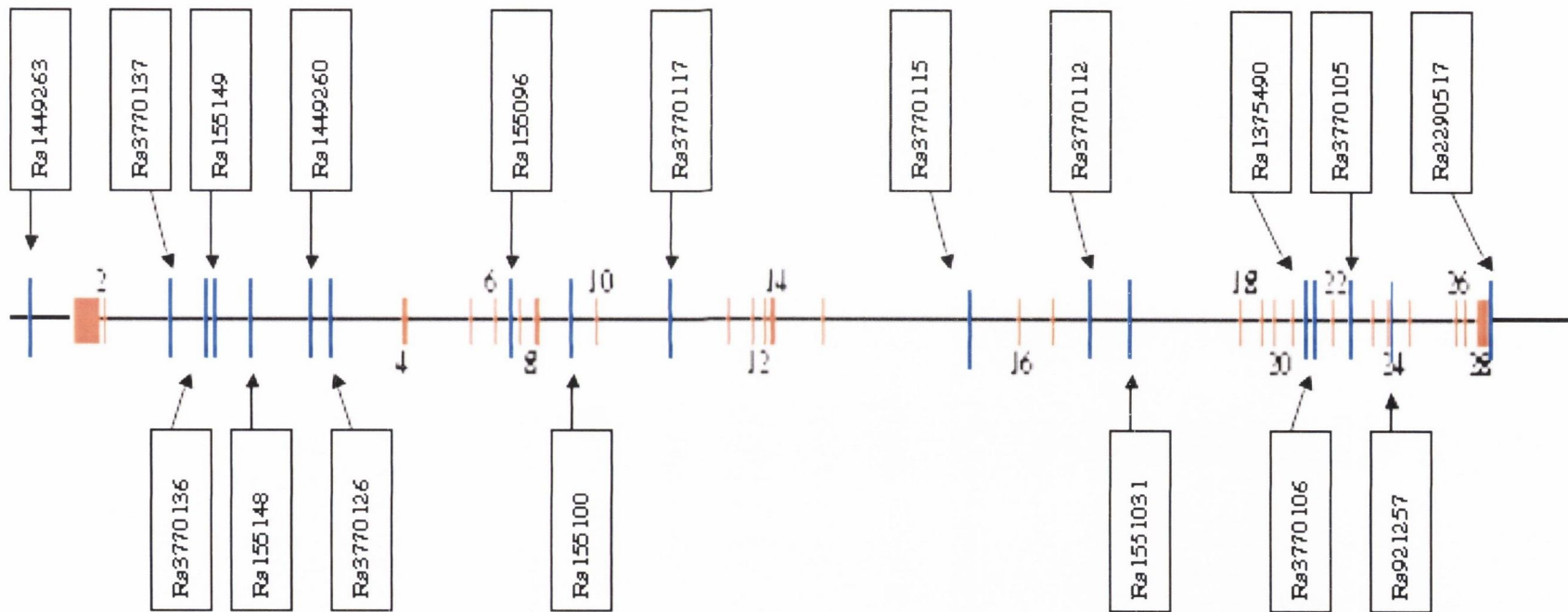


Figure 5.5 LOCATION OF EXONS AND MARKERS IN THE *ITGA4* GENE. All markers genotyped, with the exception of rs3770106, and the relative position of exons in the *ITGA4* gene. Marker rs3770106 was not genotyped due to an assay design failure. The map of exons is adapted from the Ensembl database (<http://www.ensembl.org>).

5.3.2 LD structure in ITGA4 Gene

LD was measured using the HAPLOVIEW programme. LD was analysed in three samples, our parental data, our control data and the HapMap sample (downloadable from: <http://www.hapmap.org/>). The level of variation between all three samples was minimal; see Figures 5.6, 5.7 and 5.8 for more detail. All three samples show a major haploblock in the 5' region of the gene and another major haploblock at the 3' end of the gene. The parental sample and the HapMap sample also show a minor block in the centre of the gene. The parental and HapMap figures are not directly comparable due to the differing markers used to span the gene. The HapMap sample also spans an area with greater 5' coverage, whereas, our samples (both control and parental samples) have an increased coverage of the 3' region of the gene.

In both the Irish and HapMap samples, there is a 14kb block in the 5' region of the gene, in our samples this region spans from ~1.9kb 5' of the gene (rs1449263), to intron 2 (rs1449260) (see Figure 5.5). A break in LD then occurs, after which, there is another small haploblock covering intron 6 (rs155096) to intron 9 (rs155100) (Block 2 in the Parental and Control Samples, Block 3 in the HapMap Sample) (see Figure 5.5). Finally there is a very long block, which covers the distance from intron 10 (rs3770117) to the 3' region of the gene (rs2290517) in the Irish sample. This distance spans over 50kb.

Overall the most information about LD structure came from the parental sample, due to the increased number of individuals in this sample (328 individuals, versus 94 in the control sample and 30 in the HapMap CEPH sample). There were only 5 LD comparisons with a D' value of 1 and a LOD score of <2 (these comparisons are indicated by the presence of blue blocks), i.e. 2.9% of comparisons. In the control sample there were 41 such comparisons (out of a total of 171, i.e. 24%), and in the HapMap sample there were 269 such comparisons (out of a total of 1830, i.e. 14.7% of total comparisons). This result indicates the importance of using a sample with a

sample size large enough to accurately measure the LD relationships between markers.

Overall there is a high level of LD existing in defined regions of the gene. The largest block spans 50kb from intron 10 to the 3' end of the gene (parental sample, block 3). This block is separated from the remaining blocks by two regions showing disruption in the high LD. These regions, where ancestral recombination is likely to have occurred, lie between intron 2 and intron 6, and between intron 9 and intron 10. Further genotyping would be required to further define these boundaries.

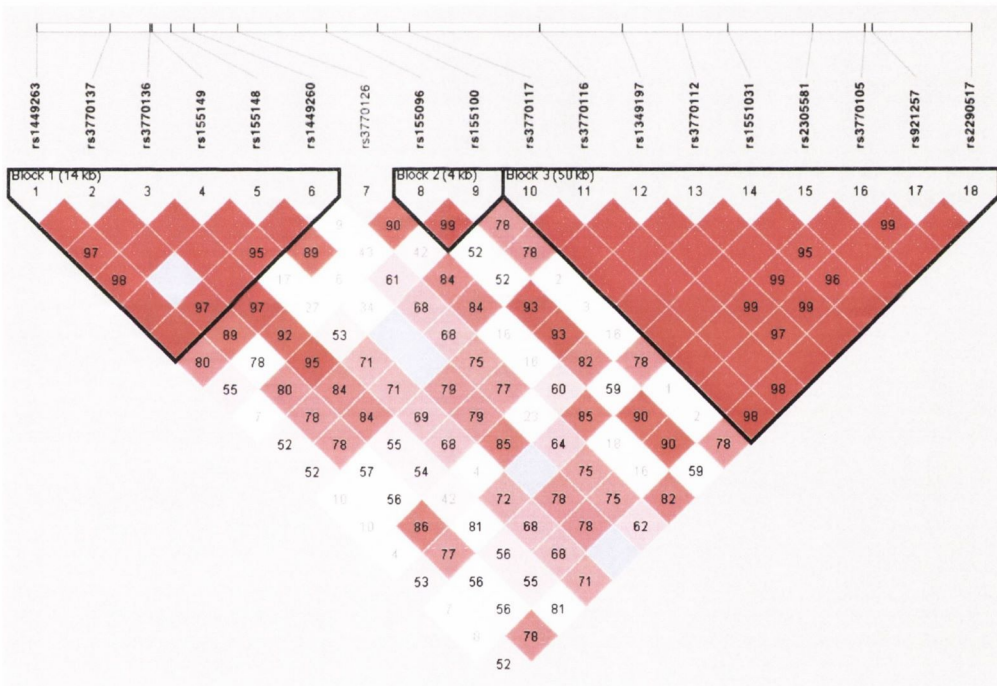


Figure 5.6 LD STRUCTURE IN PARENTAL SAMPLE. LD is measured using the D' statistic and is calculated using the Haploview Programme. This sample consisted of the parental individuals of the Irish Autism Trio Study. If the $D' = 1$ and the $LOD > 2$, the block is bright red. If the $D' = 1$ and the $LOD < 2$, then the block is blue in colour. All other D' values are represented by decreasing shades of red (where $LOD > 2$) or white (where $LOD < 2$).

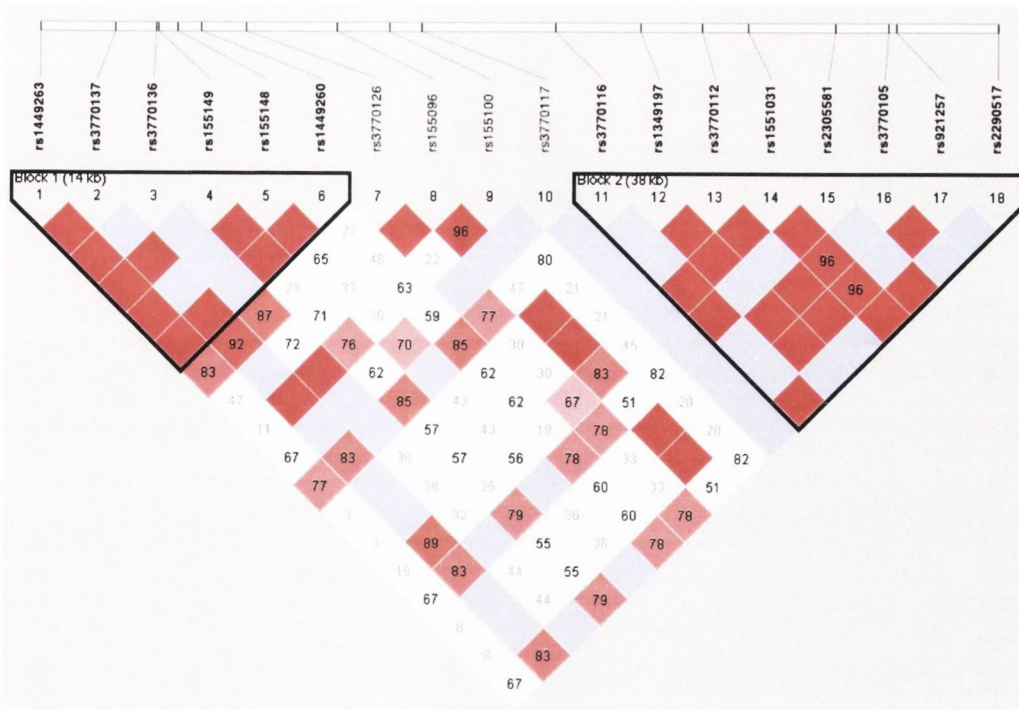


Figure 5.7 LD STRUCTURE IN CONTROL SAMPLE. This sample consisted of 94 normal healthy blood donors. LD structure is similar to that in the parental sample.

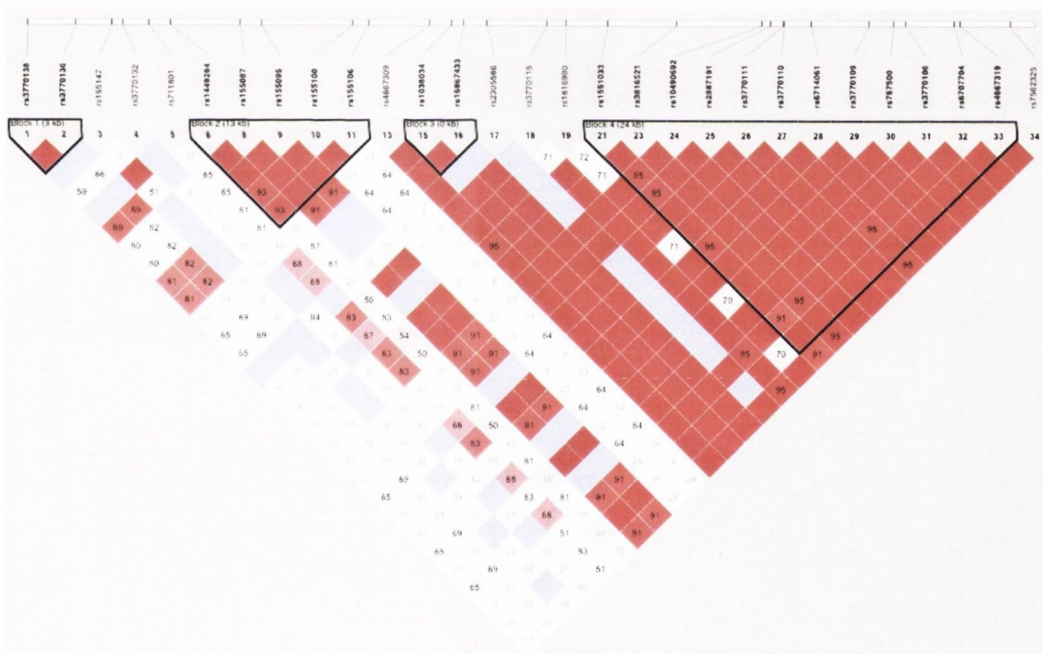


Figure 5.8 LD STRUCTURE FROM HAPMAP DATA. This sample data was downloaded from the HapMap Website (<http://www.hapmap.org/>). The sample consists of 30 CEPH families. This figure is not directly comparable to the two preceding figures, due to the use of different markers, and a greater coverage 5' and 3' to the gene.

5.3.3 TDT analysis

All SNPs were tested and found to be in Hardy-Weinberg equilibrium. TDT testing was performed on all SNPs, see Table 5.2 for a summary of results. The SNP that was 1,912bp 5' of the *ITGA4* gene showed evidence for association with autism, with the G allele being over transmitted to affected probands (86 versus 58 transmissions). This produced a χ^2 value of 5.479, an OR of 1.48 and p value of 0.0192. Another SNP located 1,040bp 3' of exon 17, rs3770112, showed stronger statistical evidence for association, producing a χ^2 of 7.809, OR of 1.73 and p value of 0.0051. Probands with autism received the C allele of this SNP significantly more often than the T allele (69 versus 40 transmissions). The two SNPs surrounding this SNP showed trends towards significance. Over-transmission of the C allele of rs1349197 (66 versus 49 transmissions), which is 2,232bp 5' of exon 16, resulted in a p value of 0.1122. Another SNP in intron 17, rs1551031, also showed distorted transmission, leading to over-transmission of the A allele, giving a χ^2 of 3.457 and p value of 0.063. The remaining SNPs did not show any associations with autism (see Table 5.2 for further details).

SNP number	SNP ID number	ALLELE 1	ALLELE 2	Allele transmitted		CHI	OR	P VALUE
				1	2			
				1	rs1449263			
2	rs3770137	C	G	28	19	1.734	1.47	0.1879
3	rs3770136	G	A	37	27	1.569	1.37	0.2104
4	rs155149	C	A	59	45	1.89	1.31	0.1692
5	rs155148	G	A	32	35	0.134	1.09	0.714
6	rs1449260	G	A	66	53	1.423	1.25	0.2329
7	rs3770126	T	C	54	41	1.785	1.32	0.1816
8	rs155096	A	G	51	51	0	1	1
9	rs155100	T	A	63	59	0.1312	1.07	0.7172
10	rs3770117	G	A	20	17	0.2435	1.18	0.6217
11	rs3770116	A	C	32	26	0.6218	1.23	0.4304
12	rs1349197	G	C	49	66	2.522	1.35	0.1122
13	rs3770112	T	C	40	69	7.809	1.73	0.0051
14	rs1551031	G	A	59	81	3.457	1.37	0.063
15	rs2305581	A	G	31	25	0.6441	1.24	0.4222
16	rs3770105	G	A	56	75	2.756	1.34	0.097
17	rs9211257	T	C	52	72	3.24	1.39	0.0719
18	rs2290517	C	A	27	24	0.1766	1.13	0.6743

Table 5.2 TDT RESULTS FOR INDIVIDUAL SNPs. Those SNPs, which show significant over transmission, are highlighted in yellow and in bold. SNPS that show a trend towards significance (i.e. $0.5 < p < 0.1$) are highlighted in light yellow. SNP rs3770112 shows very significant over transmission of the C allele to probands suffering autism ($\chi^2 = 7.809$, $p = 0.0051$).

5.3.4 Haplotype Analysis

The UNPHASED test (see Section 2.3.1.3.2) was used to construct 2-marker haplotypes and test for distortions in their haplotype transmissions. Transmission of two of these was significantly higher than expected. Both of these haplotypes contain rs3770112 (SNP13), which was the SNP that was significant in the TDT test. Haplotype 12-13 (i.e. rs1349197 – rs3770112) spans exons 16 and 17 of the *ITGA4* gene, and the haplotype containing the C-C alleles was significantly over-transmitted (56 transmissions versus 32 non-transmissions). This haplotype produced a χ^2 value of 7.022, OR of 1.75 and p value of ~0.03. The haplotype distal to this, containing rs3770112 and rs1551031, was also significantly over-transmitted. Transmission of 50 C-A haplotypes compared to 26 non-transmissions to probands produced a χ^2 value of 14.29. The p value and OR associated with this haplotype is very significant (p value = 0.0025, OR = 1.92). No other haplotypes showed any significant association (see Table 5.3 for more details). However, a haplotype containing the first two SNPs, rs1449263 (located 5' of the gene) and rs3770137 (located in intron 2), showed a trend towards significance, giving a p-value of 0.0675.

SNPs	Haplotype Over-transmitted	Transmitted	Not Transmitted	Global Chi square	OR	Global p value
1_2	A-G	65	39	5.3887	1.67	0.0675864
2_3	G_G	48	36	1.692	1.33	0.4291
3_4	G-C	66	48	2.602	1.38	0.2723
4_5	C-A	51	36	5.97	1.42	0.1131
5_6	A-G	55	42	2.001	1.31	0.3677
6_7	G-T	58	42	3.616	1.38	0.306
7_8	T-A	58	49	2.346	1.18	0.5037
8_9	A-A	50	44	1.715	1.14	0.6337
9_10	A-G	56	52	0.136	1.08	0.9343
10_11	G-C	48	36	1.718	1.33	0.4237
11_12	C-C	66	43	4.282	1.53	0.1176
12_13	C-C	56	32	7.022	1.75	0.02987
13_14	C-A	50	26	14.29	1.92	0.0025
14_15	A-G	58	43	1.852	1.35	0.3692
15_16	G-A	70	45	4.713	1.56	0.09473
16_17	A-C	61	42	3.936	1.45	0.1397
17_18	C-A	63	43	3.252	1.47	0.1967

Table 5.3 TRANSMISSION OF TWO MARKER HAPLOTYPES. SNP numbers correspond to those numbers (and hence SNP ID numbers) reported in Table 5.2. Two marker haplotypes that are significantly over transmitted are highlighted in yellow and in bold. Haplotypes showing a trend towards over-transmission ($0.5 < p < 0.1$) are highlighted in light yellow. Transmission of haplotypes was tested using the UNPHASED programme. In a fashion similar to the TDT results, the most significantly over-transmitted haplotypes are those that contain SNP13 (rs3770112).

5.3.5 Mutation Screening of Exons 16 and 17

TDT and haplotype analysis highlighted the over-transmission of alleles/haplotypes around markers 13 and 14 (rs3770112 and rs1551031). These markers lie very close to exons 16 and 17. Therefore, it was decided to screen these exons for mutations in affected probands using a denaturing high performance liquid chromatography technique (see Sections 1.4.2, 2.2.5.2 and 5.2.3). A number of samples showed chromatogram shifts under the different denaturing conditions. Fifty-two (out of a total of 156 probands screened) showed a chromatogram shift under the denaturing temperature of 51°C (see Figure 5.9). Fifty-nine probands showed chromatogram shifts at 53 °C and 56 °C (see Figures 5.10 and 5.11). No chromatogram shifts were identified in the PCR fragment containing exon 17 (see Figure 5.12).

A random collection of those fragments showing chromatogram shifts (3 shift samples + 1 control sample) at each temperature were forward and reverse sequenced using the ABI BigDye Terminator (v3.0) Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) (see Section 2.2.5.3 for further details).

Sequencing of the samples led to the identification of a G to A transition on the forward strand, and a C to T transition on the reverse strand (see Figure 5.13 for the reverse strand result). This mutation had previously been identified by Heymann et al ⁴³², and is identified by the accession number rs12690517 in the public database dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). The exact position of this SNP is unresolved at present. The dbSNP and UC Santa Cruz Genome Assembly (<http://www.genome.ucsc.edu>) databases identify it as intronic, whereas the Ensembl database (http://www.ensembl.org/Homo_sapiens) identifies it as an exonic SNP, occurring as the last base within exon 16. This is due to disagreement between the databases on the exact splice position. Hence within this thesis it will be referred to as a SNP occurring on the exon 16/intron 16 border.

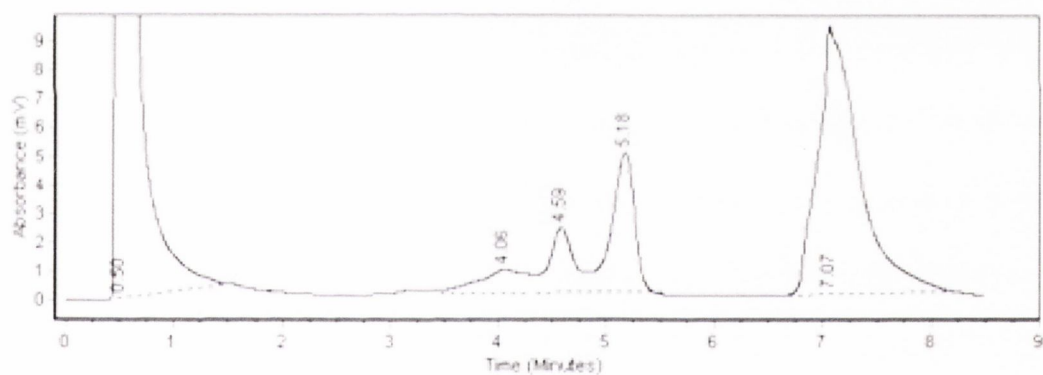
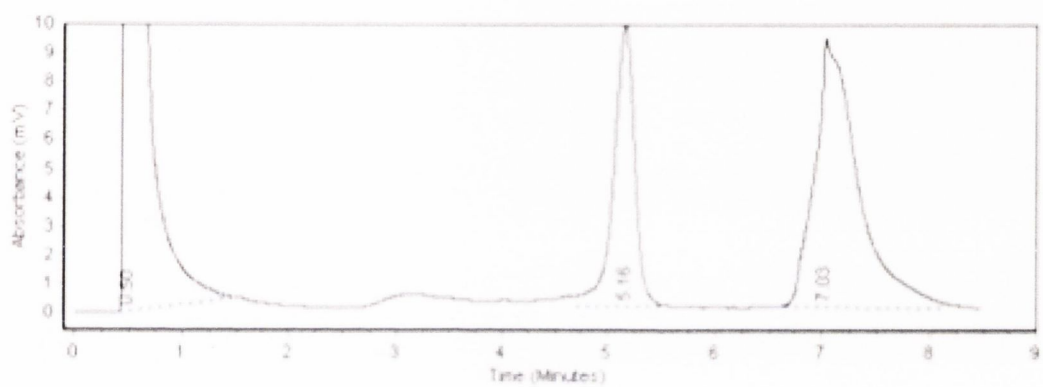


Figure 5.9 dHPLC ANALYSIS OF *ITGA4* EXON 16 AT 51°C. A number of samples showed the presence of two distinct peaks (as in the lower sample) indicative of the presence of a heteroduplexes, i.e. sequence differences.

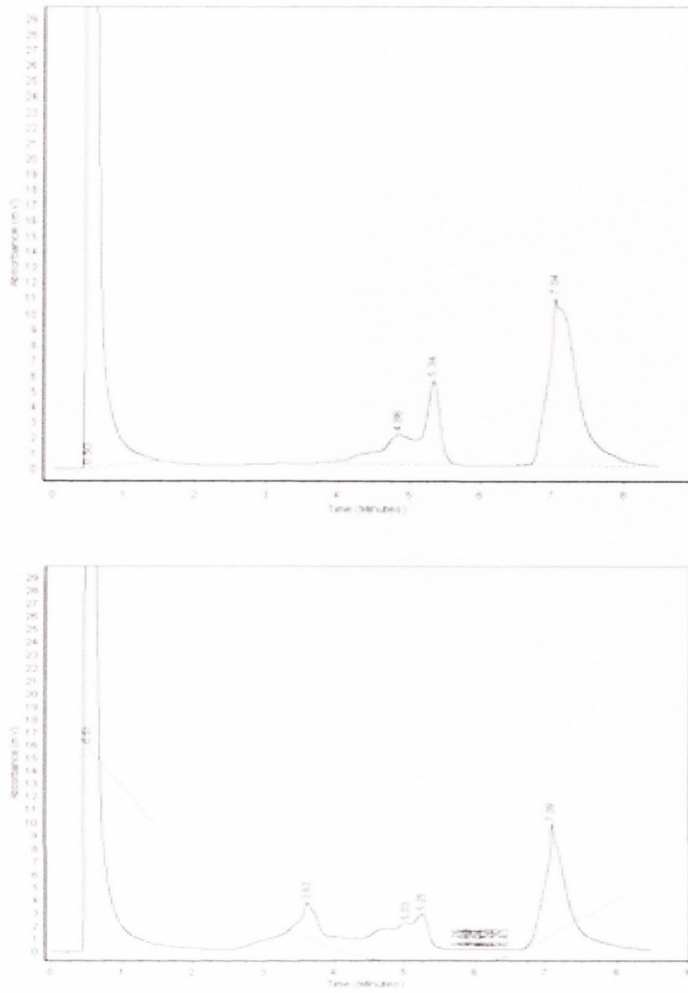


Figure 5.10 dHPLC ANALYSIS OF *ITGA4* EXON 16 AT 53°C. A number of samples showed the presence of two distinct peaks (as in the lower figure) indicative of the presence of a heteroduplexes.

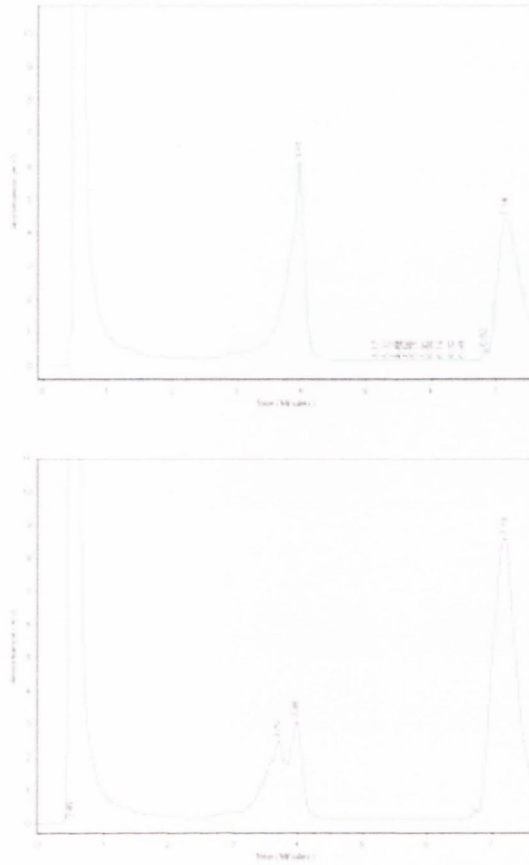


Figure 5.11 dHPLC ANALYSIS OF *ITGA4* EXON 16 AT 56°C. A number of samples showed the presence of two distinct peaks (as in the lower figure) indicative of a sequence difference.

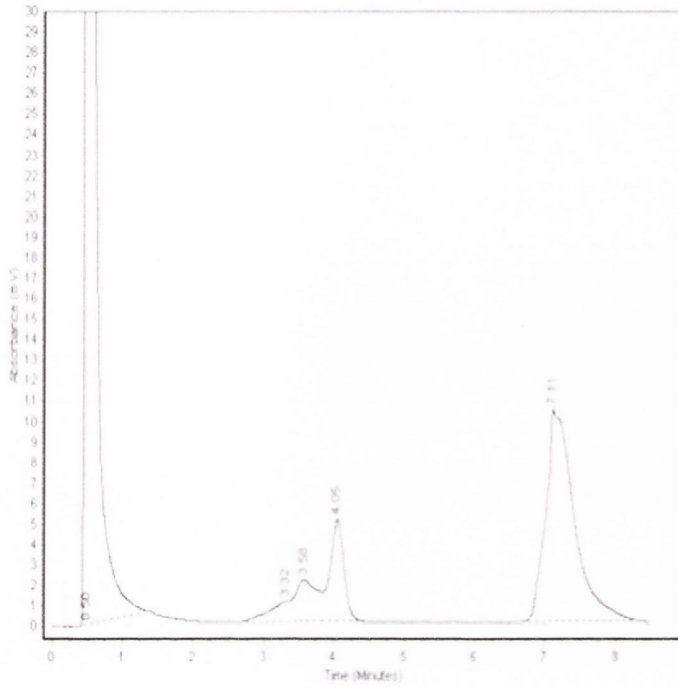


Figure 5.12 dHPLC ANALYSIS OF *ITGA4* EXON 17 AT 52°C. All samples revealed just one peak (with a small shoulder), indicating that there are no mutations present in any of the affected probands in this exon.

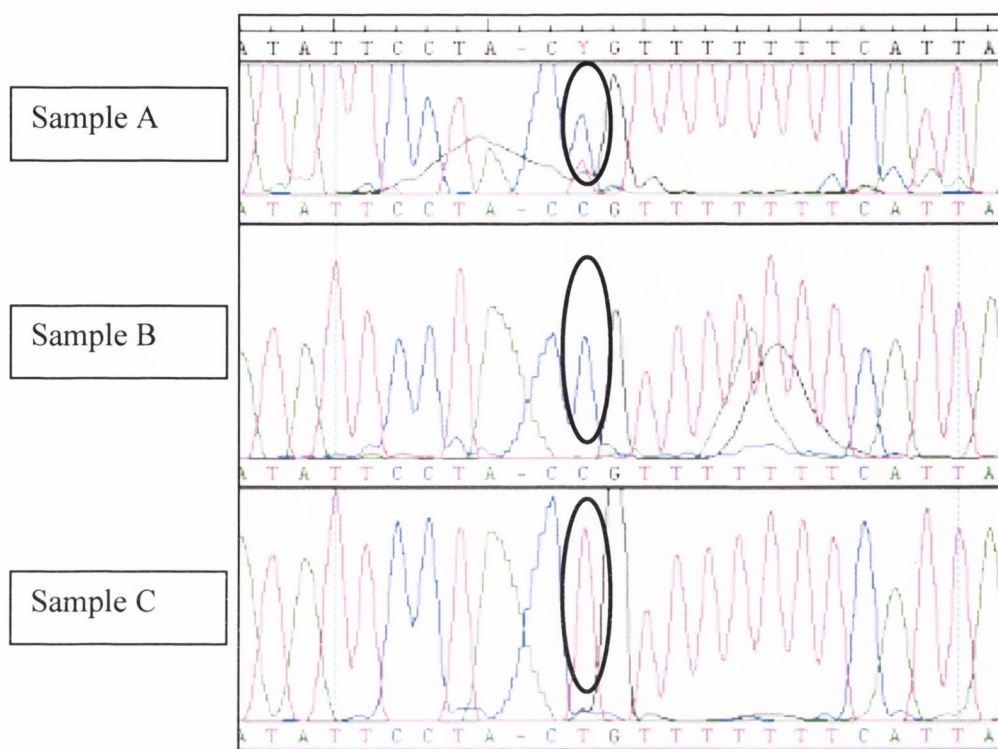


Figure 5.13 SEQUENCING RESULTS FOR MUTATION DETECTED BY dHPLC. This mutation (rs12690517) on the reverse strand has previously been identified on the exon 16 / intron 16 border. Sample A shows the presence of the C and T alleles (i.e. a heterozygote), sample B is a CC homozygote, whereas Sample C is a TT homozygote.

This SNP (coded as W) was genotyped in the complete sample using the SNaPshot method and an extension primer, on the reverse strand, specific to the SNP (see Section 2.2.5.3). Significant allelic transmission distortion to affected probands was detected. Sixty-four transmissions versus 35 non-transmissions of the C allele occurred, leading to a χ^2 value of 8.014, OR of 1.83 and p value of 0.0046. Transmissions of haplotypes containing this SNP were also highly significant. Transmission of the SNP12-W haplotype (C-C haplotype transmitted 56 times versus 32 non-transmissions) alleles produced a χ^2 value of 7.022 (OR = 1.75, p value = 0.0299). Transmission of the W-13 haplotype proved to be more significant, with 48

transmissions of the CA form of the haplotype compared to 16 non-transmissions, resulting in a χ^2 value of 14.148, OR of 3 and p value of 0.0027.

5.3.6 Attempted Replication

5.3.6.1 LD Analysis of Replication Samples

In order to validate the above findings, replication was sought in two independent samples kindly provided by colleagues at Vanderbilt University, Tennessee, USA. LD was measured in the AGRE and Vanderbilt samples using the Haploview programme (see Figures 5.14 and 5.15).

LD structure in the AGRE sample closely resembled the LD structure in the Irish sample (see Figures 5.6 and 5.14). Both samples had three blocks: 2 major haploblocks and a minor haploblock. The first haploblock in the AGRE sample also spanned 15kb from the region just 5' of the gene to intron 2. The second haploblock, the minor block, also included the same markers as in the Irish sample, rs155096 and rs155100, spanning intron 6 to intron 9. The only difference in LD and haploblock structure occurred in the last block. In the Irish sample, this block spanned 50kb, from intron 10 (rs3770117) to 3' of the gene (rs2290517). In the AGRE sample, this block only spanned 41kb, and excluded the final marker, rs3390517.

LD structure in the Vanderbilt sample showed a number of contrasts to the Irish, the AGRE sample and the HapMap sample. The most striking difference occurs in the breakdown of LD structure in the original Irish parental block 3 (from intron 10 to the 3' end of the gene). Block 1 and block 2 appear the same as in all previous examples, spanning the same regions and including the same SNPs. However, the strength of LD breaks down across haploblock 3 (in Irish, HapMap and AGRE samples). Instead of having a major block spanning ~50kb, there are three much

smaller blocks, lying side-by-side spanning 12.7kb in total. These minor blocks are in the regions from intron 15 to intron 23. Decreased LD with markers 10 (rs3770117), 11 (rs3770116), 13 (rs3770112) and 18 (rs2290517) is causing the absence of formation of the 2nd major block. Increasing the sample size of Vanderbilt sample may improve the LD relationships in this block, but it would be unlikely to restore the formation of this block.

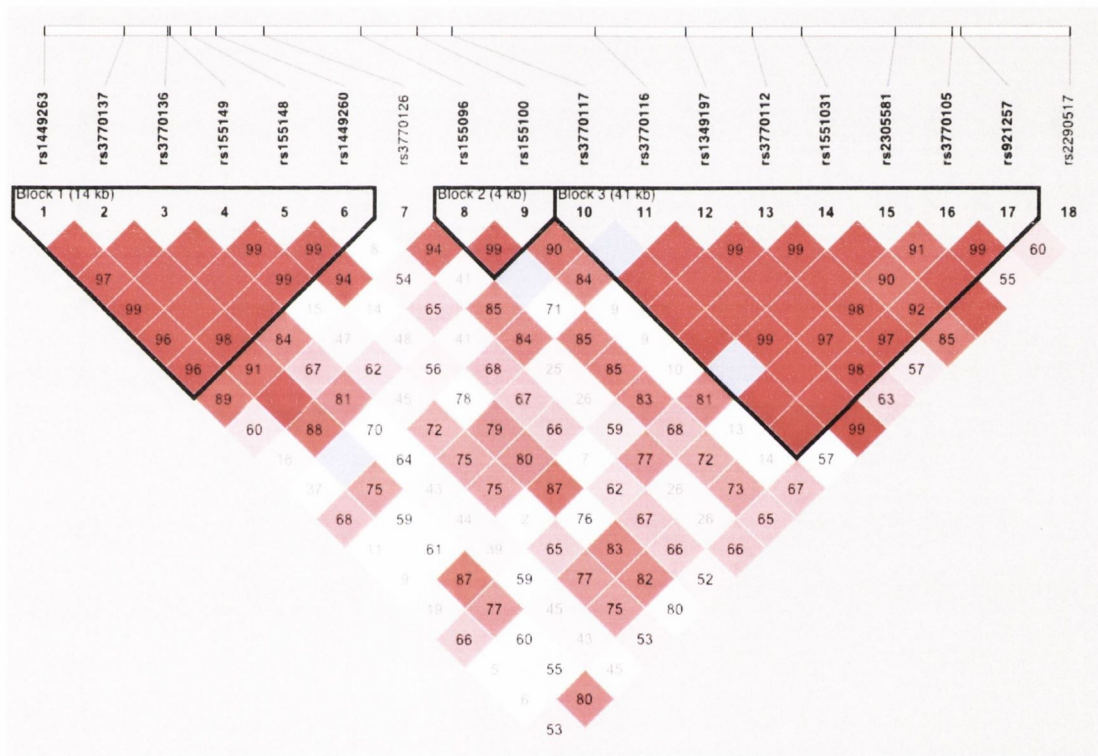


Figure 5.14 *ITGA4* LD STRUCTURE IN THE AGRE SAMPLE.

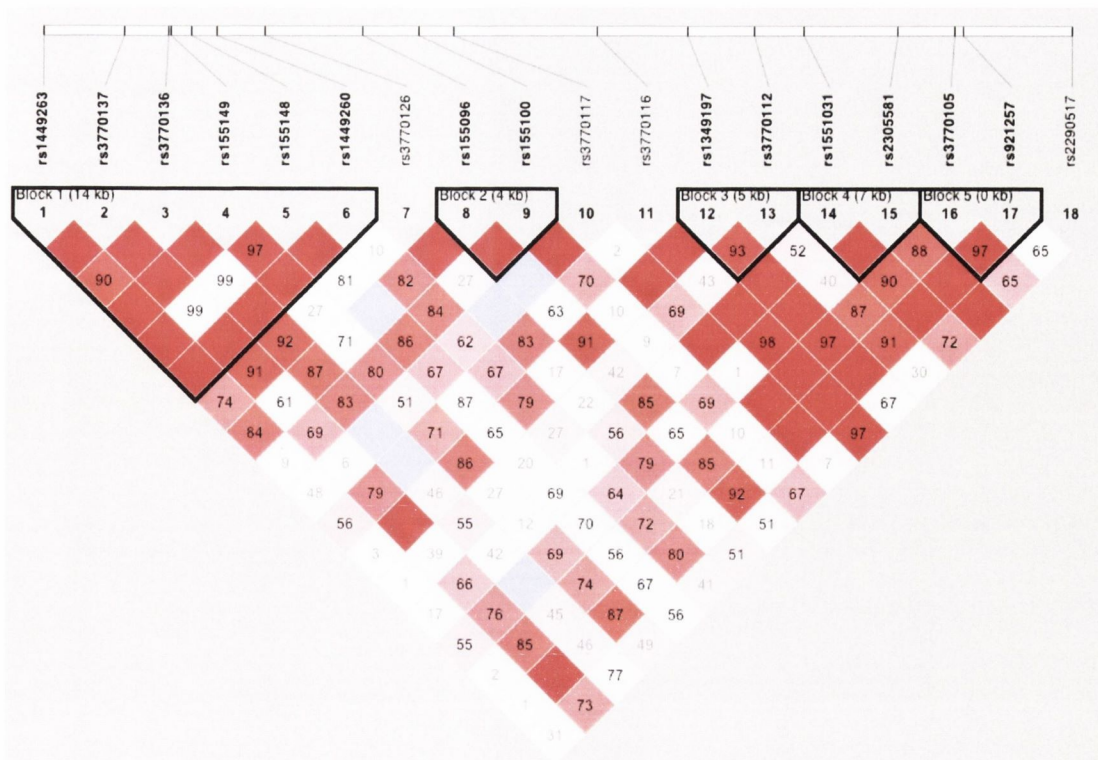


Figure 5.15 *ITGA4* LD STRUCTURE IN VANDERBILT SAMPLE

5.3.6.2 TDT Replication Results.

The transmission of alleles to affected probands was analysed using the PDT test, which tests the transmission of alleles in a general pedigree structure. The PDT-sum test was used instead of the PDT-avg test due to the increased power of this test, and all results correspond to the PDT-sum statistic.

There were no markers showing statistical evidence for over-transmission in the AGRE sample (see Table 5.4 for details).

Analysis of the Vanderbilt revealed that a number of SNPs showed significant, or very strong trends towards significant association with autism (see Table 5.5 for details). A SNP in intron 2, rs3770137, showed a trend towards over-transmission of the C allele (36 transmissions versus 22 transmissions) leading to a p value of

0.0647. Another 2 markers in introns 10 and 15, rs3770117 and rs3770116, also showed significant (and a strong trend towards significance) association with autism, arising from the over-transmission of the A alleles at each marker (rs3770117 p value = 0.0386, rs3770116 p value = 0.575). A final marker in intron 20, rs2305581 also showed significant over-transmission of the G allele to affected probands (p value = 0.0258). None of the aforementioned markers showed independent statistical association in either the Irish sample or the AGRE sample.

Marker	SNP ID	Allele	Transmitted	Not Transmitted	OR	Global p value
1	rs1449263	G	231	208	1.11	0.2772
2	rs3770137	G	82	80	1.03	0.8751
3	rs3770136	G	134	131	1.02	0.8558
4	rs155149	A	164	153	1.07	0.5367
5	rs155148	G	107	97	1.10	0.4838
6	rs1449260	A	183	167	1.10	0.3923
7	rs3770126	C	187	173	1.08	0.5961
8	rs155096	A	150	136	1.10	0.407
9	rs155100	T	220	206	1.07	0.4975
10	rs3770117	G	59	48	1.23	0.2872
11	rs3770116	C	82	80	1.03	0.8751
12	rs1349197	C	198	174	1.14	0.2742
13	rs3770112	C	205	183	1.12	0.2639
14	rs1551031	A	221	195	1.13	0.2022
15	rs2305581	A	88	85	1.04	0.8196
16	rs3770105	A	209	181	1.15	0.1561
17	rs921257	C	195	171	1.14	0.2095
18	rs2290517	A	96	75	1.28	0.1078

Table 5.4 PDT RESULTS FOR AGRE SAMPLE. PDT results revealed that no SNP was significantly over transmitted to affected probands.

Marker	SNP ID	Allele	Transmitted	Not Transmitted	OR	Global p value
1	rs1449263	G	71	64	1.11	0.546775
2	rs3770137	C	36	22	1.64	0.06469
3	rs3770136	A	41	37	1.11	0.6505
4	rs155149	A	62	53	1.17	0.4011
5	rs155148	A	39	28	1.39	0.178
6	rs1449260	A	63	62	1.02	0.9287
7	rs3770126	C	61	58	1.05	0.7833
8	rs155096	G	36	36	1.00	1.000
9	rs155100	A	82	73	1.12	0.4696
10	rs3770117	A	20	9	2.22	0.03859
11	rs3770116	A	39	24	1.63	0.05757
12	rs1349197	C	66	65	1.02	0.9304
13	rs3770112	C	61	58	1.05	0.7833
14	rs1551031	G	78	62	1.26	0.1758
15	rs2305581	G	42	24	1.75	0.02576
16	rs3770105	G	65	64	1.02	0.9298
17	rs921257	T	65	64	1.02	0.9298
18	rs2290517	C	46	39	1.18	0.4474

Table 5.5 PDT RESULTS FOR THE VANDERBILT SAMPLE. PDT testing revealed that a number of SNPs showed significant over-transmission (represented in yellow and bold) or strong trends towards over-transmission (represented in light yellow) to affected probands

5.3.6.3 Haplotype Transmissions in Replication Sample

All haplotype transmissions were tested using the UNPHASED programme. In the AGRE sample there were only two haplotypes showing evidence for over-transmission to affected probands. The haplotype containing markers 12 and 13 (C-C alleles), rs1349197 and rs3770112 respectively, was significantly over transmitted (155 transmissions versus 128 non-transmissions, OR = 1.21) leading to a p value of 0.015, see Table 5.6. The next haplotype (C-A alleles), with markers 13 (rs3770112) and 14 (rs1551031), also showed evidence of a trend towards significance, leading to an OR of 1.22 and a p value of 0.05014. These are the same two haplotypes (with the same over-transmitted alleles) that were significantly over-transmitted in the Irish population (see Table 5.3 for comparisons).

SNPs	Haplotype Over-transmitted	Transmitted	Not Transmitted	OR	Global p value
1_2	G-G	167	140	1.19	0.393
2_3	G-A	109	102	1.07	0.891
3_4	A-C	105	95	1.11	0.508
4_5	A-G	77	71	1.08	0.667
5_6	G-A	79	68	1.16	0.545
6_7	G-C	106	92	1.15	0.534
7_8	C-G	141	126	1.12	0.465
8_9	G-T	101	92	1.10	0.824
9_10	T-G	160	140	1.14	0.615
10_11	G-C	115	106	1.08	0.727
11_12	C-C	171	153	1.12	0.162
12_13	C-C	155	128	1.21	0.015
13_14	C-A	161	132	1.22	0.050
14_15	A-A	162	139	1.17	0.488
15_16	A-A	181	160	1.13	0.523
16_17	A-C	155	126	1.23	0.122
17_18	C-A	156	130	1.20	0.058

Table 5.6 TRANSMISSION OF *ITGA4* TWO MARKER HAPLOTYPES IN AGRE SAMPLE. SNP numbers correspond to those numbers (and hence SNP ID numbers) reported in Table 5.2. Two marker haplotypes that are significantly over-transmitted are highlighted in yellow and bold, whereas haplotypes showing a trend towards over-transmission are highlighted in light yellow.

Examination of transmissions of haplotypes in the Vanderbilt data revealed the significant transmission of markers different to those in the Irish and AGRE samples. Both the Irish and Vanderbilt samples show trends towards significant association with the 1-2 haplotype (rs1449263-rs3770137). However, the samples differ in the transmission of the allelic forms of the haplotype. The Vanderbilt sample shows over-transmission of the G-C form (p value = 0.0703) in comparison to the Irish sample, which shows over-transmission of the A-G form of the haplotype (p value = 0.0675). In effect, the proband samples are receiving the opposite forms of the haplotype.

The Vanderbilt sample also revealed the significant over-transmission of a haplotype involving markers 10-11 (rs3770117-rs3770116). The G-A form of the haplotype showed the transmission of 33 forms of this haplotype in comparison to 16 non-transmissions, resulting in an OR of 2.06 and a p value of 0.0185. Other haplotypes, including 14-15 (rs1551031-rs2305581) and 17-18 (rs921257-rs2290517) showed trends towards association (p values of 0.064 and 0.099 respectively).

SNPs	Haplotype Over-transmitted	Transmitted	Not Transmitted	OR	Global p value
1_2	G-C	26	13	2.00	0.0703
2_3	C-G	30	14	2.14	0.04912
3_4	G-A	43	35	1.23	0.6661
4_5	A-A	29	15	1.93	0.1082
5_6	A-A	35	27	1.30	0.5431
6_7	A-C	22	14	1.57	0.2284
7_8	C-G	40	37	1.08	0.9404
8_9	G-A	58	41	1.41	0.1075
9_10	A-G	68	52	1.31	0.2935
10_11	G-A	33	16	2.06	0.01825
11_12	A-C	28	16	1.75	0.1623
12_13	C-C	47	42	1.12	0.6501
13_14	C-G	26	14	1.86	0.1732
14_15	G-G	26	12	2.17	0.06414
15_16	G-A	29	13	2.23	0.1026
16_17	G-T	45	44	1.02	0.7299
17_18	C-C	29	16	1.81	0.09924

Table 5.7 TRANSMISSION OF *ITGA4* TWO MARKER HAPLOTYPES IN VANDERBILT SAMPLE. SNP numbers correspond to those numbers (and hence SNP ID numbers) reported in Table 5.2. Two marker haplotypes that are significantly over-transmitted or showing strong trends towards significance are highlighted in yellow and bold, and light yellow respectively.

5.3.6.4 Comparison of 4-marker haplotypes across all 3 samples

In order to test the possibility of whether there was the transmission of a disease-causing low frequency haplotype common to all 3 samples, the transmission of 4-marker haplotypes was tested in all 3 samples using UNPHASED. Transmissions were tabulated and examined by eye. Only haplotypes in those regions that showed evidence of association in any of the three samples were investigated, i.e. haplotypes around markers 1, 2, 9, 10, 12, 13, 14, 15 and 17.

There were a number of haplotypes around markers 13 and 14 that showed similarities in transmissions in both the Irish and AGRE samples (blue cells in Table 5.8). This result was expected as both groups showed over-transmission of the C-C (2-2) form of the haplotype. There were also a number of similarities between the AGRE and Vanderbilt samples (green cells) but these haplotypes did not continue to be passed as the haplotype moved distal including the next marker, i.e. transmission of a 1-2-1-1 haplotype was not followed by a 2-1-1-X haplotype. There were no similarities between the Irish and Vanderbilt samples in the 3' region of the gene (yellow cells).

There were also only three common haplotypes transmitted to affected probands in all three groups (pink cells). Two of these spanned the intron 2 – intron 10 and the intron 6 - intron 15 gene region. However this region only included the proximal region of the final major haploblock in the gene. One common haplotype occurred in the central region of the final haploblock. This haplotype contained markers 12-13-14-15. However under-transmission of this haplotype did not remain significant with the addition of extra markers.

SAMPLES			
	Irish	AGRE	Vanderbilt
marker 1-2-3-4			
1112	0.737 (33)	0.926 (104)	2.090 (34)
1211	0.667 (25)	1.16 (54)	0.235 (21)
1212	0.857 (52)	1.067 (124)	0.944 (35)
1221	0.625 (39)	1.353 (160)	1 (46)
2211	1.611 (94)	0.827 (274)	1.048 (86)
2221	0 (0)	0.75 (7)	0 (0)
marker 2-3-4-5			
1122	0.619 (34)	0.982 (109)	2.083 (37)
2112	1.463 (101)	0.963 (314)	0.868 (99)
2121	0.923 (50)	1.062 (134)	0.727 (38)
2212	0.769 (46)	1.195 (191)	1.074 (56)
marker 7-8-9-10			
1111	0.5 (3)	0.75 (7)	0.2 (6)
1121	1.263 (86)	0.813 (272)	1 (96)
1122	0.929 (27)	0.762 (74)	1.143 (15)
1211	0.970 (65)	1.123 (172)	1.083 (50)
2111	0.833 (33)	1.143 (165)	0.714 (48)
2121	0.75 (25)	1.283 (121)	1.846 (37)
2211	2 (3)	2 (6)	0 (0)
marker 8-9-10-11			
1112	0.84 (46)	1.093 (180)	0.611 (58)
1211	0.7 (34)	1.196 (112)	2.091 (34)
1212	1.255 (106)	0.862 (270)	1.122 (104)
1222	0.867 (28)	0.814 (78)	0.857 (13)
2111	3 (4)	0.75 (7)	2 (3)
2112	0.971 (67)	1.129 (181)	1 (58)
marker 9-10-11-12			
1112	1 (6)	0.75 (7)	1.5 (5)
1121	0.75 (35)	1 (132)	0.864 (41)
1122	1.056 (74)	1.098 (214)	0.8 (72)
2112	0.632 (31)	1.229 (107)	1.75 (33)
2121	0.9 (57)	0.821 (153)	1.208 (53)
2122	1.533 (76)	0.981 (210)	1 (80)
2222	0.8 (27)	0.944 (70)	0.5 (12)

Similar RR in Irish & Vanderbilt Samples
Similar RR in AGRE & Vanderbilt Samples
Similar RR in Irish & AGRE samples
Similar RR in all 3 samples

Table 5.8(a) COMPARISON OF 4 MARKER HAPLOTYPES TRANSMITTED TO AFFECTED SAMPLES. Transmissions are considered “similar” if the ORs for the haplotypes are both / all > 1.1 or both / all < 0.9. Similarities between the Irish and Vanderbilt samples are represented in yellow. Similarities between the Irish and AGRE samples are in blue, whereas those between the Vanderbilt and AGRE samples are in green. If a common haplotype was over or under-transmitted in all three groups, the cell is pink in colour. The numbers in brackets represent the number of haplotypes transmitted.

SAMPLES			
	Irish	AGRE	Vanderbilt
marker 10-11-12-13			
1122	0.654 (43)	1.048 (127)	2.077 (40)
1211	0.527 (84)	0.845 (262)	0.974 (75)
1212	0 (0)	2 (2)	1.333 (7)
1222	1.947 (112)	1.194 (316)	0.912 (109)
2222	0.882 (32)	0.878 (77)	0.385 (18)
marker 11-12-13-14			
1221	0.583 (38)	1.020 (103)	2.375 (27)
2111	0.652 (76)	0.831 (249)	1.1 (63)
2121	0 (0)	2 (3)	1 (3)
2221	2 (3)	0.667 (5)	1 (2)
2222	1.839 (88)	1.223 (269)	0.689 (76)
marker 12-13-14-15			
1112	0.66666667	0.851 (248)	0.727 (76)
1212	0	2 (2)	0.75 (7)
2211	0.60869565	1.057 (109)	2.11 (28)
2222	1.74193548	1.185 (271)	0.814 (78)
marker 13-14-15-16			
1122	0.630 (75)	0.896 (256)	1.030 (67)
2111	0.619 (34)	1 (116)	2.5 (28)
2121	0 (0)	1 (4)	0.25 (5)
2122	0 (0)	1 (2)	2 (3)
2221	1.719 (87)	1.186 (282)	0.778 (80)
2222	1 (2)	0.818 (20)	0.5 (3)
marker 14-15-16-17			
1112	0.682 (37)	1.057 (109)	2.333 (30)
1221	0.792 (86)	0.889 (238)	1.030 (69)
2212	1.395 (91)	1.198 (266)	0.783 (82)
2221	0.667 (5)	0.467 (22)	0 (0)
2222	0 (0)	0.667 (5)	1 (4)
marker 15-16-17-18			
1121	0.857 (39)	0.982 (113)	2.455 (38)
2122	1.381 (100)	1.226 (276)	0.763 (104)
2211	0 (0)	0.25 (10)	0.333 (4)
2212	0.765 (90)	0.927 (237)	1.128 (83)
2222	0 (0)	0.667 (5)	1 (4)

Similar RR in Irish & Vanderbilt Samples
Similar RR in AGRE & Vanderbilt Samples
Similar RR in Irish & AGRE samples
Similar RR in all 3 samples

TABLE 5.8(B) COMPARISON OF 4 MARKER HAPLOTYPES TRANSMITTED TO AFFECTED SAMPLES. Transmissions are considered “similar” if the ORs for the haplotype are both / all > 1.1 or both / all < 0.9. Similarities between the Irish and Vanderbilt samples are represented in yellow. Similarities between the Irish and AGRE samples are in blue, whereas those between the Vanderbilt and AGRE samples are in green. If a common haplotype was over or under-transmitted in all three groups, the cell is pink in colour. The numbers in brackets represent the number of haplotypes transmitted.

5.3.6.5 Joint Analysis

To test if the association found with rs3770112 remained in a sample consisting of all three samples, a post-hoc TDT and haplotype transmission test was undertaken. This amalgamated sample consisted of 1231 affected individuals, within 524 individual pedigrees.

PDT testing of the transmission of individual alleles from individual markers was undertaken. Only one marker showed evidence to support association with autism. The C allele of rs3770112 was significantly over-transmitted to affected probands (335 times versus 281 times non-transmitted) leading to a χ^2 of 4.74, OR of 1.19 and p value of 0.02947. Another marker, rs9211257, which showed a trend towards association in the Irish sample (see Table 5.2), but no trend in either the AGRE or Vanderbilt samples (see Tables 5.4 and 5.5), showed a trend towards over-transmission of the C allele (p value = 0.08381). Another marker, this time in intron 10, rs3770117, also showed a trend towards association. However this did not reach statistical significance (p value = 0.0838). In the joint analysis no other markers showed any evidence of association with autism (see Table 5.9).

Transmission of 2-marker haplotypes was also tested in this amalgamated sample using UNPHASED. In a fashion similar to the original Irish sample, the haplotypes containing marker rs3770112 (SNP13) showed significant over-transmission to affected probands. The C-C form of haplotype SNP12-SNP13 was significantly over-transmitted (259 transmissions versus 213 non-transmissions) resulting in a χ^2 of 15.0 (3 df) and p value of 0.001815. The OR of this haplotype was relatively low at 1.22, indicating that if this is a true OR, the ITGA4 only plays a minor role in the development of autism. Another haplotype containing the rs3770112 marker, this time in association with the rs1551031 marker, also showed evidence of distorted transmissions. Over-transmission of the C-C haplotype proved to be statistically significant, leading to a χ^2 of 10.91, OR of 1.05 and p value of 0.01221.

A third, independent haplotype also showed evidence for association. The final 2-marker haplotype in the gene, involving markers rs921257 and rs2290517, showed non-significant over-transmission of the C-A form of the haplotype, producing a χ^2 of 7.753 and p value of 0.05141. This haplotype showed a strong trend towards association with autism in the AGRE sample (p value = 0.05845, see Table 5.6). It showed no trend towards association in the Irish sample, although the C-A form of the haplotype was over-transmitted (p value = 0.1967, see Table 5.3). The Vanderbilt sample showed a trend towards association with this haplotype also. However the trend in this sample was not the over-transmission of the C-A form of the haplotype, but instead the CC form of the haplotype, yielding a p value of 0.09924 (see Table 5.10 for further information). The information gained in the post-hoc testing resulted in the identification of a potentially important haplotype that may otherwise have gone unnoticed.

Marker	SNP ID	Allele	Transmitted	Not Transmitted	OR	Global p value
1	Rs1449263	G	360	358	1.01	0.9405
2	Rs3770137	C	135	132	1.02	0.8543
3	Rs3770136	G	208	199	1.05	0.6556
4	Rs155149	A	271	265	1.02	0.7955
5	Rs155148	A	171	167	1.02	0.8278
6	Rs1449260	A	299	295	1.01	0.8696
7	Rs3770126	C	285	285	1.00	1.0000
8	Rs155096	G	237	223	1.06	0.5139
9	Rs155100	T	352	351	1.00	0.9699
10	Rs3770117	G	99	74	1.34	0.0569
11	Rs3770116	A	145	138	1.05	0.6773
12	Rs1349197	C	327	288	1.14	0.1157
13	rs3770112	C	335	282	1.19	0.0295
14	Rs1551031	A	362	333	1.09	0.2712
15	Rs2305581	A	152	143	1.06	0.6003
16	Rs3770105	G	329	321	1.03	0.7537
17	Rs921257	C	331	288	1.15	0.0838
18	Rs2290517	A	162	145	1.12	0.3318

Table 5.9 JOINT PDT ANALYSIS OF *ITGA4* IN THE AMALGAMATED SAMPLE. PDT analysis was undertaken in a sample comprising of all three samples (Irish, AGRE and Vanderbilt). The results indicate that rs3770112 remains significant (p value = 0.295). Two other SNPs, rs3770117 and rs921257, also show a strong trend towards association (p values of 0.0569 and 0.0838 respectively).

SNPs	Haplotype Over-transmitted	Transmitted	Not Transmitted	OR	Global Chi Square	Global p value
1_2	G-C	102	98	1.04	0.0788	0.9614
2_3	C-G	110	104	1.06	0.2565	0.8796
3_4	G-A	197	192	1.03	1.552	0.6704
4_5	A-G	127	124	1.02	1.463	0.6908
5_6	G-A	128	123	1.04	0.1019	0.9916
6_7	G-C	163	155	1.05	0.2847	0.9629
7_8	T-G	188	166	1.13	1.773	0.6208
8_9	G-T	166	158	1.05	2.64	0.4504
9_10	T-G	257	249	1.03	1.337	0.7203
10_11	G-A	126	118	1.07	5.147	0.1614
11_12	A-C	112	105	1.07	1.723	0.4224
12_13	C-C	258	202	1.28	15	0.0018
13_14	C-A	259	213	1.22	10.91	0.0122
14_15	A-G	260	233	1.12	4.649	0.1994
15_16	A-G	103	86	1.2	3.099	0.3767
16_17	G-C	201	172	1.17	4.661	0.1984
17_18	C-A	265	235	1.13	7.753	0.0514

Table 5.10 JOINT ANALYSIS FOR TWO-MARKER HAPLOTYPES IN THE AMALGAMATED SAMPLE. The transmission of two-marker haplotypes in the amalgamated sample (Irish, AGRE and Vanderbilt) was analysed using the UNPHASED programme. Significant transmissions are highlighted in yellow and bold. Trends towards significance are highlighted in light yellow. There was significant transmission of the haplotypes containing the markers spanning (and just 3' to) exons 16 and 17 (haplotypes 12_13 and 13_14). A haplotype containing the last two SNPs also showed a very strong trend towards significance (p value = 0.0514)

5.3.4 Testing Validity of htSNPs

Haplotype tagging SNPs (htSNPs) were chosen in the control sample using the Haploview Programme. This was undertaken to test if genotyping a small number of controls and choosing htSNPs in this sample, followed by the genotyping of these htSNPs in the autism population would detect genetic associations. In the first major block, which contained 6 SNPs, 5 SNPs were chosen as htSNPs (rs1449263, rs3770137, rs3770136, rs155148 and rs1449260).

Within the final haploblock there were 8 SNPs. Only two of these, rs1551031 (in intron 17) and rs1349197 (in intron 21) were highlighted as htSNPs. Genotyping both of these SNPs allow representation of 98.9% (the sum of the three major haplotypes) of all haplotypes within the control block (see Figure 5.16).

The relationship between haplotypes is also produced by the Haploview Programme, see Figure 5.16. Thick black lines linking haplotypes from different blocks indicate that these haplotypes are observed to be in LD with a frequency of at least 10%, indicating the presence of a ‘two-block haplotype’.

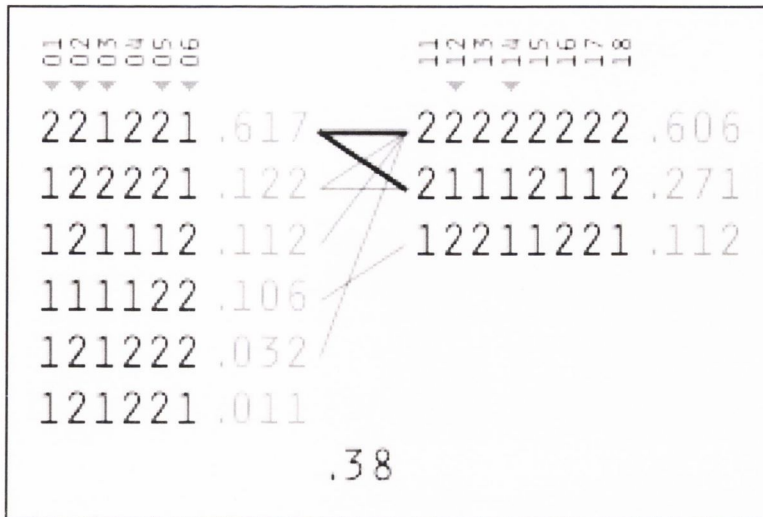


Figure 5.16 HAPLOTYPE TAGGING SNPs AND INTER-BLOCK RELATIONSHIPS.

Haplotype tagging SNPs are indicated by the presence of arrow-heads. Thick black lines indicate the presence of ‘two-block haplotypes’ that occur with a frequency of at least 10%. The frequencies of the haplotypes are indicated in grey numbers, and the frequency of ‘two-block haplotypes’ is indicated in black numbers.

The transmission of single marker alleles and the haplotypes consisting only of those htSNPs was tested in the Irish sample. TDT analysis and two-marker haplotype analysis using the htSNPs in the first block would not have produced results different to the initial haplotype analysis. The trend towards significant over-transmission of both the first two SNPs and the rs1449263-rs3770137 haplotype (SNP1-SNP2 haplotype) would have been identified as important and followed up if necessary. Similarly, TDT analysis of individual htSNPs in the final block would have highlighted the possible association with the *ITGA4* gene and follow up analysis would have been undertaken. Surprisingly, in the case of the htSNPs in this second major haploblock, in which many of the significant associations in all three samples were found, transmission of the htSNP haplotypes (using rs1349197 and rs1551031) was not significant. Nor did it show any significant trend towards association. Transmission of this haplotype resulted in a χ^2 of 4.301 (3df) and p value of 0.2307.

The selection of htSNPs was also studied in a “test” sample from the AGRE and Vanderbilt samples. These comprised the first 94 parents in each dataset, as control samples from these samples were not available. In the case of the AGRE sample (see Figure 5.17) different htSNPs were chosen in all blocks in contrast to the Irish control sample. In the final major haplotype block, SNPs 14 and 16 (in comparison to SNPs 12 and 14 in the Irish sample) were chosen to capture the most variation. TDT analysis of just these two markers would have resulted in p values of 0.2022 and 0.1561 (rs1551031 and rs3770105 respectively). Haplotype analysis of just these two markers would have resulted in a p value of 0.154. Given the significance of the TDT results, further fine-mapping would have been undertaken. In the case of the htSNPs in blocks one and two, TDT and haplotype analysis did not produce any significant results.

Selection of htSNPs in the Vanderbilt sample included SNPs 1, 2, 3, 4 and 6 in block 1, which was the same as in the Irish control sample (see Figures 5.16 and 5.18). Individual TDT and haplotype analysis would have highlighted the importance of rs2770137 (SNP2) in this sample (see Table 5.5). Haplotype analysis of htSNPs in block 2, SNPs 8 and 9 would have also produced a significant result, which would lead to follow up testing. SNPs 14, 15, 16 and 17 were all selected as htSNPs to help describe variation over the two small blocks (see Figure 5.15), and TDT and haplotype analysis would have resulted in significant results in this region also.

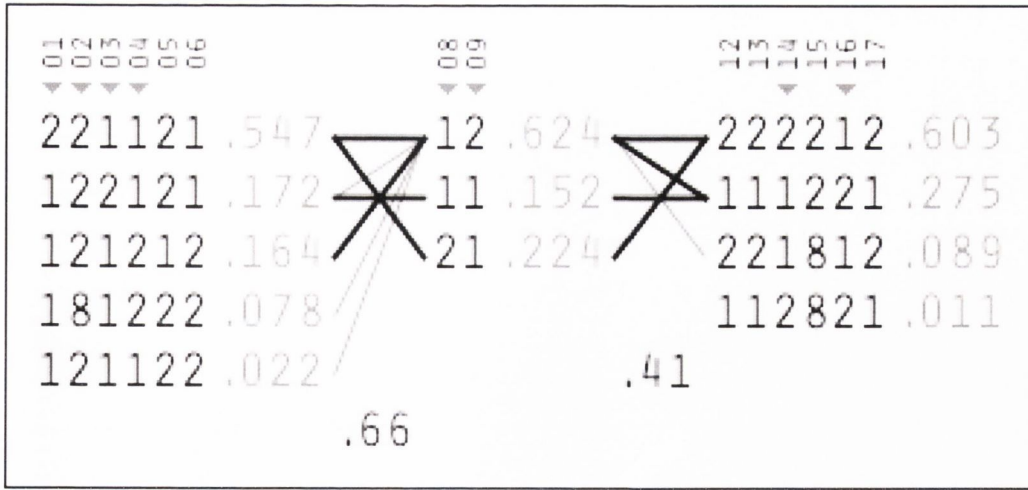


Figure 5.17 HAPLOTYPE TAGGING SNPs AND INTER-BLOCK RELATIONSHIPS IN AGRE SAMPLE.

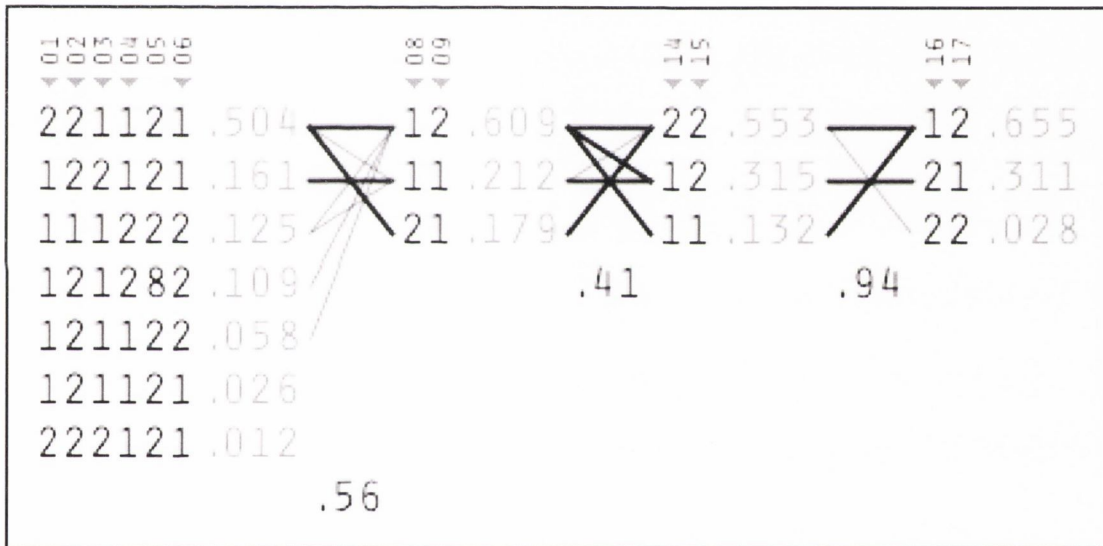


Figure 5.18 HAPLOTYPE TAGGING SNPs AND INTER-BLOCK RELATIONSHIPS IN VANDERBILT SAMPLE.

5.4 Discussion

5.4.1 LD Structure of ITGA4 Gene

It has been suggested that disparate populations can show different LD patterns. We had an opportunity to test this in the comparison of the Irish data, HapMap data, AGRE data and Vanderbilt data. The Irish LD structure was compared to data from the HapMap project. Both samples showed the presence of a number of regions with the gene showing extensive LD. Due to the difference in markers the results produced are not directly comparable. However, both samples show the presence of a haploblock at the 5' region of the gene (our block 1 versus the HapMap block 2), followed by a break in LD, which in turn is followed by the presence of a small haploblock. The largest haploblock in the HapMap sample likewise occurs at the end of the gene. In the HapMap sample, this final block is shorter than the same block in the Irish sample. This results from the inclusion of two SNP (rs3770115 and rs1816990) both in intron 15, which disrupts the formation of a larger block. Apart from these minor differences, the structure of LD in the Irish sample is comparable to that of the HapMap CEPH sample.

5.4.2 TDT, Haplotypes and Mutation Detection

TDT analysis revealed the presence of a number of associations significant at the $p < 0.05$ level, especially surrounding rs3770112 in intron 17. Haplotypes involving this SNP had transmissions were also statistically over-transmitted to probands with autism (p values = 0.0299 and 0.0027, see Table 5.3).

The strength of results, in close proximity to exons 16 and 17, led to the speculation that there may be 'disease-causing' variants located in these exons. This theory was investigated using dHPLC mutation analysis. No mutations were observed in exon

17. A number of samples showed chromatogram shifts (see Figures 5.9, 5.10 and 5.11) for exon 16. Sequencing of these samples revealed the presence of an A to G transition on the exon 16 / intron 16 border. This SNP (rs12690517) had been previously described by Heymann et al ⁴³². The exact nature of the SNP is still unclear. Two databases (dbSNP and Santa Cruz Genome Assembly) describe it as intronic, whereas another (Ensembl) describes it as exonic. Despite the ambiguity of its exact location, this A to G transition may alter a potential splicing sequence. The consensus splicing sequence is AG/GUPuAGU ⁴³². In the human databases, the sequence analogous to this is (in the presence of the G allele) CG/GUAGGU (differences are underlined). In the presence of the A allele, the splicing sequence is altered one step further away from the consensus sequence than the original sequence, to become CA/GUAGGU. The original report describing this SNP found no difference in mRNA sequence in individuals carrying both forms of the splice variant ⁴³². However the analysis was undertaken using mRNA extracted from blood. It is not yet known if splicing is altered by the SNP in the CNS.

TDT analysis of this SNP revealed significant over-transmission of the C (genotyped on the reverse strand, hence G allele on forward strand) allele. There were 64 transmissions of the C allele to affected probands compared to 35 non-transmissions. These transmissions resulted in a χ^2 of 8.014, p value of 0.0046, and an OR of 1.82, which was the most significant result of all single TDT analyses. Haplotype transmissions involving the C allele of this SNP were also highly significant. Over-transmission of a C-C haplotype with SNPs rs1349197-rs3770112 occurred with 56 transmissions in comparison to 32 non-transmissions ($\chi^2 = 7.022$, p value = 0.0299, OR = 1.75). Transmission of the C-C haplotype of SNPs rs3770112 – rs12690517 proved to be even more significant, with 48 transmissions compared to 16, resulting in a χ^2 of 14.148, p value of 0.0027 and OR of 3. These results indicate that this haplotype is an important risk haplotype for individuals with autism, at least in the Irish population.

5.4.3 Attempted Replication

Replication was sought in two independent samples (AGRE and a local Vanderbilt sample) kindly provided by colleagues at Vanderbilt University.

5.4.3.1 AGRE Sample

The AGRE sample, consisting of 267 families, revealed no significant association with any single marker using the PDT test. Haplotype analysis revealed the over-transmission of just two haplotypes, the 12-13 and 13-14 haplotypes. Over-transmission of the C-C form of haplotype 17-18 occurred, as in the Irish sample, giving a global p value of 0.015 and OR of 1.21.

Although the effect size (OR) of the haplotypes is decreased compared to the equivalent OR in the Irish sample, the over-transmission of the same haplotypes is supporting evidence for the potential role of *ITGA4* in the development of autism. Also, it is a common occurrence in the field of complex genetics, for the first study finding an association with a gene to show an increased risk (as described by p values and ORs) in comparison to subsequent studies⁴³³. Therefore the decreased significance of the AGRE sample is not surprising.

5.4.3.2 Vanderbilt Sample

The results of the Vanderbilt sample may also show evidence of replication. However, in this sample the replications do not arise from the SNPs rs1349197, rs3770112 or rs1551031. Instead, there was evidence of distorted transmission of alleles in SNP rs3770117 (located in intron 10), with increased transmission of the A allele to affected probands (p value = 0.039) and also increased transmission of the G

allele of rs2305581 (p value = 0.026). Both of these SNPs are located in the last major haploblock in the Irish sample. Haplotypes involving these SNPs also proved to be significant; the haplotype rs3770117-rs3770116 (increased transmission of the G-A form) had a global p value of 0.018, and the haplotype rs1551031-rs2305581 (increased transmission of the G-G form) showed a trend towards significance, having a p value of 0.064.

Both the Irish and Vanderbilt samples show trends towards significant association with the 1-2 haplotype (rs1449263-rs3770137). However, the samples differ in the transmission of the allelic forms of the haplotype. The Vanderbilt sample shows over-transmission of the G-C form (p value = 0.0703) in comparison to the Irish sample, which shows over-transmission of the A-G form of the haplotype (p value = 0.0675). In effect, the proband samples are receiving the opposite forms of the haplotype. Further work is required to establish the nature of associations in the 5' region of the *ITGA4* gene.

It is important to point out at this stage that these differences in allelic and haplotypic over-transmissions in the Vanderbilt sample may have arisen from the differences in LD structure within the *ITGA4* gene. LD structure in the first half of the gene (from the region just 5' of the gene to intron 10) in the Vanderbilt sample is comparable to that in the Irish sample (see Figures 5.6 and 5.15). However, from marker rs3770117 onwards there seems to be a breakdown in LD. Instead of a haploblock spanning 50kb (as in the Irish sample) there are only three smaller LD haploblocks (each comprising of just 2 markers). It is therefore possible that the associations in the Irish and AGRE samples are in LD with a functional variant elsewhere in the gene that itself is in LD with the haplotypes mentioned above in the Vanderbilt sample.

5.4.3.3 Joint Analysis.

A joint analysis was undertaken post-hoc to assess whether the associations found in the Irish sample still remained in a larger amalgamated sample. All three samples tested for association were combined into one large sample consisting of 1231 affected individuals in 524 families. In a fashion similar to the original results in the Irish sample, significant over-transmission of the C allele of rs3770112 was observed (see Table 5.9). The two haplotypes containing this marker also remained significant in the amalgamated sample. The haplotype spanning exons 16 and 17 (and hence the splice variant) was the most significantly over-transmitted haplotype (see Table 5.10). The OR for the C-C haplotype in this amalgamated sample (OR = 1.22) was much smaller than the OR in the original Irish sample (OR= 1.75), but the *ITGA4* gene, based on the observations in the amalgamated sample, still shows strong evidence for association.

Another SNP that showed a trend towards association in the amalgamated sample was rs9211257. This SNP was one of the original SNPs used in the fine-mapping experiment that highlighted *ITGA4* as a potential gene of interest within the deleted region of chromosome 2q in patient KM. In the amalgamated sample, there was increased transmission of the C allele. This SNP showed increased transmission in the Irish sample (p value = 0.0719) and the AGRE sample (p value = 0.21), but not in the Vanderbilt sample (p value = 0.93). Transmission of the rs9211257-rs2290517 haplotype was also significant, with over-transmission of the C-A form (p value = 0.05). Once again, this was the haplotype over-transmitted in the Irish and AGRE samples (p value of 0.1967 and 0.058 respectively), but not in the Vanderbilt sample, which showed increased transmission of the C-C haplotype (p value = 0.099). It is possible that the breakdown in LD structure in the 3' region of the gene is partly responsible for the discrepancies between the Irish / AGRE and Vanderbilt samples (see Figure 5.15).

5.4.4 Possible Theories for Differences in Haplotype / Allele Transmission

Both the AGRE and the Vanderbilt samples show association with haplotypes in haploblock 3. However, in the case of the Vanderbilt sample the pattern of association is not the same as the Irish sample. It is possible that the association in the *ITGA4* gene in autism is similar to the association of the Dysbindin gene to schizophrenia, i.e. that there are a number of susceptibility haplotypes within the gene. A report by Williams et al ⁴³⁴ describes 1 risk and 2 protective dysbindin haplotypes associated with schizophrenia in 2 independent samples. Other associated haplotypes have been identified in other populations at this locus but they differ to those described in the Williams paper. Another report by Bray et al found that the “Williams risk haplotype” tags cis-acting variants (as yet unknown variants upstream of the gene), and these variants result in reduced levels of dysbindin mRNA ⁴³⁵. In addition, they showed that the different risk haplotypes from other samples also tagged these cis-acting variants. This demonstrates that there can be different associated haplotypes causing the same effect (reduction in dysbindin mRNA). All this could be more fully explained if the cis-acting variants that affect gene expression were identified and their LD relationship with risk and protective haplotypes determined. Further replications in other samples would be necessary to test if there are multiple risk haplotypes in the *ITGA4* gene associated with autism.

It is also possible that these associations in different haplotypes may occur due to clinical differences within the samples, i.e. an *ITGA4* variant is associated with a trait that is more prevalent in the Irish and AGRE populations. Due to the differences in LD structure, there also remains a possibility that the association in the Irish and AGRE sample is in LD with the same functional mutation as the Vanderbilt sample, but this functional mutation is in LD with differing haplotypes.

Finally it is possible that all the associations found above are merely chance findings. Due to the fact that “significance” is deemed / accepted to occur at a p value less

than 0.05, this still means that there is still a 1 in 20 chance of finding an association by chance. Therefore with 19 SNPs (in the Irish sample) tested, one would expect ~1 to show an association, even if it is not a true association. In the Irish sample, there were 3 SNPs tested that met the criteria for significance (rs1449263, rs12690517 and rs3770112) with p values ranging from 0.0192 to 0.0046. Two haplotypes (rs1349197-rs3770112 and rs3770112-rs1551031) showed evidence for association in the Irish and AGRE samples. While it remains a possibility that these haplotypes are both false positives in their respective samples, the possibility is much lower than the original p value of 0.05. The chance that given 17 possible haplotypes that the same two haplotypes will show evidence for association is in fact 0.0138 (i.e. $2/17 \times 2/17$). The joint analysis also highlights these haplotypes as significant. The power of a genetic study is often a limiting factor in any study. In the joint analysis however, the sample size was adequate to detect the association reported here. Therefore, while chance findings are still possible, the evidence for association in the study reported here reduces this possibility.

5.4.5 Biological Hypothesis

The role of integrins in brain development and working function has been investigated in recent years. Integrin acts as an intermediary protein linking the actin cytoskeleton within the cell to the extra-cellular matrix (see Figure 5.4)⁴³⁶. It is through these interactions that integrin proteins may play a part in neural cell adhesion, neural cell spreading, synaptogenesis, hippocampal synaptic plasticity and spatial memory (see Table 5.11 for a list of integrin protein and brain development reports, also see Milner et al⁴³⁷ for a full review of the role of integrin proteins in the CNS)^{415 438}.

ITGA4 has been reported to be involved in a number of neural processes (see section 1.1.4). Experimentation has shown that the ITGA4 cytoplasmic tail is important for determining the extent of neural cell migration. The presence of the $\alpha 4$ tail (in

comparison to the $\alpha 2$ or $\alpha 5$ tails) leads to increased cell migration, reduced cell spreading, impaired localised and diminished adhesion strengthening⁴³⁹.

It is not implausible that the splicing variant described in section 5.4.2, which is over-transmitted to affected children, could lead to aberrations in the normal development and functioning of the brain.

Another possibility is that those SNPs and haplotypes are in LD with a functional variant elsewhere in the gene, perhaps in intron 16 or intron 17. There is increasing evidence of the importance of introns to influence gene expression. Transcription enhancer and repressor sequences can be located within intronic sequences. Introns are also required for specific modification of some exonic sequences by RNA editing (see Le Hir et al for a review on the role of introns in gene expression⁴⁴⁰).

Expression of ITGA4 in the brain has been localised to the hippocampus and the piriform and entorhinal cortex⁴¹⁵. These structures form part of the limbic system, which is a region of the brain that is implicated in social learning. Deficits in social learning have been linked to autism⁴⁴¹. There have been a number of reports of abnormalities (reduced cell size and increased cell packing density) associated with autism in the hippocampus and to a lesser extent in the entorhinal cortex (see Palmén et al for a full review of neuropathological findings³⁵⁶). Integrin proteins are a requirement of hippocampal synaptic plasticity and spatial memory, and are also responsible for modulating fast excitatory transmission at synapses in this region of the brain^{442 443} (see Table 5.11). Integrin proteins also have a reported role in epilepsy, a co-morbid condition associated with autism (see Section 4.1.3). Experimentation with integrin matrix interactions have shown that disruption of these interactions results in an increased spontaneous bursting period of *in vitro* kindling in the hippocampus (kindling a model of epileptogenesis, whereby repeated exposure to subconvulsive stimuli eventually results in spontaneous electrical discharges)⁴⁴⁴.

There has also been increasing evidence for ITGA4's possible role in immunological reactions in the brain. A study using rats suffering transient focal cerebral ischemia reported that blocking $\alpha 4$ integrin action, led to a decrease in the infarct size, even if administered after the onset of ischemia ⁴⁴⁵. Lupus prone mice, in which an autoimmune reaction leads to leukocyte recruitment throughout the body including the brain, showed that blocking $\alpha 4$ integrin action completely eliminated P-selectin-independent leukocyte rolling, while also significant inhibition of leukocyte adhesion occurred ⁴⁴⁶. Perhaps more importantly, is a study using a mouse model of encephalomyelitis, showed a significant decrease in the number of VLA-4+, CD4(+), T-cells and macrophages in the spinal cord following the administration of anti-sense oligos specific for $\alpha 4$ integrin ⁴²⁶. Hydrocephalus is a condition that occurs more frequently in individuals with autism (see Kielinen et al for further details ⁴⁴⁷). These and other immune reactions in the brain (see Ashwood et al for a review ⁴³⁰) may be in part responsible for the development of the aetiology of autism.

ITGA4 has a number of functions within the brain, and alteration of the normal expression and formation of the protein, may be association with autism.

Name	Year	Integrin subunit	Animal model	Effects
Chan C-S et al ⁴⁴²	2003	$\alpha 3$, $\alpha 5$ and $\alpha 8$	Mouse	Reduced expression leads to defective hippocampal LTP and spatial memory
Dulabon L et al ⁴²¹	2000	$\alpha 3\beta 1$	Mouse	$\alpha 3\beta 1$ binds to Reelin
Emsley JG & Haag T ⁴⁴⁸	2003	$\alpha 6\beta 1$	Mouse	Neutralizing antibodies against the $\alpha 6\beta 1$ int subunit. Inhibition of $\beta 1$ int subunit causes disruption of cohesive nature of rostral migratory system.
Fasen K et al ⁴⁴⁹	2003	$\alpha 1$, $\alpha 3$, $\alpha 4$, $\alpha 5$, $\beta 1$, $\beta 3$ and $\beta 4$	Rat	After induced Status Epilepticus, strong immunoreactivity for $\alpha 1$, $\alpha 2$, $\alpha 4$, $\alpha 5$, $\beta 1$, $\beta 3$ and $\beta 4$ was observed in reactive astrocytes
Gall CM et al ⁴³⁸	2003	RGD-binding integrins	Rat tissue culture	Integrin engagement regulates expression of a subset of growth related genes, at least in part through calcium influx
Georges-Labouesse E et al ⁴⁵⁰	1998	$\alpha 6$	Mouse	Abnormalities in laminar organization of the developing cerebral cortex and retina
Grooms SY and Jones LS ⁴⁴⁴	1997	RGDS - integrin antagonist	Rat	Disruption of integrin matrix interactions increases spontaneous bursting period of in vitro kindling
Kil SH et al ⁴¹⁹	1998	$\alpha 4$	Chicken and mouse	In the presence of an $\alpha 4$ blocking antibody, there was significant reduction in level of neural crest cell migration
Mercado ML et al ⁴⁵¹	2004	$\alpha 7\beta 1$		$\alpha 7\beta 1$ encourages cerebellar granule neurites to grow longer
Rohrbough J et al ⁴²³	2000	Volado	Drosophila	Volado regulates functional synaptic plasticity processes
Vogelezang MG et al ⁴³⁹	2001	$\alpha 4$	Mouse	$\alpha 4$ expressed during peripheral nerve regeneration. A4 cytoplasmic tail increased cell migration but decreased cell spreading and diminished adhesion strengthening under conditions of shear flow and impaired localization of $\alpha 4\beta 1$ integrin into focal adhesion complexes
Werner A et al ⁴²⁰	2000	$\alpha 7$	Mouse	$\alpha 7$ null mice show decreased rate of axonal outgrowth and a delay in the re-innervation of the whisker-pad, a peripheral target of facial mono-neurons

Table 5.11 ROLE OF INTEGRIN PROTEINS IN BRAIN. This is a list of some of the published papers in which the role of integrin proteins in development and neuron processes has been reported.

5.4.6 The Use of HapMap Tagging SNPs

The use of haplotype tagging SNPS (htSNPs) is one proposed way of increasing genotyping efficiency through genotyping those markers that best describe a haploblock. In the case of the ITGA4 gene in the Irish population, there are 3 haploblocks. In the first haploblock, there were 6 SNPs, of which 5 were tagged to be htSNPs. The genotyping of these 5 SNPs in comparison to the complete set, caused no association information to be lost. The association with rs1449263 and the trend towards association with the first 2-marker haplotype (rs1449263-rs3770137) was still identified (p value = 0.0676) and would have been followed up in a larger sample.

The second haploblock only contained 2 SNPs, rs155096 and rs155100, and both of these would also have been genotyped in a follow up sample, leading once again to no loss of association information.

The largest haploblock was in the 3' region of the gene. It spanned 50kb, and included 8 SNPs. Only two of these, rs3770115 and rs1551031, were chosen as htSNPs by the Haploview programme. TDT analysis of these single markers would have led to the identification of a 'region of interest' (p values of 0.112 and 0.063 respectively). Surprisingly, the 2-marker haplotype containing these two markers did not show any strong evidence of association (p value = 0.2307). Based on the TDT results, a follow up study would include fine mapping this haploblock with an increased density, which would highlight the regions of interest.

The choice of htSNPs also varied from sample to sample. The choice of htSNPs in block 1 was identical in both the Irish control and the Vanderbilt "test" samples. However in the case of the AGRE sample, one less SNP (SNP 6, rs1449260) was included (see Figures 5.16, 5.17 and 5.18). In the case of the final block, different SNPs were chosen in each population. In the Irish population, SNPs 12 and 14 were

selected, in the AGRE sample SNPs 14 and 16 were chosen, while due to the breakdown of LD within the Vanderbilt sample, and the lack of a comparable block 3 (see Figure 5.15), SNPs 14, 15, 16 and 17 were chosen. Although there may have been differences in the choice of htSNPs, each selection would have produced results that would have been followed up in the affected sample (see Section 1.4.7).

The evidence produced here indicates that the use of htSNPs is an efficient and economical method of covering a gene of interest. It is important however to ensure that one carefully chooses a correct control sample and determine the minimum level of significance required before a region is excluded from follow up analysis. Other reports have highlighted a few other caveats to take into consideration when using htSNPs. Firstly, htSNPs are most reliable for tagging SNPs that are of equal or greater frequency than the htSNP. Tagging is much less successful when untagged SNPs were rarer than the htSNP itself⁴⁵². This same study by Ke et al also reported that the SNP density, ranging from 1SNP/2.3kb to 1SNP/5kb, did not have a serious impact on the efficiency of describing genetic variation. Almost 80% of htSNPs selected at the 1/5kb map were also chosen in the 1/2.3kb map⁴⁵². However another study has found that mapping a region with a SNP density of 1SNP/10kb only captured 78% of the variation in the region. Even at a lower density of 1SNP/2kb, only 88% of variation was captured⁴⁵³. Meng et al reported that an initial sample size (from which tagging SNPs are selected) of 50 – 100 individuals is necessary⁴⁵⁴. From the results reported above it is also important that htSNPs in one population are not assumed to represent variation in the same region in another population. Haploblocks vary between samples / populations, see Figures 1.8, 1.9, 1.10, 1.15 and 1.16). HtSNPs are generally similar, but not always, for populations with similar haploblock structure, but the choice of htSNPs may differ if the haploblock structure differs.

5.4.7 Conclusions

A few points should be taken into consideration with respect to this study. Firstly the splicing variant (rs12690517) was not genotyped in the 2 independent replication samples due to time and financial constraints. Genotyping of this variant may strengthen the evidence for the splicing variant being partly responsible for the development of autism.

Also, none of the results above have been corrected for multiple testing. However the attempted replication in two independent samples was used to limit the possibility of false positives. As in any report of association with a complex disorder, further replication in other samples is required.

Mutation screening of all exons within the *ITGA4* gene would be beneficial, as the haploblock in the region of the gene showing the most evidence for association contains 18 exons, anyone of which may contain mutations which are in LD with those markers already tested in this sample.

Finally, the effect of the splicing variant on protein structure and function within the CNS should be tested. Further mapping of the regions 5' and 3' of the gene should also be undertaken to further test the haplotype associations that are present with the first and last two-marker haplotypes in the gene.

That said the results presented here show evidence of association with the *ITGA4* gene and autism. Significant association was found to exist between autism and a number of SNPs and haplotypes, especially markers rs1349197, rs3770112 and rs1551031. All these SNPs are located close to exons 16 and 17. Mutational analysis was undertaken and a SNP (rs12690517) was identified in a number of affected probands. This SNP was genotyped and transmissions of the two alleles were tested for association. Significant association was also found with rs12690517 on the exon 16/ intron 16 border (p value = 0.0046, OR = 1.82). This SNP causes a change in the

splice site for this intron / exon boundary. Potential changes in mRNA splicing could be responsible for alterations in brain development or functioning.

Replication in two independent samples was undertaken. Within the AGRE sample, there was also over-transmission of the rs1349197-rs3770112 haplotype (p value = 0.015). However in the Vanderbilt sample, there was a breakdown in LD structure, and there were a number of associations with SNPs in the 3' region of the gene, e.g. rs2305581 (p value = 0.026), although none of the associations were the same as in the Irish or AGRE samples. This may occur for a number of reasons. Firstly, there may be numerous haplotypes associated with autism, as is the case with the dysbindin gene and schizophrenia. Also, due to the differences in LD it could be possible that the associations in all three samples are in LD with the same functional variant, but different markers are picking up the association. Finally it could be a false positive result. However the associations in the AGRE and Vanderbilt samples reported here would indicate that this is not the case.

This study also shows that the use of htSNPs is an economical and efficient method for choosing SNPs for use in large-scale association studies, although care is required to ensure maximum variation is captured.

CHAPTER 6

THE WNT PATHWAY

6.1 Introduction

6.1.1 An Overview of the Wnt Pathway

The Wingless-type MMTV Integration (Wnt) family of proteins is a large family of secreted signalling proteins. Inclusion within this family is determined by sequence similarity instead of functionality. Wnt proteins contain a signal sequence followed by a highly conserved distribution of 23 – 24 cysteines⁴⁵⁵. The Wnt pathway is best known for its role in several developmental processes and oncogenesis (see Logan & Nusse for a review on this topic⁴⁵⁶). Wnt proteins are involved in two pathways. Binding of a Wnt protein to a Frizzled receptor and the subsequent involvement of the β -catenin is known as the canonical Wnt pathway. The non-canonical Wnt pathway does not involve the β -catenin protein.

Following the release or presentation of Wnt proteins they bind to Frizzled receptors on the cell surface. This causes a signal to pass through the cell membrane and transduced to a protein complex that includes Dishevelled (Dsh), glycogen kinase-3 β (GSK-3), Axin, Adenomatous polyposis coli (APC) and β -catenin.

In the absence of Wnt signalling β -catenin is phosphorylated by casein kinase α (CKI α) and glycogen kinase 3 β (GSK3 β). Following phosphorylation, β -catenin is targeted by the proteasome for degradation. In the presence of a Wnt signal, Axin is recruited to a frizzled bound Dsh protein. This acts to remove Axin from the destruction complex, and thus leads to β -catenin stabilisation. As β -catenin levels rise, both intracellularly and within the nucleus, transcription factors (TCF and LEF) become activated, and Wnt target genes are transcribed. This is due to the nuclear accumulation of β -catenin causing the displacement of the *Groucho* protein from the TCF/LEF complex and the recruitment of a histone acetylase complex binding protein. In the absence of Wnt signalling TCF acts as a repressor of target genes

through its interaction with *Groucho*. These proteins associate with histone deacetylases. This association leads to the prevention of nucleosomes binding to the gene's TATA boxes.

There are a number of downstream genes that the Wnt pathway activates, including Engrailed 2, TCF1, LEF1 and Axin2 (see Table 6.1 and Figure 6.1 for further information). A feedback mechanism is also present, i.e. the Wnt signalling system causes other Wnt components to be actively transcribed. The majority of these Wnt signalling components that are transcribed cause the down-regulation of the Wnt pathway, thereby acting as a control to prevent excess Wnt pathway effects.

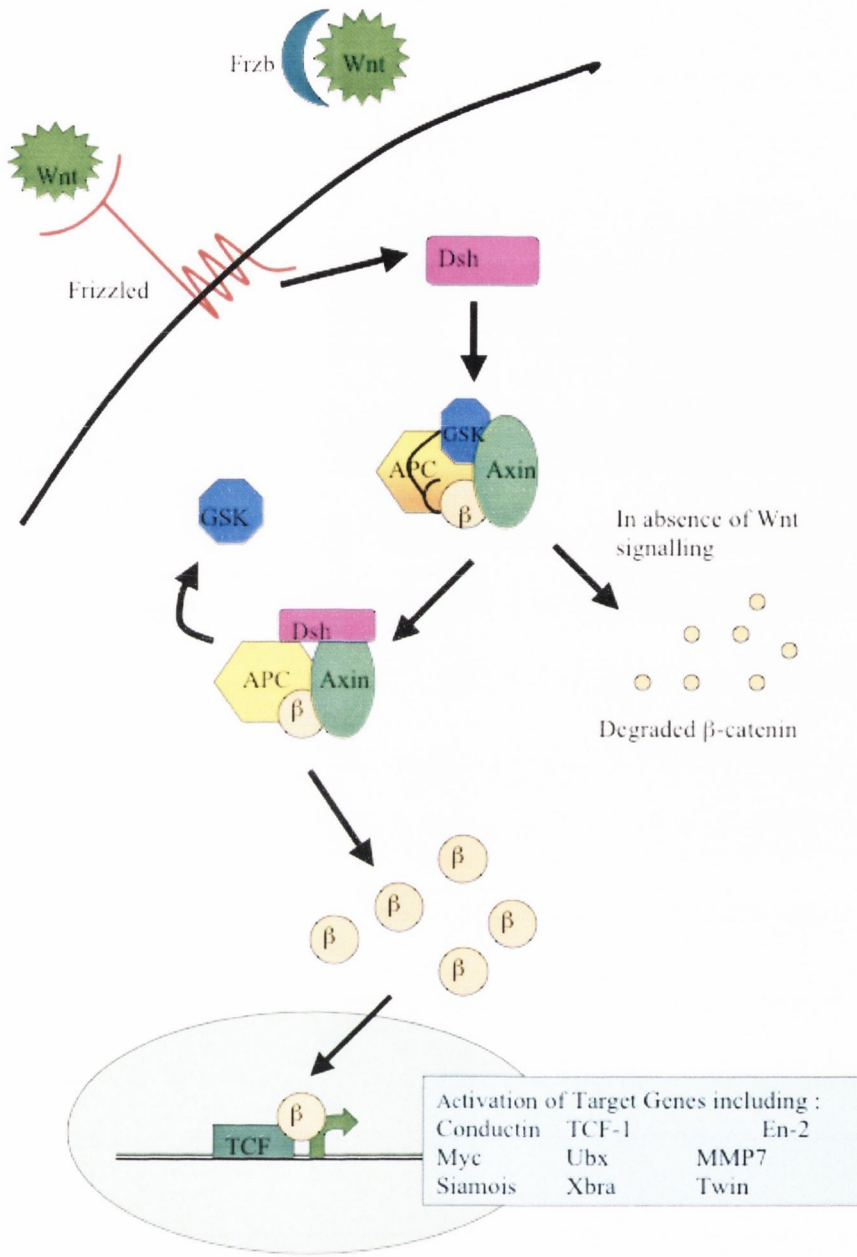


Figure 6.1 CANONICAL WNT SIGNALLING PATHWAY. In the absence of Wnt signalling, β -catenin becomes phosphorylated and degraded. However in the presence of Wnt signalling, β -catenin degradation is inhibited, leading to increased intracellular and nuclear concentrations of β -catenin. This in turn leads to the transcription of target genes. Diagram is adapted from Logan & Nusse, 2004⁴⁵⁶.

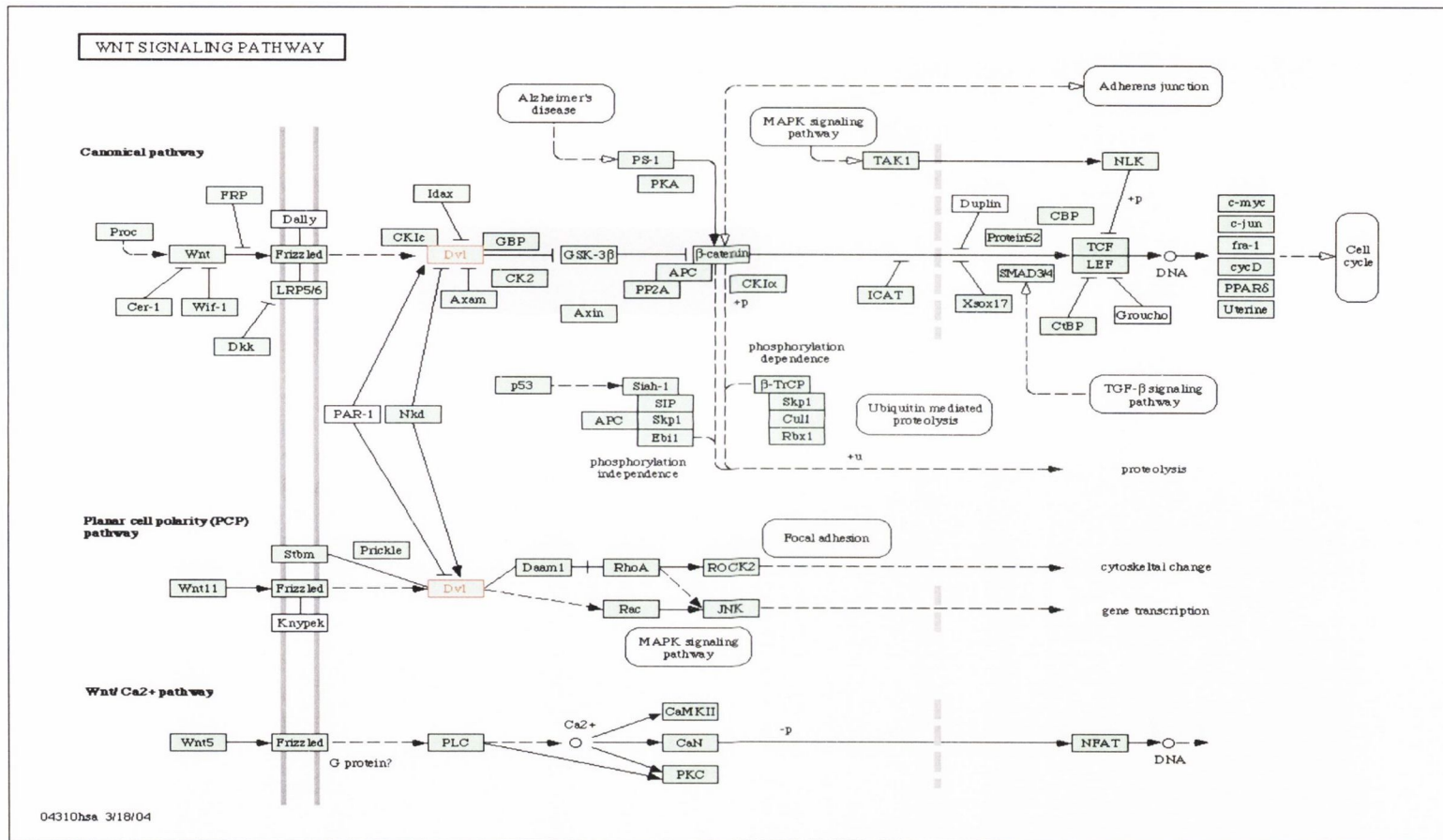


Figure 6.2 WNT CANONICAL AND NON-CANONICAL PATHWAY (from the KEGG database (<http://www.genome.ad.jp/kegg/pathway/hsa/hsa04310.html>)).

Target Gene	Model Organism	Effect of Wnt signal on target gene expression	Effect of changes in target gene expression on Wnt pathway	Target gene interacts with	Reference
<i>Fz</i>	Drosophila	Down	Inactivate	Wnt	Muller et al ⁴⁵⁷
<i>Dfz2</i>	Drosophila	Down	Inactivate	Wnt	Cadigan et al ⁴⁵⁸
<i>Dfz3</i>	Drosophila	Up	Activate	Wnt	Sato et al ⁴⁵⁹
<i>Arrow / LRP</i>	Drosophila	Down	Inactivate	Wnt	Wehrli et al ⁴⁶⁰
<i>naked</i>	Drosophila	Up	Inactivate	Dsh	Rousset et al ⁴⁶¹
<i>Axin2</i>	Mouse	Up	Inactivate	B-catenin	Jho et al ⁴⁶²
<i>Beta-TCRP</i>	Human Cell Study	Up	Inactivate	B-catenin	Spiegelman et al ⁴⁶³
<i>TCF1</i>	Human Cell Study	Up	Inactivate	TCF	Roose et al ⁴⁶⁴
<i>LEF1</i>	Human Cell Study	Down	Activate	B-catenin	Hovanes et al ⁴⁶⁵
<i>Nemo</i>	Drosophila	Up	Inactivate	B-catenin / LEF / TCF	Zeng & Verheyen ⁴⁶⁶
	Zebrafish		Activate		Thorpe & Moon ⁴⁶⁷

Table 6.1 WNT SIGNALLING COMPONENTS AS WNT PATHWAY TARGETS.

Activation of the Wnt pathway leads to the transcription of a number of genes, including those involved in the Wnt pathway itself. As a result, many down-regulators of the Wnt pathway are transcribed, hence leading to a dampening of the Wnt pathway. This table is adapted from Logan and Nusse ⁴⁵⁶.

6.1.2 Expression in the Brain, Animal Models and Autism

The Wnt pathway is best known for its role in development, and cancer, but there is increasing evidence for its the role of the Wnt pathway in brain development and synaptogenesis.

In the development of an embryo, it is necessary for a number of secreted proteins and their receptors to provide information regarding their position within the embryo. This information sets up the polarity of the developing embryo. The Wnt proteins are part of this complicated network of proteins. A small number of epithelial cells secrete *Wg* (the *Drosophila* form of Wnt), and through the mechanisms of endocytosis, cell divisions and transcytosis (the transport of macromolecules across a cell, consisting of endocytosis of a macromolecule at one side of a monolayer and exocytosis at the other side), a gradient of *Wg* protein is set up across every section of the embryo, leading to differential transcription of target genes. The aim of this process is to develop a number of cells with differing identities. More recent investigations have begun to elucidate the role of the Wnt pathway in the brain, and the correct formation of synaptic specialisations.

The first report of a possible role for Wnt proteins in the developing brain was in 1990 ⁴⁶⁸. Targeted disruption of the mouse *int-1* gene (shows 54% amino acid homology to the *Drosophila Wg* protein) resulted in a very high mortality rate of homozygous mutants. Of the 10 original *-/-* births, 6 had died in less than 12 hours. Only 1 survived beyond 36 hours and into adulthood. This one mice exhibited severe ataxia, which was as a consequence of the absence of the cerebellum and a large portion of mesencephalon.

Another Wnt gene, *Wnt3*, is also expressed in the cerebellum. Expression has been localised to the Purkinje cell layer. Abnormalities in these cells have been associated with autism, with such individuals having fewer Purkinje cells in the cerebellum, and those present of a larger size ⁴⁶⁹. Purkinje cells form during embryogenesis, but their

maturation and the formation of synapses continues during postnatal life. The expression of *Wnt3* is linked to the initial development of the cell layer, starting at P6 until it reaches a maximum level at P15, which coincides with phase II of Purkinje cell maturation as defined by Altman⁴⁷⁰. This is the period during which the formation of synaptic contacts between Purkinje cells occurs. Deletion of this gene in mice does not lead to malformation of the Purkinje cell layer however. Instead, abnormalities in axis formation, gastrulation and hair structure are observed^{471,472}. The absence of cerebellar abnormalities highlights the presence of functional redundancy, whereby another Wnt protein may act in place *Wnt3* place, thereby reducing the levels of disruption caused by the absence of one particular Wnt protein. Wnt proteins can be divided and placed in two separate groups depending on their effect on cell transformation (with regards to tumorigenicity); the transformers (WNT1, WNT2, WNT3, WNT3a, WNT7a and WNT7b) and the non-transformers (WNT4, WNT5a, WNT5b and WNT6). This raises the possibility of two distinct signalling mechanisms, with functional redundancy only occurring within the two groups.

Wnt8 gene expression is identical in mouse and human brain. In both cases the expression is restricted to the developing brain, including the telencephalic ventricles (including the developing hippocampus), dorsal thalamus and regions of the posterior hypothalamus (mammillary and retromammillary regions). Interestingly, in humans the *Wnt8* gene localises to chromosome 10q24, a region that has shown linkage in families suffering partial epilepsy, which as stated in section 4.1.3 occurs at a higher frequency in individuals with autism. One genome wide scan in autism has highlighted the 10q region for linkage, although this did not remain significant after corrections for multiple testing⁹⁴.

Mutations in the frizzled receptors likewise cause brain abnormalities. Deletion of the *Fz3* gene causes a number of defects in mice. Variable loss of the corpus callosum is observed in *-/-* mice. The *-/-* mice also exhibit complete loss of the thalamocortical, corticothalamic, and nigrostriatal tracts of the anterior commissure.

The cell death occurs late in the gestation period, and may occur as a result of the almost complete absence of long-range connection. Another possible mechanism causing the neuronal cell death may be the failure of the budding axons to polarize correctly, which leads to abortive axonal outgrowth⁴⁷⁴. Knockout mice for the *Fz4* gene also show brain abnormalities, in addition to inner ear and oesophagus defects. These abnormalities are only apparent in the postnatal life of the *Fz4* *-/-* mice (after 3rd postnatal week). Before this time cerebellar cell proliferation, migration and arborization continue as normal. After three weeks however, granule cell death starts to occur, and this is eventually followed by Purkinje cell death⁴⁷⁵.

A downstream target of the Wnt pathway is the Engrailed 2 gene (*En2*). *En2* is a homeobox gene that plays an important role in segmentation processes, neurogenesis and neuronal differentiation⁴⁷⁶. Deletion of the *En2* gene causes cerebellar abnormalities to occur during embryogenesis. Delays in cerebellar fusion in mice are observed by 15.5 d.p.c (days post coitum), in addition to a decrease in the size of the cerebellum. Cerebellum abnormalities continue to develop postnatally, especially concerning fissure formation. Close examination of the cerebellum reveals the loss of Purkinje cells, with a decrease in the number of granule cells⁴⁷⁷. Surprisingly, over-expression of the *En2* gene also leads to disruption of normal Purkinje cell development. When the *En2* gene was ectopically expressed during late embryogenesis and postnatally, a 40% reduction in the number of Purkinje cells was observed in the cerebellum. This loss was distributed uniformly across the mediolateral axis. However, cells with the cerebellar fissures were most sensitive to this loss. The remaining cells appeared biochemically and morphologically normal, indicating that the majority (60%) of Purkinje cells remain impervious to the persistent expression of *En2*⁴⁷⁸. The *En1* (2q13) and *En2* (7q36) genes are 64% homologous, and therefore it is not surprising that knockout mice for the *En1* gene closely resemble *En2* knockouts. The mid-hindbrain in *En1* *-/-* mutants is the most severely affected but both animals show a decrease in cerebellum size, while *En1* *-/-* mice also show defects in limb, sternum and rib development. *En2* *-/-* mutants exhibit a greater decrease in size⁴⁷⁹. The difference in severity (i.e. size decrease)

may arise for two reasons. Firstly, the two engrailed proteins may be functionally equivalent, with the *En1* mutant phenotype being attributed to defects arising at early embryonic stages, when *En1* but not *En2* is not expressed. The other alternative is the two proteins are not functionally equivalent. *En1* may be responsible for the development and generation in mid-hindbrain cells, and *En2* may function later in cerebellum patterning. Expression of engrailed genes is maintained into adult life in the mid and hindbrain, and is crucial to the survival of mesencephalic dopaminergic (MeDa) neurons. Surprisingly, deletion of the engrailed genes (both *En1* and *En2*) does not prevent the initial development of the MeDa neurons. Embryos start to develop normally, with typical brain development including the development of MeDa neurons. However in the absence of normal engrailed expression the MeDa neurons die by apoptosis, leading to a large deletion of the mid and hindbrain structures and eventual death of mice after birth ⁴⁷⁶. The effects of lack of *En2* expression on a single cockroach neuron was examined by Marie et al ⁴⁸⁰. The effects of *En2* knockout were similar whether RNAi was added before or after the development of the neuron, thereby showing the *En2* functions post-mitotically. Abnormalities in axonal anatomy were observed, illustrating the importance of *En2* to the control of axonal branching patterns ⁴⁸⁰.

Deletion of the dishevelled (*dvl*) in mice does not lead to major brain abnormalities. Mice homozygous for the deletion were normal in development, growth, size and appearance. Deletion of the gene was not unremarkable however. The adults showed abnormal social behaviour ⁴⁸¹. Whisker trimming is a social behaviour, which is common to both male and female mice. *Dvl* *-/-* mice had full sets of whiskers, and this absence of normal barbering was also associated with less social interaction. Homozygous mutant mice interacted with other mice less often, slept in random scattered patterns (in contrast to huddling like wild-type mice), and built shallower nests. Sensorimotor gating was also reported to be affected by the *dvl* mutation, with *-/-* mice showing significantly lower levels of prepulse inhibition and responses to acoustic startle stimuli ⁴⁸¹. These results are an illustration of a gene that is involved in complex social behaviours, and defects in social interaction is a

definable trait in autism. The importance of this fact highlights that developmental gene defects can cause subtle abnormalities in complex social behaviour and not just major structural brain abnormalities.

All these animal models exhibit alterations to the normal phenotype, which may be related to the autism phenotype. Numerous studies have highlighted Purkinje cell abnormalities in the cerebellums of affected individuals, with patients showing a decreased number of Purkinje cells (see Palmen et al for a full review) ³⁵⁶. Other reported cerebellar abnormalities include decreased nicotinic receptor levels in patients compared to controls ⁴⁸². Similarly alterations in the limbic system have also been reported. A detailed review of several case reports based on the analysis of the brain sections of six autistic individuals, highlighted that all 6 cases showed “increased cell packing density and reduced cell size in hippocampus, subiculum and amygdala, and although to a lesser extent, in entorhinal cortex, mammillary bodies and septal nuclei” ⁴⁸³. Taken together, there is sufficient animal model and neuropathological evidence to investigate the role that the Wnt pathway plays in determining the development of the autistic aetiology.

6.1.3 The Wnt2 Gene

Wnt genes are defined as genes with sequence homology to the homologues *Wnt-1* (in the mouse) and *wg* (in *Drosophila*). The minimum homology required is 18% (mean = 35%), but conservation of a pattern of 23-24 cysteines is paramount for inclusion in the Wnt family. 19 Wnt genes have been identified in the human to date (see Miller et al, 2001 for further details) ⁴⁸⁴. *Wnt2* is located on chromosome 7q31, is 46.06 Kb in length with five exons ⁴⁸⁵. Transcription of the gene produces a protein of 360 amino acids long.

The *Wnt2* gene was chosen as a model Wnt gene for analysis in the Wnt pathway for a number of reasons. Firstly, it is located on chromosome 7q31, a region in which there has been a number of linkage reports in autism samples^{78 90 77,84,89}. Secondly, there have been a number of reports of cytogenetic abnormalities reported on this region of 7q, including one report in which a translocation breakpoint occurred in the gene (*RAY1*) proximal to the *Wnt2* gene^{113,114,486,487}. It is also a protein expressed in the brain, mainly in the thalamus region²⁶³. Finally an association with autism and *Wnt2* had been reported by Wassink et al²⁶³. Wassink et al undertook a mutation screening experiment, and found 3 mutations (2 non-conservative missense mutations, 1 synonymous codon change), two of which were only transmitted to affected individuals. One mutation (a C to T transition) caused an arginine, at position 299, to change to tryptophan. This mutation was found in a father and his two affected sons. The mutation was not found in the mother and was not transmitted to the two unaffected sons. It was also observed that the father did not fit the criteria for autism, but showed deficits in childhood conversation. His reading and speech were also impaired. A second mutation, in the signalling domain of exon 1, caused a leucine to arginine change (Leu5Arg). This mutation was found in one mother and her two affected children, but not the father. Both of these mutations caused amino acid changes in regions of the gene that are highly conserved, and therefore likely to be of functional significance to the protein. Wassink et al also examined the transmission of two markers, one in the 5'UTR and one in the 3'UTR, to probands diagnosed with autism. Transmission of the T allele of the 3'UTR SNP was found to be significantly greater than the transmission of the C allele, leading to a likelihood ratio of 22 (equivalent to χ^2 value of 6.2, and corresponding p value of 0.013).

6.1.4 The Frzb Gene

The *Frzb* gene lies on chromosome 2q31.2. It is 32.2 kb long, with 6 exons. Transcription of the gene produces a 36.3kDa hydrophilic protein of 325 amino

acids. Comparison of its amino acid sequence reveals that it has 4 domains, including 2 hydrophobic stretches in the 5' region of the gene, one of which is hypothesised to be a signal domain, and a stretch of amino acids that contains several potential serine/threonine phosphorylation sites. The Frzb protein also has a cysteines-rich amino terminal domain, which shows similarity to the putative wnt binding domain of frizzled receptors⁴⁸⁸. Experimentation by Wang et al and Leys et al showed that Frzb has the ability to interact directly with Wnt proteins (Xwnt8 and Wnt1), acting as a Wnt antagonist and thereby blocking Wnt signalling^{489,490}.

A deletion study undertaken by Lin et al revealed that it is the cysteines rich domain of the protein that is required and sufficient for interaction with Wnt proteins⁴⁹⁰. Deletion of this domain led to the loss of inhibitory action of Wnt-1 resulting in the induction of *Xenopus* embryonic axis duplication. This region of the protein has a predominantly alpha helical structure, with the cysteines forming disulfide bonds⁴⁹¹.

Frzb is expressed in a number of organs in the human. Its expression is strongest in the placenta and heart, with moderate expression in the brain, skeletal muscle and pancreas. It exhibits low level of expression in the lung and liver⁴⁹². In *Xenopus* embryos, *frzb* is also expressed during gastrulation in a region of the embryo called Spemann's organiser. This region induces the formation of a new body axis, complete with notochord, neural structures and muscle. Given the expression of *Frzb* in the Spemann's organiser, it is possible that *frzb* is involved in modulating neural cell fate or endodermal cell fate in addition to its likely roles in the mesoderm^{489,493}.

The *Frzb* gene was chosen to be analysed due to its location on chromosome 2q. It is one of a number of genes, which has shown to be contained in a region in a patient in our study, patient KM, that has been translocated to chromosome 9q31⁴⁰⁷. There have been a number of linkage studies reporting association with this region (see Section 5.1.1 for further details). As an antagonist to Wnt activity, it is also worthy of investigation.

6.1.5 The Dishevelled Gene

The dishevelled gene (*DVL*) has been mapped to 1p36, is 14.102kb in length with 15 exons ⁴⁹⁴. The DVL protein is 670 amino acids in length and has a molecular weight of 72.893kDa. The human form of the protein is 92% identical to the mouse *Dvl* gene. Like its *Drosophila* counterpart, it encodes a cytoplasmic phosphoprotein that has a role in cell proliferation, and developmental regulation. Pizzutti et al found expression of this gene to be strongest in adult skeletal muscle and pancreas. Moderate expression was also reported in the heart, with expression noted in the adult brain, placenta, lung, liver and kidney ⁴⁹⁴.

Expression in the brain was further examined by Rosso et al ⁴⁹⁵. Expression was reported in mouse hippocampal neurons, especially within the cell bodies and neurites of the neurons. *Dvl* expressing neurons were shown to have a more elaborate dendritic arborization pattern compared to controls. Deletion of the PDZ region (a protein-protein interaction domain) of the *Dvl* gene led to fewer secondary and tertiary axonal branches in comparison to controls. Examination of the hippocampal neurons in the *Dvl* *-/-* mice with social behavioural problems revealed that these mice had mild defects in primary axon branching and substantial defects in secondary and tertiary axonal branching. The social problems observed in these mice included an absence of whisker trimming and barbering, decreased social interaction with other mice, and abnormal sleeping and nesting behaviour (the knockout mice preferring to sleep in scattered random patterns in shallow nests as opposed to huddling in deep nests with other mice) ⁴⁸¹. Rosso et al concluded that *Dvl* plays an important role in dendritic arborization, and sends its signal through the Wnt non-canonical pathway via Ras-related C3 botulinum toxin substrate (Rac) and the jun-terminal kinase (JNK) ⁴⁹⁵.

The *Dvl* gene makes a good candidate for autism based on a number of observations. Firstly, the knockout model for *Dvl* shows social and behavioural characteristics that

may be similar to the autistic phenotype⁴⁸¹. Decreased dendritic branching has been reported in the hippocampus in an individual with autism⁴⁹⁶.

6.1.6 The EN2 Gene

The Engrailed 2 (*EN2*) gene is a small gene located on chromosome 7q36.2^{497,498}. It is 2 exons long with one intron, and is 3.3kb in length. The protein produced by transcription is a homeobox protein 333 amino acids in length with a molecular weight of 34210 Da. The human form of *EN2* is 100% identical to the mouse form, indicating that evolution has acted to conserve the sequence of this gene. Expression of the gene has been reported in both foetal and adult brain⁴⁹⁸. In *Drosophila* the *EN2* protein acts as a developmental gene involved in the segmentation process, where it is required for the formation of the posterior compartment.

EN2 is one of the downstream target genes activated by the Wnt pathway⁴⁹⁹. Its inclusion in this study of the Wnt pathway was based on a number of criteria. Firstly, an early study reported the association of the *EN2* gene with autism in a sample of 100 French patients suffering autism²¹⁸. Secondly deletion of this gene in mice leads to cerebellar abnormalities, including a decreased number of neurons in the cerebellum and gross structural changes in the cerebellar structure^{477 520}. These both show similarities to the neuroanatomical observations in individuals with autism (see Section 1.1.4). Finally, there has been a number of reports of suggestive evidence for linkage to this region of chromosome 7^{82 79,83,191}. All these reasons make *EN2* an interesting candidate gene for autism within the Wnt pathway.

6.1.7 Aims

The aims of this study were two fold.

- 1) Firstly, to attempt to replicate the findings of Wassink et al ²⁶³, by screening for mutations within the gene, and testing the transmission of a number of variants to affected individuals. An attempted replication of Petit et al association between autism and the *EN2* gene will also be undertaken ²¹⁸.
- 2) Secondly, to test known polymorphisms for association with autism in 4 genes involved in the Wnt pathway, *WNT2*, *FRZB*, *DVL* and *EN2*.

6.2 Material and Methods

6.2.1 Mutation Screen of Wnt2 Gene

All affected probands were screened by SSCP (see Section 2.2.5.1) for the mutations reported in exons 1 and 5 by Wassink et al ²⁶³. The fragments spanning the mutation were amplified by PCR, for conditions and primer information please refer to Appendix II, sections 1.1 and 1.2.

6.2.2 Genotyping of SNPs

Three *WNT2* SNPS were commercially genotyped by K-bioscience. They included a SNP 219bp 5' of the gene (rs39315), a SNP in intron 3 (rs2285545) and a SNP in the 3'UTR region of the gene (rs2024233).

Eight *FRZB* SNPs were chosen to span the gene. The choice of SNPS was based primarily on whether they caused a non-synonymous change, and whether the polymorphism had been confirmed sufficiently common in other populations. Two SNPs fulfilling this criterion were found (rs7775 and rs288326). Rs288326 is a G/A SNP, which leads to a tryptophan amino acid changing to an arginine (amino acid number 200). Its heterozygosity value is 0.145518. The SNP rs7775 causes amino acid 324 to change from a Glycine to Arginine as a result of a G to C nucleotide change and has a heterozygosity value of 0.18795. The remaining 7 SNPs (rs288324, rs2242070, rs722738, rs2118590, rs6433993 and rs10206992) are all intronic SNPs, with an average inter-marker distance of 3868bp and average heterozygosity value of 0.424. All of the above markers were genotyped commercially by K-bioscience.

Four *DVL* SNPs were selected for allelic transmission analysis. They included a SNP that mapped within 2kb of the 5' of the gene (rs307354), and three intronic SNPs (rs307371, rs307359 and rs307356). The average inter-marker distance was ~5.6 kb, with an average heterozygosity value of 0.27. PCR and genotyping information can be found in Appendix II, sections 1.1 and 1.2.

Five *EN2* SNPs were genotyped for association with the *EN2* gene with autism. They included the four SNPs reported in the Gharani paper (rs3735653, rs1861972, rs1861973 and rs2361689)¹⁹¹. An additional 2 SNPs were included, one 383bp 5' of the gene (rs3757846), and another, rs1861958, which was 399bp 3' of the gene. All markers were genotyped using the SNaPShot method of allele discrimination (see Section 2.2.4.3, and Appendix II, sections 1.1. and 1.2, for information).

6.2.3 STATISTICS

All markers were tested for Hardy-Weinberg equilibrium. Deviation from expected allelic and haplotype transmissions were tested using the TDT test (See Section 2.3.1.2) and UNPHASED (see Section 2.3.1.3.2) respectively.

6.3 Results

6.3.1 WNT2 Gene

All markers were tested for and found to be in Hardy-Weinberg equilibrium.

6.3.1.1 LD STRUCTURE

The extent of linkage disequilibrium was measured by D' values. Moderate LD was found to exist between rs39315 and rs2285545 ($D' = 0.76$). No LD was found to exist between rs2285545 and rs2024233 ($D' = 0.09$) (see Figure 6.3).

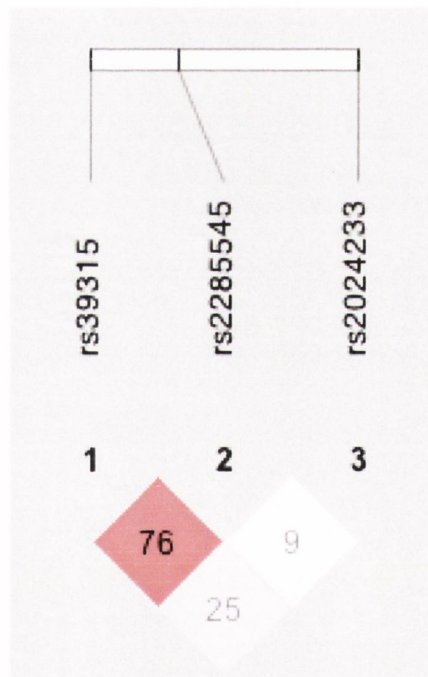


Figure 6.3 LD MEASUREMENTS WITHIN THE *WNT2* GENE. LD measurements are in the form of D' values, and are measured by the programme, Haploview. Moderate to strong LD was found to exist between the marker 5' to the gene and the intron 3 marker (rs2285545) (markers 1 and 2). The marker in the 3'UTR of the gene was not in LD with either of the two remaining markers.

6.3.1.2 TDT ANALYSIS

Screening of exons 1 and 5 of the gene did not reveal the presence of any mutations. TDT testing of the three markers, rs39315 (219bp 5' of gene), rs2285545 (intron 3) and rs2024233 (3'UTR), revealed over-transmission of the C allele of marker rs2024233. The G allele was transmitted 71 times in comparison to 50 transmissions of the A allele. Transmission of alleles from this marker produced a χ^2 value of 3.663, and a p value of 0.0556, which did not reach the level of statistical significance (p-value > 0.05), see Table 6.2 for further details. Another trend towards significance was observed with the marker 5' of the gene, rs393315, with 79 transmissions of the C allele compared to just 63 transmissions of the T allele. This produced a χ^2 value of 1.807 and p value of 0.1789. No distortions in allelic transmissions were observed for the intronic marker, rs2285545 (p value = 0.7257).

Marker	Allele	Transmitted	Not Transmitted	OR	χ^2	p value
rs39315	C	79	63	1.25	1.807	0.1789
rs228545	C	67	63	1.06	0.1231	0.7357
rs2024233	G	71	50	1.42	3.663	0.0556

Table 6.2 RESULTS OF TDT TESTING OF *WNT2* ALLELIC VARIANTS. Markers showing a trend towards significance (at the p < 0.05 level) are highlighted in light yellow.

6.3.1.3 HAPLOTYPE ANALYSIS

Transmission of haplotypes was also investigated using the UNPHASED programme. No significant transmissions of any haplotypes were observed, see Table 6.3 for more detail. However, there was a trend towards increased transmission of the C-A haplotype of markers rs39315 and rs2285545, leading to an OR of 2.5 and p value of 0.0784. The global χ^2 and p value for the above marker did not reach the threshold for significance.

Markers	Alleles showing excess transmission	Transmitted	Not Transmitted	Global χ^2	Global p-value
rs39315- rs2285545	C-A	15	6	4.294	0.2314
rs2285545 rs2024233	A-G	36	23	4.7596	0.1903

Table 6.3 RESULTS OF HAPLOTYPE TESTING OF *WNT2* VARIANTS. There was no significant distortion in allelic transmissions with the markers genotyped in the *Wnt2* gene. Haplotype testing was carried out using the UNPHASED programme. Details for the haplotypes with the highest ORs are presented in this table.

6.3.2 Frzb Gene

All markers were tested and found to be in Hardy-Weinberg Equilibrium.

6.3.2.1 LD Measurements

D' values were calculated and haploblocks (according to the definition of a haploblock as stated by Gabriel et al ⁵⁰¹) were constructed using the Haploview programme. High levels of LD were found to exist within the 5' region of the gene (see Figure 6.4). A haploblock comprising of 5 markers, ranging from the beginning of intron 1 to the beginning of intron 3, of length 23kb was found to exist. High LD was also shown to exist between rs2242070 (in intron 3) and rs288326 (in exon 4), with a D' value of 0.91. High LD was also apparent between rs288326 and rs288324 (in intron 5) (D' = 0.95). The final marker in the gene, in exon 6, was found not to be in LD with any of the previously mentioned markers (see Figure 6.4).

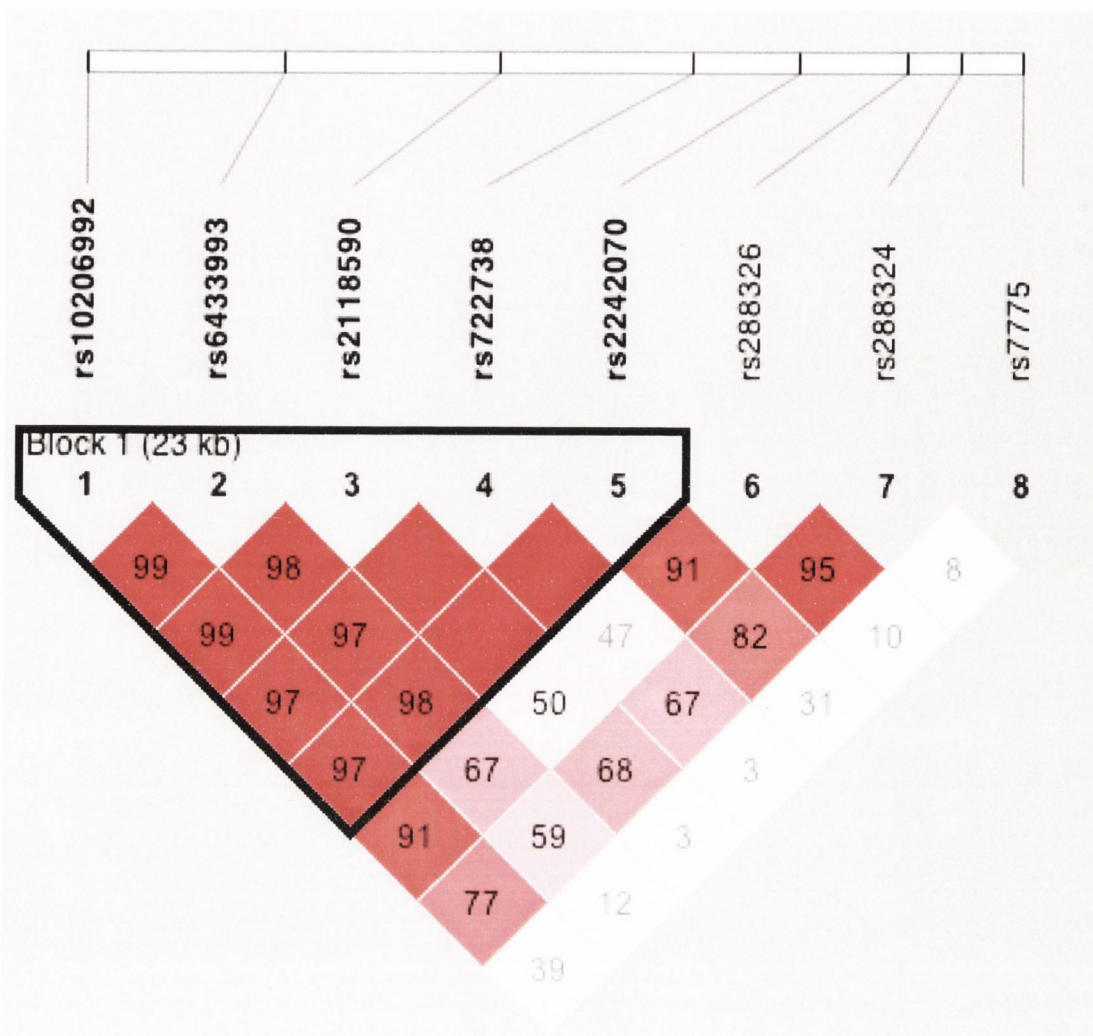


Figure 6.4 MEASURES OF LD AND PRESENCE OF HAPLOBLOCKS WITHIN THE *FRZB* GENE. LD is measured using the D' statistic and calculated using the Haploview programme.

6.3.2.2 TDT Analysis

All markers were tested for significant distortions in allelic transmissions. A significant trend towards over-transmission of the G allele of marker rs288326 was observed ($\chi^2 = 3.994$, OR = 1.71, p value = 0.04566). This is a synonymous transition in exon 3 that leads to the production of an arginine residue (G allele), a basic amino acid, or a tryptophan residue (A allele), an aromatic amino acid.

Increased transmission of the G allele may therefore be of functional significance and related to the development of the autistic phenotype. The marker located proximal to this (rs2242070), located at the end of intron 3, also showed significant association with autism. There was increased transmission of the G allele to affected probands (69 transmissions versus 41 non-transmissions), leading to a χ^2 value of 7.206, OR of 1.68 and corresponding p value of 0.007265. The remaining markers showed no evidence of over-transmission to probands, see Table 6.4 for more details.

Marker	Allele	Transmitted	Not Transmitted	OR	χ^2	p value
rs1020699	G	59	51	1.16	0.5823	0.4454
rs6433993	G	63	49	1.29	1.755	0.1853
rs2118590	C	68	67	1.01	0.007	0.9314
rs722738	C	63	55	1.15	0.5428	0.4613
rs2242070	G	69	41	1.68	7.206	0.0073
rs288326	G	36	21	1.71	3.994	0.0457
rs288324	G	65	61	1.07	0.127	0.7216
rs7775	G	19	14	1.35	0.7605	0.3832

Table 6.4 TDT TESTING OF *FRZB* MARKERS. Those markers highlighted in yellow and bold show significant association with autism. The ORs calculated per marker are for the over-transmitted allele in each case.

6.3.2.3 Transmission of *Frzb* Haplotypes

Transmission distortion of haplotypes was tested using the UNPHASED programme. Two haplotypes showed significant transmission distortion to affected probands. In a fashion similar to the TDT results, the two significant haplotypes contained

rs2242070 and rs288326. Over-transmission of the C-G haplotype of markers rs722738 and rs2242070, spanning exon 3 showed significant association with the probands, $\chi^2 = 10.28$, OR = 1.86 and p value of 0.0163. The over-transmission of the haplotype from rs2242070 to rs288326, which includes most of intron 3 and exon4, was significantly associated with autism also. Increased transmission of the G-G form of the haplotype resulted in an OR of 2.29 for individuals with autism. The global χ^2 of this haplotype is 14.92, with a corresponding p value of 0.0019. The transmission of all other haplotypes was deemed to be non-significant (see Table 6.5).

Markers	Alleles showing excess transmission	Transmitted	Not Transmitted	OR	Global χ^2	Global p-value
rs10206992-rs6433993	T-A	39	31	1.26	2.595	0.4584
rs6433993-rs2118590	G-T	42	32	1.31	2.061	0.3567
rs2118590-rs722738	C-C	47	36	1.31	2.267	0.322
rs722738-rs2242070	C-G	39	21	1.86	10.28	0.0163
rs2242070-rs288326	G-G	71	31	2.29	14.92	0.0019
rs288326-rs288324	G-G	50	41	1.22	4.758	0.1904
rs288324-rs7775	A-G	9	4	2.25	2.015	0.5692

Table 6.5 TRANSMISSION OF *FRZB* HAPLOTYPES. Transmission of haplotypes was tested using UNPHASED. Significant Transmission are highlighted in yellow and bold, and involve the two markers that showed significant transmission in the TDT test, i.e. rs2242070 and rs288326. Details are given in this table for the allelic forms of the haplotype that generated the highest OR.

6.3.3 DISHEVELLED GENE

All markers were found to be in Hardy-Weinberg equilibrium.

6.3.3.1 LD STRUCTURE

The strength of LD was measured throughout the gene. Strong LD was observed from the region covered by markers rs307356 (in intron 1) to rs307354 (located 6.12kb distal to the gene) (see Figure 6.5). D' values range from 0.67, between rs307359 and rs307354, to the markers being in complete linkage disequilibrium (markers rs307356 –rs307354, and rs307356-rs307359). There wasn't the occurrence of a haploblock as defined by Gabriel et al ⁵⁰¹ as the lower boundary for a 95% confidence interval must be 0.7. In this case it was 0.67, see Figure 6.5 for further information.

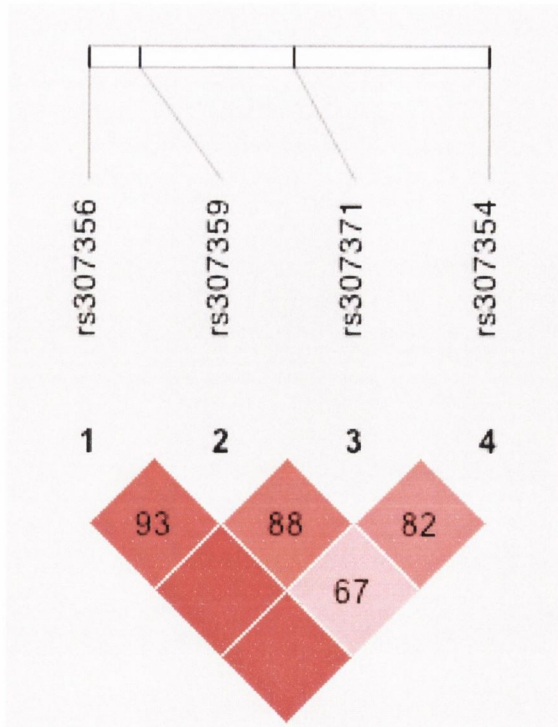


Figure 6.5 LD MEASUREMENTS WITHIN *DVL* GENE. Strong LD is found to exist throughout the gene. LD was measured using the D' statistic in the programme Haploview.

6.3.3.2 TDT TESTING

Allelic transmissions were analysed by the TDT test, while transmission of haplotypes were analysed by the UNPHASED test. TDT testing revealed that the G allele of marker rs307354 is significantly over transmitted to probands with autism, producing a χ^2 value of 5.28 and p value of 0.02155 (see Table 6.6 for further information). No further distortions in allelic transmission were uncovered.

Marker	Allele	Transmitted	Not Transmitted	χ^2	OR	p value
rs307356	C	10	6	1.011	1.67	0.3147
rs307359	G	18	13	0.81	1.38	0.3681
rs307371	C	12	11	0.044	1.09	0.8348
rs307354	G	26	12	5.281	2.17	0.0216

Table 6.6 RESULTS OF TDT TESTING OF *DVL* ALLELIC VARIANTS. One marker in the dishevelled gene (rs307354) showed significantly increased transmission of the G allele to affected offspring, highlighted in yellow and bold.

6.3.3.3 TRANSMISSION OF HAPLOTYPES

Transmission of two marker haplotypes was also investigated. Excess transmission of a G-C haplotype (markers rs307354-rs307371) was found to occur, leading to a χ^2 of 3.642 (p value = 0.03).

Markers	Alleles showing excess transmission	Transmitted	Not Transmitted	OR	Global χ^2	Global p-value
rs307356-rs307359	C-G	17	9	1.89	4.901	0.17
rs307359-rs307371	C-G	13	7	1.86	8.318	0.0398
rs307371-rs307354	C-G	15	9	1.67	3.642	0.3027

Table 6.7 RESULTS OF HAPLOTYPE TESTING IN *DVL* GENE. Distorted transmission of rs307359-rs307371 is reported here. Haplotype testing was carried out using the UNPHASED programme.

6.3.4 EN2 GENE

All markers were tested for Hardy-Weinberg Equilibrium. Despite numerous attempts to amplify and genotype marker rs3735653, it remained in Hardy-Weinberg disequilibrium (p value = 0.0282), with an increased frequency of TT homozygotes and decreased frequency of CT heterozygotes. It was therefore excluded from all subsequent analysis.

6.3.4.1 LD Structure

The strength of LD throughout the *EN2* gene was low. Only two markers, both located only 152bp apart, were shown to be in strong LD (see Figure 6.6). The remaining markers, 5' of the gene, in exon 2, and 3' of the gene were not in LD with any other marker ($0.10 < D' < 0.74$).

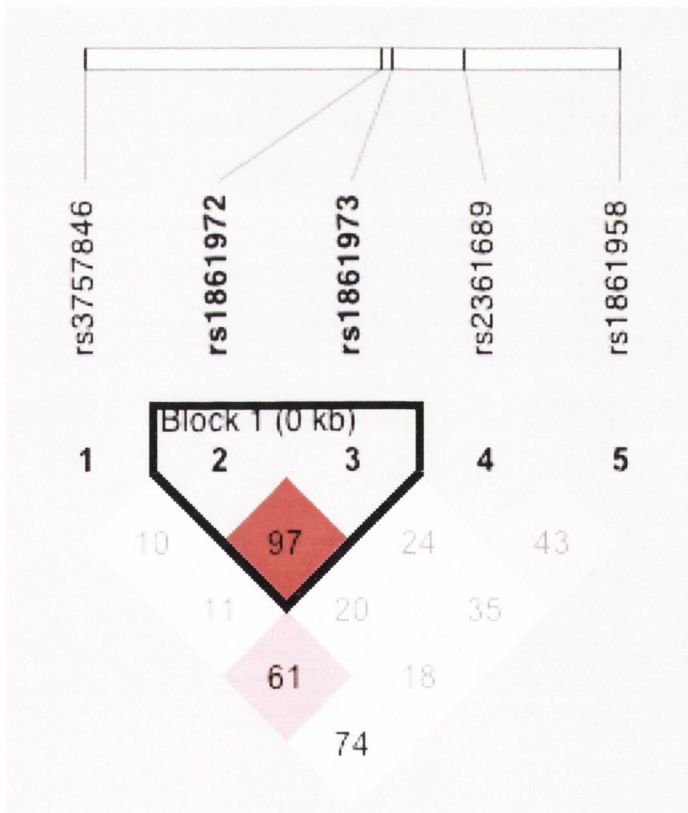


Figure 6.6 LD MEASUREMENTS IN EN2 GENE.

6.3.4.2 TDT ANALYSIS

TDT testing revealed a non-significant trend towards association of the C allele of marker rs3757846, which is located 2.6kb from the first exon of the *EN2* gene. There were 51 transmissions of the C allele versus only 36 non-transmissions from heterozygous parents. This produced an OR of 1.42 and p value of 0.107. There were no other transmission distortions using any of the other markers in the *En2* gene (see Table 6.8).

Marker	Allele	Transmitted	Not Transmitted	OR	χ^2	p value
Rs3757846	C	51	36	1.42	2.599	0.107
Rs1861972	A	55	49	1.22	0.3463	0.556
Rs1861973	T	37	36	1.03	0.014	0.907
Rs2361689	T	66	53	1.25	1.423	0.233
Rs1861958	A	22	14	1.57	1.793	0.1806

Table 6.8 RESULTS OF TDT TESTING IN *EN2* GENE. A non-significant trend towards association was found with marker rs3757846. This however did not reach the threshold of significance (p value < 0.05). It is highlighted in light yellow.

6.3.4.3 HAPLOTYPE TESTING

In a fashion similar to the transmission of single markers, there was only one haplotype that showed a trend towards significance. Once again it involved rs3757846, this time in conjunction with rs1861972, resulting in the over-transmission of the C-A allele (OR = 1.59, p value = 0.102). It did not reach the

threshold of significance at p value < 0.05. No other 2-marker haplotypes showed evidence of association with autism (see Table 6.9).

Markers	Alleles showing excess transmission	Transmitted	Not Transmitted	OR	Global χ^2	Global p-value
rs3757846- rs1861972	C-A	43	27	1.59	6.200	0.102
rs1861972- rs1861973	A-C	22	19	1.05	5.625	0.131
rs1861973- rs2361689	C-C	21	16	1.31	1.915	0.590
rs2361689- rs1861958	T-A	43	27	1.59	4.771	0.189

Table 6.9 TRANSMISSION OF HAPLOTYPES IN *EN2* GENE.

6.4 DISCUSSION

6.4.1 WNT2 GENE

This study presents evidence for a trend towards association with the G allele of rs2024233 in the 3'UTR of the *WNT2* gene with autism (see Section 6.3.1). This result is a replication in part of Wassink's original report, which also found an association with this SNP in a sample of patients with autism²⁶³. However our association was due to the over-transmission of the G allele of this SNP as opposed to the A allele (reported as a T on the forward strand) reported by Wassink et al. Another association study of the *WNT2* gene undertaken by Li et al, found no significant association with either allele of this SNP, but over-transmission of the G allele was also found⁵⁰². In Wassink et al's original study, the association reported was almost exclusively due to a subgroup of individuals with language impairments²⁶³. In our study and the study by Li et al, subgroup analysis was not undertaken. This is due to the difficult nature of classifying individuals based on the qualitative nature of language impairments. Therefore it is possible that we both may have missed true replication of the original report. A study by McCoy et al performed an association test of the two original SNPs (rs39315 and rs2024233) in an overall sample that was subsequently divided into two groups, one presenting with language impairment, and the other presenting with no language impairment⁵⁰³. Neither group, nor the overall sample showed any evidence for association with autism. None of the groups, nor in our sample, reported the occurrence of the two original mutations in exons 1 and 5^{263,502,503}. It is therefore unlikely that these original two mutations make a common contribution to the development of autism. It remains a possibility that the associations in the 3'UTR region in both our study and the study by Wassink et al are chance findings²⁶³. Replication by other groups and subgroup analysis for speech and language impairments may help better explain the role of the *WNT2* gene in autism or highlight the associations mentioned above as chance findings.

6.4.2 FRZB GENE

To date there have been no reports of association studies or otherwise with the *FRZB* gene and autism. This is a little surprising, as it is located on a region of chromosome 2q31-33 that has shown evidence of linkage to autism (see Section 5.1.1). As an antagonist to the Wnt proteins (see Section 6.1.4), and therefore a possible modulator of the Wnt pathway, it presents as a good candidate gene for autism. Within our Irish sample, we found association with a functional mutation in exon 4 (rs288326). Occurrence of the G allele leads to the insertion of an arginine amino acid (hydrophilic with a strong positive charge), whereas occurrence of the A allele leads to insertion of a tryptophan amino acid (larger hydrophobic amino acid with no charge). This non-synonymous transition could lead to potentially deleterious effects in the protein. However in the Irish sample reported here, the more common allele (G) is transmitted to affected probands. Therefore, it is possible that this mutation here is in LD with another mutation elsewhere in the gene, and is merely a “marker” and not of functional significance in this case. A more significant association with autism using another marker, which is more proximal to rs288326, supports this. Another marker in intron 3, rs2242070, showed strong association with autism, as a result of over-transmission of the G allele, resulted in an OR of 1.68 and p value of 0.0073 (see Section 6.3.2.2). Haplotype analysis involving this marker also led to significant associations (see Section 6.3.2.3). The haplotypes associated with autism span exon 3 (rs7722738-rs2242070) and include the non-synonymous transition in exon 4 (rs2242070-rs288326) having ORs of 1.86 (p value = 0.0163) and 2.29 (p value = 0.0019) respectively.

6.4.2 DVL GENE

Expression of the *DVL* gene in the hippocampal regions of the brain⁴⁹⁵, its effect on dendrite branching, and the social and behavioural deficits in knockout mice for the *Dsh* gene makes the *DVLI* gene an interesting candidate gene for autism. There have been no association studies investigating the role of this gene in autism to date.

In the Irish sample described here, there was increased transmission of the G allele of marker rs307354, which is located 6.2kb distal to the *Dvl1* gene. Over-transmission of this allele led to an OR of 1.86 (p value = 0.02), which is just within the boundaries of statistical significance. A haplotype involving rs307359-rs307371 (spanning all exons except exons 1 and 15) showed a significant trend towards association with autism, having a p value of 0.04 (see Section 6.3.3.3) and OR of 1.86. Both of these results indicate that there may be functional variants within this gene predisposing to autism. Unfortunately the low heterozygosity of the markers used (the only markers with heterozygosity information at the time of choosing SNPs), ranging from 0.26 to 0.38, may not be powerful enough to detect association with other unknown functional mutations within the gene. Replication by other groups and also increased mapping of this region may be beneficial in eliminating the chance of a false-positive result, and highlighting regions of the gene harbouring potential functional mutations.

6.4.3 EN2 GENE & AUTISM

The results presented here, do not suggest a conclusive association for *En2* and autism. Replication of the study by Gharani et al, who found an association with autism with two markers in the intron of *EN2* (rs1861972 and rs1861973), did not occur¹⁹¹. In our sample there was no distorted transmission of either of these markers (see Table 6.8). A trend towards association of the C allele of marker rs3757846 did occur, but this did not reach the accepted level of significance (p value < 0.05), see Section 6.3.4.2. Interestingly, this marker is close (~500bp) proximal to a *PvuII* polymorphism with the MP-4 probe originally reported by Petit et al²¹⁸. Due to the low level of LD between markers, also identified by Gharani et al¹⁹¹, it is not possible to exclude the presence of functional mutations within the exons and were captured by other markers in LD with them. A study by Zhong et al failed to find any association using TDT testing and linkage using non-parametric sib-pair analysis with rs3735653 in a sample of 204 AGRE families⁵⁰⁴. It is unclear

why there are discrepancies between these studies. It is possible that the findings of Petit et al and Gharani et al are false positives^{191,218}. It is also possible that any underlying association is resistant to replication in differing populations due to either lower effect size or allelic heterogeneity.

In the four Wnt pathway genes examined here, there were 3 significant associations and one trend towards significance in a total of 17 markers tested. Given a false positive rate of 5%, the chance of all three of these markers being false positive findings is 0.000125. Therefore, while not impossible, it is unlikely that all of these findings are false positives. The evidence supporting the Wnt pathways involvement in the development of autism comes from a wide range of studies, from genetic association studies to mouse knockout experiments (see Section 6.1.2). The analysis presented here also supports the involvement of the some of the genes in the Wnt pathway. Many more genes in this pathway remain to be tested, including the *WNT1*, *WNT3* and *WNT8* genes, all of which play a role in the development of the cerebellum. A collaboration between Dr. M DeRiggi-Ritchie of Vanderbilt University, Tennessee, is currently underway to test if gene-gene interactions between the variants tested above are important in autism. Dr. DeRiggi-Ritchie is adapting the Multifactor-Dimensionality Reduction method (MDR), which is used to identify possible gene-gene interactions in case-control studies, for use with pedigree data. MDR is a non-parametric and genetic model-free alternative to logistic regression for detecting and characterizing nonlinear interactions among genetic and environmental attributes. Preliminary analysis indicates that there may be an interaction rs307371 (*DVL*), rs1861973 (*EN2*) and rs722738 (*FRZB*). However, further testing, analysis and simulations of this method are required to ensure that the new MDR method is reliable to test for such interactions.

CHAPTER 7

GENERAL DISCUSSION

7.1 Aims and Findings

This thesis describes an investigation of the role of genetic factors in the aetiology of autism, a highly heritable, complex neuropsychiatric disorder of childhood. The aim of the work was to:

- 1) To identify a number of potential candidate genes for autism using evidence from *inter alia*, published linkage findings and candidate gene studies (Section 1.3).
- 2) To investigate these candidate genes, using a range of approaches, including genetic association studies, mutation screens, protein activity measurements, collaborations with other groups and gene-gene interaction tests (see Section 1.5).

ITGA4, *DBH*, two genes in the serotonin system (the serotonin transporter and *TPH2*) and 4 genes in the Wnt pathway (*WNT2*, *FRZB*, *DVL1* and *EN2*) were identified as prospective candidate genes for autism and investigate further using the approaches mentioned above.

7.1.1 Findings

The literature has been conflicting regarding the possible role of the serotonin transporter in autism. A number of groups have found genetic association with the long promoter allele, other groups have reported associations with the short promoter allele and some found no association with either allele (see Section 3.1.4 for details). In the Irish sample, association with the short promoter was found, in addition to a number of haplotypes (see Sections 3.3.1.2 and 3.3.1.3). However, it remains possible that this association, and the others mentioned above are merely false positive results. A joint analysis of 11 published reports containing information regarding the transmission of promoter alleles to affected individuals was

undertaken. Overall, there was non-significant increased transmission of the short promoter allele to affected offspring (see Section 3.3.1.5). Thus, in the sample of individuals with autism as a whole, the serotonin transporter does not appear to play a consistent aetiological role. Further work is required to determine if this might be associated with particular clinical characteristics.

Another gene in the serotonin system, tryptophan hydroxylase 2, has also been proposed as a potential candidate gene for autism by Coon et al ⁷. Both the Irish sample and the sample described by Coon et al showed association with a marker located in intron X, rs11179000, although associations were with different alleles. The difference in design of both studies (case-control versus family design based study) means that the results are not directly comparable (see Section 3.4.1.2). The validity of these results remains to be determined.

Dopamine is a ubiquitous neurotransmitter in the CNS, and abnormalities in its concentration, and that of related enzymes and its metabolites has been reported in autism, including dopamine beta hydroxylase (DBH) (see Section 4.1.1). A striking and highly significant reduction in DBH activity in plasma from probands compared to their parents or to population controls (see Sections 4.3.3 and 4.3.4). These decreases were not due to the over-transmission of “low activity” alleles and/or haplotypes. Regression analysis indicated that the variants studied played a significantly smaller role in determining DBH activity in probands compared to their parents (see Section 4.4.5). Possible explanations include mutations/variations in other genes, leading to the downstream effect of decreasing DBH activity, or environmental or developmental switches (see Section 4.4.6).

A number of reasons (reword) led to the investigation of the Integrin alpha 4 gene (*ITGA4*) as a candidate gene for autism, including linkage studies, fine-mapping of the surround region of chromosome 2q in a patient with a translocation in this region and a literature search (see Section 5.1). Dense SNP mapping of the gene was undertaken, and statistical testing revealed significant over-transmission of markers

(rs3770112) and haplotypes close to exons 16 and 17 (see Sections 5.3.3 and 5.3.4). Mutation screening of exon 16 identified a possible splice site variant, rs12690517, which was also significantly over-transmitted to probands with autism (see Section 5.3.5). An attempted replication study in two other samples, AGRE and Vanderbilt samples, was also carried out. The AGRE sample provided support for the original findings in the Irish sample (see Sections 5.3.6.2 and 5.3.6.3). The Vanderbilt sample also showed significant over transmission of variants and haplotypes (see Sections 5.3.6.2 and 5.3.6.3). However these were different to those in the Irish sample. A joint analysis of the three samples combined indicated significant over-transmission of the C allele of rs3770112, and the haplotypes spanning exons 16 and 17 (see Section 5.3.6.5). A reason for the differences in haplotypes associated with autism may be that there are multiple risk haplotypes (see Section 5.4.4). Further replication of the work presented here is essential.

Finally, four genes in the Wnt pathway, *WNT2*, *FRAB*, *DVL* and *EN2*, were tested for association with autism based on their roles in brain development, in mouse models exhibiting social interaction deficits, in cytogenetic locations and in previous reports of associations in samples diagnosed with autism (see Section 6.1). Statistical associations (at the $p < 0.05$ level) were found with variants in the *FRZB* and *DVL* genes (see Section 6.3). These associations, in addition to evidence from other sources (see Section 6.1.2) highlight the potential importance of these genes, and the possible role of the Wnt pathway in the development of autism.

7.2 Strengths and Limitations

7.2.1 Strengths of this Study

7.2.1.1 Irish Sample

Ascertainment of an appropriate sample is the most crucial step in a genetic association study. The ideal sample for a genetic association study is one that is clinically and genetically homogenous, definitively diagnosed for the phenotype being examined, large enough in size to detect associations with low frequency variants and variants that result in moderate ORs, and finally be drawn at random from the total population of individuals with the phenotype in question (excludes bias). In reality the collection of the ideal sample is unachievable. However, even modest deviations from the above desired characteristics could be accommodated in most studies.

The location of Ireland, an island on the periphery of Europe, and its social history of limited migration into the country, provides a population that is considered genetically quite homogenous. A number of studies have reported clines in genetic diversity that peak in Western Europe, including the “hg1” Y chromosome haplotype⁵¹², the Q188R mutant allele in transferase-deficient galactosaemia⁵²¹ and blood group O⁵²². Scandanavian mutations causing hyperphenylalaninaemia, mediated by Viking incursions from North Western Norway, account for 6.1% of mutant alleles detected in an Irish sample diagnosed with hyperphenylalaninaemia⁵²³. The Ulster region of Ireland, unlike the rest of Ireland, shows a high genetic similarity to Scotland. Scottish mutant hyperphenylalanaemia mutations accounted for 46% of all mutations detected in Ulster. This is as a result of the Ulster-Irish settelements (3rd to

5th century AD), Viking settlement and exchange between Western Scotland and Ireland (8th to 12th century) and immigration from Scotland in the Plantation of Ulster (17th century)⁵²³. Since the potato famine in Ireland (1845-1849) there has been a constant migration out of the country, resulting in genetic drift and gene flow out of the country. Until the last 15 years, there has been little to no migration into the country. As a result, the Irish population may possess less genetic variation than other more multi-ethnic countries (e.g. the USA and UK). A homogenous population may have fewer susceptibility loci, with potentially higher ORs, than more genetically heterogeneous populations⁵¹³. This may aid in the search for autism susceptibility genes.

In any complex disorder, there are difficulties in sample collection, definition of phenotype and measurement of the traits involved. As autism is a relatively rare disorder, there is already a limitation on the collection of the sample. The ascertainment and phenotypic measurement of samples for analysis in neuropsychiatric disorders is costly, time consuming and requires individuals who are specifically trained to evaluate the phenotype being investigated. There are also a number of necessary ethical restraints involved in dealing with a neuropsychiatric childhood disorder. As stated in the introduction, it is likely that there are a number of functional variants in genes that have a relatively small effect in their own right. It is the accumulation of and/or interaction between these variants that leads to the development of autism. Large samples are required to detect these variants. The Irish sample used in this study is a relatively large and homogenous and thus is a valuable resource for the purpose of genetic association investigations. Despite this, definitive results are difficult to obtain. This is due to the complex nature of the disease, and the likely small effect size of susceptibility genes. As stated in the introduction (Section 1.2.3), autism is very likely a multi-gene disorder, with up to 15 or more genes involved in the development of the autistic aetiology. Therefore, it is unsurprising that definitive results require studies with very large numbers of affected individuals. The collaboration study undertaken at the *ITGA4* gene is an example of a positive finding in a large sample, which even with 1231 affected

individuals and their families, studies are subject to statistical chance (see Section 5.3.6.5).

7.2.1.2 Design and Analysis

As mentioned in the introduction, the use of family based association testing is one method used to reduce the possibility of a false positive result due to population stratification (see Section 1.4.1.2). A disadvantage of this method is the reduced power to detect associations with low frequency variants. However, a similar problem also occurs in case-control studies. Association studies remain more powerful than linkage studies in detecting genes of modest effect that contribute to complex disorders¹⁷⁸. For variants with genotype relative risks (GRRs) greater than 4, the number of affected sib-pairs required to detect linkage is between 185 and 297 for variants with an allele frequency between 5% and 75%. In such cases, linkage studies are preferable. However, in the case of variants with GRRs less than 2 (as is likely in autism), then the number of individuals required to detect linkage is greater than 2000, making sample collection extremely problematic. For family based association studies however, even GRRs less than 1.5 can be detected in samples less than 1000⁵¹⁴.

Collaborations between groups, as in the *ITGA4* association study (see Chapter 5), are a useful and relatively easy way of increasing sample size. Autism is not a common disorder, affecting only 10 per 10,000¹⁶. This in addition to the high number of genes involved in the development of the aetiology makes the detection of many genes of small effect difficult. In the case of the initial association between the *ITGA4* gene and autism, collaboration resulted in associations in a joint analysis in a sample of 1231 affected individuals in 524 pedigrees. The OR of the rs3770112 was 1.73 in the original Irish sample (p value = 0.0051), but in the joint analysis the OR decreased to 1.19 (p value = 0.029). The haplotype containing markers rs1349197 and rs3770112 produced an OR of 1.75 in the Irish population (p value = 0.02987),

and this was also reduced in the joint sample to 1.22 (p value = 0.001815). These reductions in ORs may represent two situations. Firstly it is possible that the original result in the Irish population is a false positive result, and the reduction of OR in the joint analysis merely reflects the tendency of larger samples to give a more accurate reflection of the true effect of the gene in the population. However it may also reflect that the *ITGA4* is a gene of small effect, of which there are hypothesised to be many (see Section 1.2.3), a larger OR in the Irish sample is therefore “required” to be detected at all. The increased OR in the Irish population may have arisen by chance (in any large multi-sample study, some samples will have higher ORs than others) or that the Irish population OR for *ITGA4* is higher, perhaps due to genetic homogeneity. Given that statistical modelling has suggested that there are between 3⁶⁵ and greater than 15 genes⁶⁶, the average OR ranges from 1.33 to less than 1.067. An OR of 1.22 lies within this range, and therefore it is not surprising that a significant finding, such as the rs1349197-rs3770112 haplotype (p value = 0.001815) would produce only a modest OR. If genes for autism are all genes of small effect, then very large samples are required to detect such associations.

7.2.1.3 Gene Coverage

To ensure complete coverage of a gene when undertaking association studies, one must genotype all variants in a gene, including those in promoter and long-range enhancer regions. Given the difficulties in identifying promoter and long-range enhancer regions, very few genes have been covered completely in any association study. Methods of selecting and prioritising variants are thus important for association studies. Within this thesis, preference was given to genotyping putative functional polymorphisms and ensuring LD coverage. In the *DBH* gene (see Section 4.2.2), the putative promoter variant, C-1021T, was genotyped, in addition to 4 exonic SNPs and another 2 promoter SNPs were chosen for genotyping. In the case of the *ITGA4* gene, 18 SNPs were chosen to cover the gene, from 1.9kb 5' of the gene to 2.4kb 3' of the gene. Within this region, markers were selected based upon

their heterozygosity (those with the highest heterozygosity being the most informative), and the inter-marker distance. High LD between markers, with the exception of rs1449260 and rs3770136 (due to a lack of suitable markers between these two), ensured maximum coverage using the selection criteria. A similar strategy was used in the selection of markers for the *FRZB* and *DVL1* genes. The testing of haplotype tagging SNPs (htSNPs) in the *ITGA4* gene also proved to be an effective method to minimise genotyping, while maximising transmission information for an initial screening process. Follow-up analysis would be undertaken around any marker (or haplotype) showing trends towards positive association. In the case of replication studies, i.e. the serotonin transporter, *TPH2*, *WNT2* and *EN2*, the aim of these studies was the replication of initial results, and therefore complete coverage of the gene was not being attempted. It is therefore important to state that as no gene was completely covered, negative findings do not eliminate the possibility that the genes tested contain other variants associated with autism. Instead, it merely indicates that the tested variants are not associated with autism in the Irish population.

It is important also to mention the possible effect of long-range control of gene expression that may play a role in leading to different markers/haplotypes being associated with autism, e.g. the serotonin transporter, *DBH* or *ITGA4* genes. Many disease-causing variants are located within the coding region of genes, splice site or in the promoter region of the genes. However there are a small number of diseases/syndromes, e.g. Deafness type 3 (the most common form of X-linked deafness), that can also occur due to the deletion of regulatory regions many kilobases from the gene involved in the pathogenesis of the disease/syndrome (see Kleinjan and van Heyningen, for a review on this topic)⁵¹⁵. In the case of deafness type 3, the usual spectrum of mutation types within the gene (*POU3F4*) lead to the observed loss of hearing. However, there are also a number of cases where this form of deafness occurs do to chromosomal deletions, inversions and translocations and other mutations in a region approximately 900kb upstream of the *POU3F4* gene⁵¹⁶. At this location there is a 2kb sequence that shows 80% conservation between human

and mouse, and therefore it is assumed to be of functional significance. It is possible that there is a regulatory sequence upstream of the serotonin transporter gene that is associated with autism. Different mutations within this sequence could theoretically be in LD with different alleles or haplotypes showing association. It is not yet known how far upstream of a gene, markers should be placed to capture association by such enhancer sequences. It is also assumed that the majority of variants associated with complex disorders such as autism, are located close to or within the genes themselves.

7.2.2 Limitations of Study

7.2.2.1 Multiple Testing

The problem of multiple testing remains one of the major concerns in the field of gene mapping in genetically complex disorders. According to formal statistical theory, the p value of any positive result should be corrected for the number of tests performed on the data, and arguably, for the number of tests that might be performed. The simplest and most conservative method for correcting for multiple tests is the Bonferroni method. This states that if the significance level for the entire set of n independent comparisons is α , then the significance level for each comparison should be α/n . This is because with each additional test, there is an increased likelihood of finding a false positive result. When correcting for multiple tests it is necessary to determine the number of independent tests that have to be corrected for. Cases of sub-group analysis and multiple association tests do require correction, but the situation becomes more complicated when tests are not independent. Examples of non-independent tests include markers in LD with each other and haplotype tests. Each test is not providing a completely independent chance of false positive error. Bonferroni corrections in this case are too conservative⁵¹⁷. In these cases Nyholt suggests that “it may be best simply to report the disease-marker association results without correction, stating the presence of marker-marker association”⁵¹⁷. This is the strategy followed throughout this thesis and stated clearly in advance. Taking a conservative approach, Risch and Merikangas suggest that this value be changed to 5×10^{-8} ⁵¹⁴. This value was based upon 100,000 genes being tested, each with 5 diallelic markers, leading to 10×10^5 tests, i.e. 10^6 tests and the equivalent false positive rate of 0.05 for a million association tests can be obtained with a significance level of 5×10^{-8} .

In this thesis, all results are presented in the form of raw p-values that are uncorrected for multiple testing. Replication is one of the best and most common

methods used to address the problem of multiple testing in the field of neuropsychiatric genetics. Replication of a positive finding increases the likelihood of it being a true finding. Also, a replication study consists of an *a priori* hypothesis, and is therefore only one test. This overcomes the need for corrections. Possibly more serious, and not often discussed, is the possibility of false negative results, particularly with small samples. Aggregating studies in a systematic and unbiased way may allow a clearer picture to emerge.

Given the high number of associations presented in this thesis, it is likely that some of them at least are “true positives”. Fifty-four variants were examined in total. Given a false positive rate of 5%, one would expect 2.7 false positive results. A basic chi square test, based on an observed finding of 8 and an expected finding of 2.7, the number of positive findings was significant ($\chi^2 = 10.40$, p value = 0.0013). Thus, one would expect that at least some of the genes reported here, do play a role in the development of autism.

7.3 General Conclusions

The results of this thesis present evidence for the association of a number of gene variants with autism. The results of the *ITGA4* fine-mapping experiment and *DBH* analysis look particularly interesting. All of the studies utilised association tests as a principal component of investigation. Association tests remain the most common analysis in the search for genes involved in complex disorders. With the decreasing cost of genotyping variants, and the increasing speed of genotyping, the number of genes that can be analysed for any one disorder is rapidly increasing. Therefore the number of false-positive results will also increase. False positive results are a “necessary evil” in the field of complex genetics. Replications and meta-analysis are methods used to confirm initial associations or reveal false positives.

Collaboration between groups is also of extreme importance, as is the multi-faceted approach to testing candidate genes. A number of collaborations have been undertaken with the work presented here. A purely genetic association study of *DBH* variants would not have highlighted the decreased DBH activity in probands with autism. Without the DBH activity measurements, carried out in collaboration with Dr. George Anderson, one would have concluded “no association” between the *DBH* gene and autism. While this remains the case in terms of genetic association, it is obvious that the decrease in DBH activity in individuals with autism may play a very important role in the development of autism. This vindicates the essential role of approaching a candidate gene test, e.g. undertaking biochemical tests, from multiple angles. The reason for this decreased DBH activity remains unclear, but for the collaboration, further investigation of this system would not be undertaken, and an important biochemical finding would not have been investigated.

Similarly, cytogenetics undertaken in the National Centre for Medical Genetics, uncovered a chromosome 2q;9q translocation, which in turn led to the investigation of the *ITGA4* gene as a candidate gene for autism. Positive associations in the Irish sample, were followed up with replications in two other samples, kindly provided by colleagues at Vanderbilt University, Tennessee. Association of *ITGA4* variants was thus confirmed in two other samples. Association in a large joint sample was also confirmed. This large sample was comprised of 1231 affected individuals in 524 families, a size that far exceeded the initial Irish sample size of 184 families. Large sample sizes are required to find associations with genes of small effect. In the case of autism, there are presumed to be many genes, each of moderate to small effect, and hence large sample sizes are crucial for genetic association tests.

Gene-gene interaction studies may also be important in untangling and helping resolve the complex nature of autism. Autism remains one of the least understood neuropsychiatric disorders of childhood. Unlike ADHD, which has shown positive association with a number of genes, conclusive associations remain elusive in

autism. It is this complexity that makes autism an interesting disorder for candidate gene searches.

7.4 Future Directions

The work undertaken in this thesis presented evidence for the association of a number of genes with autism, very significantly more than might have been expected by chance. However, it should be cautioned that in all cases, replication by independent groups is required to test the association of these genes in other samples. These replications would help strengthen the statistical support for these genes, or produce evidence that the initial findings in the Irish sample may have been false positive results.

Testing of other genes that may influence *DBH* expression will continue. Association studies on the *ARIX* and *PHOX2A* genes are top priority. If time and money were unlimited, a full genome scan could be undertaken, in order to find additional loci that may influence *DBH* expression. This could identify the genes hypothesised to trigger an increase in *DBH* activity between 2-3 years of age, and to be defective in children with autism. Efforts to identify environmental triggers or developmental switches having an effect on *DBH* expression may also shed light on the reasons for decreased *DBH* activity. An increased sample size would also allow testing the hypothesis that *DBH* activity may influence susceptibility to epilepsy.

With respect to *ITGA4*, another group in Portugal, is currently involved in an attempted replication of the results presented here in a similarly genetically homogeneous Portuguese sample. An investigation of the role of the gene in neurodevelopment is planned (with additional funding) in an animal model using a tissue specific (CNS) mouse null mutant. This might be complemented in the future by behavioural studies of these null mutants. Expression studies in brain tissue that

investigated the potential role of this (and other) variant detected in this work would also be of merit.

A form of gene-gene interaction analysis undertaken using multiple genes in the Wnt pathway may be useful in determining if multiple variants increase the risk for developing autism. Autism is a complex disease, which doesn't follow the simple Mendelian patterns of inheritance. As stated in the introduction (Section 1.2.3), it is most likely that a number of genes may play a role in the autistic aetiology. Each of these genes may have a small, possibly synergistic effect on phenotype. Therefore it is necessary to consider the simultaneous effect of several SNP genotypes at different loci on phenotype. Gene gene interaction analysis has two aims. Firstly, to identify the SNPs that are associated (singularly or with other SNPs) with the phenotype, and secondly to classify patients based on genotype in a manner that captures information about disease risk. Multifactor dimensional reduction (MDR) is a comparison method, which is designed to uncover complex relationships without relying on a specific model for interactions⁵¹⁸. MDR is a special case of recursive partitioning in which patterns are used as predictors, tree growth is restricted to a single split and misclassification error is used as a measure of impurity. The advantage of MDR analysis is that it has the ability to identify high order interactions among SNPs, and that it does not require a specific interaction model. This method has been used to find possible gene-gene interactions in a case-control sample of individuals with breast cancer⁵¹⁹. Modifications of this technique are currently being developed to allow the same analysis to be used in trio-based samples (Dr. M DeRiggi Ritchie, Vanderbilt University, Tennessee, personal communication). Development of this statistical technique will be useful in determining the role of the Wnt pathway in the development of autism. A number of other Wnt genes remain to be tested for association with autism, e.g. *WNT1*, *WNT3* and *WNT8*. These genes are good potential candidates for the development of the cerebellum. Mutation screening of exons 3 and 4 in the *FRZB* is also important, as this may uncover functional mutations which are in LD with the two markers showing association in the Irish sample (see Section 6.3.3).

Finally, expansion of the clinical data on the sample recruited, including further phenotypic measures, neuropsychological assessments and functional imaging would allow sub-phenotyping to be undertaken. It is important to note that within any sample diagnosed with autism there still remains a spectrum of affectedness for certain traits. Levels of IQ and obsessiveness can vary from individual to individual. The use of sub-phenotyping may help detect associations that otherwise may have remained hidden. 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 analysis¹⁷⁶. 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. Currently there has been an application made for funding of a multi-centre collaborative study (involving this sample and others based at twelve other international sites) that would involve pooling of samples to give greater power and allow these types of investigations to occur. The plan for this investigation is to use the largest ever autism sample (1200 multiplex families and 2000 trios) to pursue genetic investigations in autism. This will allow for multiple approaches to be utilised, and it is envisioned that this continue the process of teasing out the complex interactions between genotype and phenotype for a particular sub-phenotype, e.g. hyperactivity features or obsessiveness.

The work presented would not be possible without the unwavering support and participation of the families, who have donated their time and DNA to facilitate these studies. The way forward for investigations into the aetiology of autism will, no doubt be difficult, but should, gradually, and with intelligent experimental design and some measure of insight, bear fruit. The hope and expectation, of researchers in the field, is that the achievement of advances can be of direct benefit to the patients and their families.

REFERENCES:

1. Betancur, C. et al. Serotonin transporter gene polymorphisms and hyperserotonemia in autistic disorder. *Mol Psychiatry* **7**, 67-71 (2002).
2. Coutinho, A.M. et al. Variants of the serotonin transporter gene (SLC6A4) significantly contribute to hyperserotonemia in autism. *Mol Psychiatry* **9**, 264-71 (2004).
3. Kim SJ et al. Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder. *Mol Psychiatry* **7**, 278-88 (2002).
4. Klauck SM, Poustka F, Benner A, Lesch KP & A., P. Serotonin transporter (5-HTT) gene variants associated with autism? *Hum Mol Genet* **6**, 2233-8 (1997).
5. Maestrini E et al. Serotonin transporter (5-HTT) and gamma-aminobutyric acid receptor subunit beta3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. *Am J Med Genet* **88**, 492-6 (1999).
6. McCauley, J.L. et al. Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. *Am J Med Genet B Neuropsychiatr Genet* **127**, 104-12 (2004).
7. Coon, H. et al. Possible association between autism and variants in the brain-expressed tryptophan hydroxylase gene (TPH2). *Am J Med Genet B Neuropsychiatr Genet* (2005).
8. Oxenstierna, G. et al. Concentrations of monoamine metabolites in the cerebrospinal fluid of twins and unrelated individuals--a genetic study. *J Psychiatr Res* **20**, 19-29 (1986).
9. Levitt, M. & Mendlewicz, J. A genetic study of plasma dopamine beta-hydroxylase in affective disorder. *Mod Probl Pharmacopsychiatry* **10**, 89-98 (1975).
10. Kanner, L. Autistic disturbances of affective contact. *Nervous Child* **2**, 217-250 (1943).
11. Asperger, H. Die autistischen Psychopathen im Kindersalter. *Archiv fuer Psychiatrie und Nervenkrankheiten*. **177**, 76 (1944).
12. World Health Organization, W. *International Statistical Classification of Diseases and Related Health Problems*, (Geneva, 1992).
13. American, P.A. *Diagnostic and Statistical Manual of Mental Disorders*, (American Psychiatric Association, 2000).
14. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659-85 (1994).
15. Lord, C. et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* **19**, 185-212 (1989).

16. Fombonne, E. The epidemiology of autism: a review. *Psychol Med* **29**, 769-86 (1999).
17. Lotter, V. Epidemiology of autistic conditions in young children: I. Prevalence. *Social Psychiatry* **1**, 124-137 (1966).
18. Newschaffer, C.J., Falb, M.D. & Gurney, J.G. National autism prevalence trends from United States special education data. *Pediatrics* **115**, e277-82 (2005).
19. Treffert, D.A. Epidemiology of infantile autism. *Arch Gen Psychiatry* **22**, 431-8 (1970).
20. Kadesjo, B., Gillberg, C. & Hagberg, B. Brief report: autism and Asperger syndrome in seven-year-old children: a total population study. *J Autism Dev Disord* **29**, 327-31 (1999).
21. Association, A.P. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. (American Psychiatric Association, Washington, DC, 1998).
22. Croen, L.A., Grether, J.K. & Selvin, S. Descriptive epidemiology of autism in a California population: who is at risk? *J Autism Dev Disord* **32**, 217-24 (2002).
23. McCarthy, P., Fitzgerald, M. & Smith, M.A. Prevalence of childhood autism in Ireland. *Ir Med J* **77**, 129-30 (1984).
24. Wing, J.K. Epidemiology of Early Childhood Autism. *Dev Med Child Neurol* **27**, 646-7 (1963).
25. Foistein SE & B, R.-S. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet.* **2**, 943-55 (2001).
26. Fryer, A.E. et al. Evidence that the gene for tuberous sclerosis is on chromosome 9. *Lancet* **1**, 659-61 (1987).
27. Haines, J.L. et al. Genetic heterogeneity in tuberous sclerosis. Study of a large collaborative dataset. *Ann N Y Acad Sci* **615**, 256-64 (1991).
28. Povey, S. et al. Two loci for tuberous sclerosis: one on 9q34 and one on 16p13. *Ann Hum Genet* **58 (Pt 2)**, 107-27 (1994).
29. Serajee, F.J., Nabi, R., Zhong, H. & Mahbulul Huq, A.H. Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism. *J Med Genet* **40**, e119 (2003).
30. Bowden, J.A. & McArthur, C.L., 3rd. Possible biochemical model for phenylketonuria. *Nature* **235**, 230 (1972).
31. Mulley, J.C. et al. FRAXE and mental retardation. *J Med Genet* **32**, 162-9 (1995).
32. Cohen, I.L. et al. Why are autism and the fragile-X syndrome associated? Conceptual and methodological issues. *Am J Hum Genet* **48**, 195-202 (1991).
33. Hjalgrim, H., Gronskov, K. & Brondum-Nielsen, K. [Fragile X syndrome. Diagnosis, genetics and clinical findings]. *Ugeskr Laeger* **160**, 5330-4 (1998).
34. Klauck, S.M. et al. Molecular genetic analysis of the FMR-1 gene in a large collection of autistic patients. *Hum Genet* **100**, 224-9 (1997).
35. Meyer, G.A., Blum, N.J., Hitchcock, W. & Fortina, P. Absence of the fragile X CGG trinucleotide repeat expansion in girls diagnosed with a pervasive developmental disorder. *J Pediatr* **133**, 363-5 (1998).

36. Parazzini, C. et al. Spontaneous involution of optic pathway lesions in neurofibromatosis type 1: serial contrast MR evaluation. *AJNR Am J Neuroradiol* **16**, 1711-8 (1995).
37. Eliason, M. Neuropsychological patterns: neurofibromatosis compared to developmental learning disorders. *Neurofibromatosis* **1**, 17-25 (1988).
38. Gillberg, C. & Forsell, C. Childhood psychosis and neurofibromatosis--more than a coincidence? *J Autism Dev Disord* **14**, 1-8 (1984).
39. Steffenburg, S. Neuropsychiatric assessment of children with autism: a population-based study. *Dev Med Child Neurol* **33**, 495-511 (1991).
40. Nicolson, R. & Szatmari, P. Genetic and neurodevelopmental influences in autistic disorder. *Can J Psychiatry* **48**, 526-37 (2003).
41. Bailey, A. et al. A clinicopathological study of autism. *Brain* **121**, 889-905 (1998).
42. Courchesne, E. Brainstem, cerebellar and limbic neuroanatomical abnormalities in autism. *Curr Opin Neurobiol* **7**, 269-78 (1997).
43. Abell, F. et al. The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. *Neuroreport* **10**, 1647-51 (1999).
44. Sparks, B.F. et al. Brain structural abnormalities in young children with autism spectrum disorder. *Neurology* **59**, 184-92 (2002).
45. Giedd, J.N., Castellanos, F.X., Rajapakse, J.C., Vaituzis, A.C. & Rapoport, J.L. Sexual dimorphism of the developing human brain. *Prog Neuropsychopharmacol Biol Psychiatry* **21**, 1185-201 (1997).
46. Murphy, D.G. et al. Asperger syndrome: a proton magnetic resonance spectroscopy study of brain. *Arch Gen Psychiatry* **59**, 885-91 (2002).
47. Carper, R.A. & Courchesne, E. Inverse correlation between frontal lobe and cerebellum sizes in children with autism. *Brain* **123** (Pt 4), 836-44 (2000).
48. Courchesne, E. et al. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. *Neurology* **24;57**, 245-54 (2001).
49. Ritvo, E.R. et al. The UCLA-University of Utah epidemiologic survey of autism: the etiologic role of rare diseases. *Am J Psychiatry* **147**, 1614-21 (1990).
50. Chakrabarti, S. & E., F. Pervasive developmental disorders in preschool children. *JAMA* **285**, 3093-9 (2001).
51. Chudley, A.E., Gutierrez, E., Jocelyn, L.J. & Chodirker, B.N. Outcomes of genetic evaluation in children with pervasive developmental disorder. *J Dev Behav Pediatr* **19**, 321-5 (1998).
52. Gillberg, C. & M, C. *The Biology fo the Autistic Syndromes*, (Mac Keith Press, Distributed by Cambridge University Press, London, Uk, 2000).
53. Smalley, S.L., Asarnow, R.F. & Spence, M.A. Autism and genetics. A decade of research. *Arch Gen Psychiatry* **45**, 953-61 (1988).
54. Bolton, P. et al. A case-control family history study of autism. *J Child Psychol Psychiatry* **35**, 877-900 (1994).
55. Pickles, A. et al. Variable expression of the autism broader phenotype: findings from extended pedigrees. *J Child Psychol Psychiatry* **41**, 491-502 (2000).

56. Faraone, S.V. et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* **57**, 1313-23 (2005).
57. Sullivan, P.F., Kendler, K.S. & Neale, M.C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* **60**, 1187-92 (2003).
58. McGuffin, P. et al. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry* **60**, 497-502 (2003).
59. Folstein, S. & Rutter, M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* **18**, 297-321 (1977).
60. Ritvo, E.R., Freeman, B.J., Mason-Brothers, A., Mo, A. & Ritvo, A.M. Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* **142**, 74-7 (1985).
61. Steffenburg S et al. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry* **30**, 405-16 (1989).
62. Bailey A et al. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**, 63-77 (1995).
63. Greenberg, D.A., Hodge, S.E., Sowinski, J. & Nicoll, D. Excess of twins among affected sibling pairs with autism: implications for the etiology of autism. *Am J Hum Genet* **69**, 1062-7 (2001).
64. Jorde LB et al. Complex segregation analysis of autism. *Am J Hum Genet* **49**, 932-8 (1991).
65. Pickles A, B.P., Macdonald H, Bailey A, Le Couteur A, Sim CH, Rutter M. 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 (1995).
66. Risch N, S.D., Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, McCague P, Dimiceli S, Pitts T, Nguyen L, Yang J, Harper C, Thorpe D, Vermeer S, Young H, Hebert J, Lin A, Ferguson J, Chiotti C, Wiese-Slater S, Rogers T, Salmon B, Nicholas P, Myers RM, et al. A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* **65**, 493-507 (1999).
67. Wakefield, A.J. et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* **351**, 637-41 (1998).
68. Madsen, K.M. et al. A population-based study of measles, mumps, and rubella vaccination and autism. *N Engl J Med* **347**, 1477-82 (2002).
69. Taylor, B. et al. Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association. *Lancet* **353**, 2026-9 (1999).
70. Betancur, C., Leboyer, M. & Gillberg, C. Increased rate of twins among affected sibling pairs with autism. *Am J Hum Genet* **70**, 1381-3 (2002).
71. Hallmayer, J. et al. On the twin risk in autism. *Am J Hum Genet* **71**, 941-6 (2002).
72. Phillips, D.I. Twin studies in medical research: can they tell us whether diseases are genetically determined? *Lancet* **341**, 1008-9 (1993).

73. Stromland, K., Nordin, V., Miller, M., Akerstrom, B. & Gillberg, C. Autism in thalidomide embryopathy: a population study. *Dev Med Child Neurol* **36**, 351-6 (1994).
74. Christianson, A.L., Chesler, N. & Kromberg, J.G. Fetal valproate syndrome: clinical and neuro-developmental features in two sibling pairs. *Dev Med Child Neurol* **36**, 361-9 (1994).
75. Moore, M.L., Eichner, S.F. & Jones, J.R. Treating functional impairment of autism with selective serotonin-reuptake inhibitors. *Ann Pharmacother* **38**, 1515-9 (2004).
76. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
77. IMGSAC. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. *Hum Mol Genet* **7**, 571-8 (1998).
78. Philippe A, M.M., Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M, van Malldergerme L. Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. *Hum Mol Genet* **8**, 805-12 (1999).
79. Liu, J. et al. A genomewide screen for autism susceptibility loci. *Am J Hum Genet* **69**, 327-40 (2001).
80. Buxbaum JD et al. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* **68**, 1514-20 (2001).
81. IMGSAC, I.M.G.S.o.A.C. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* **69**, 570-81 (2001).
82. Alarcon, M., Cantor, R.M., Liu, J., Gilliam, T.C. & Geschwind, D.H. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* **70**, 60-71 (2002).
83. Auranen, M. et al. Analysis of autism susceptibility gene loci on chromosomes 1p, 4p, 6q, 7q, 13q, 15q, 16p, 17q, 19q and 22q in Finnish multiplex families. *Mol Psychiatry* **5**, 320-2 (2000).
84. Auranen, M. et al. A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25-27. *Am J Hum Genet* **71**, 777-90 (2002).
85. Yonan, A. et al. A genomewide screen of 345 families for autism-susceptibility Loci. *Am J Hum Genet.* 2003 Oct;73(4):886-97. **73**, 886-9 (2003).
86. Shao Y et al. Genomic screen and follow-up analysis for autistic disorder. *Am J Med Genet* **114**, 99-105 (2002).
87. McCauley, J.L. et al. Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. *BMC Med Genet* **6**, 1 (2005).

88. Lamb, J.A. et al. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet* **42**, 132-7 (2005).
89. Barrett, S. et al. An autosomal genomic screen for autism. Collaborative linkage study of autism. *Am J Med Genet* **88**, 609-15 (1999).
90. (IMGSAC), I.M.G.S.o.A.C. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* **69**, 570-81 (2001).
91. Beyer, K.S., Klauck, S.M., Wiemann, S. & Poustka, A. Construction of a physical map of an autism susceptibility region in 7q32.3-q33. *Gene* **272**, 85-91 (2001).
92. Shao, Y. et al. Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am J Hum Genet* **70**, 1058-61 (2002).
93. Sutcliffe, J.S. et al. Allelic Heterogeneity at the Serotonin Transporter Locus (SLC6A4) Confers Susceptibility to Autism and Rigid-Compulsive Behaviors. *Am J Hum Genet* **77**, 265-79 (2005).
94. Stone, J.L. et al. Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet* **75**, 1117-23 (2004).
95. Cantor, R.M. et al. Replication of Autism Linkage: Fine-Mapping Peak at 17q21. *Am J Hum Genet* **76**(2005).
96. Gillberg, C. et al. Autism associated with marker chromosome. *J Am Acad Child Adolesc Psychiatry* **30**, 489-94 (1991).
97. Robinson, W.P. et al. Clinical and molecular analysis of five inv dup(15) patients. *Eur J Hum Genet* **1**, 37-50 (1993).
98. Baker, P., Piven, J., Schwartz, S. & Patil, S. Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. *J Autism Dev Disord* **24**, 529-35 (1994).
99. Bunday, S., Hardy, C., Vickers, S., Kilpatrick, M.W. & Corbett, J.A. Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. *Dev Med Child Neurol* **36**, 736-42 (1994).
100. Leana-Cox, J. et al. Molecular cytogenetic analysis of inv dup(15) chromosomes, using probes specific for the Prader-Willi/Angelman syndrome region: clinical implications. *Am J Hum Genet* **54**, 748-56 (1994).
101. Schinzel, A.A. et al. Intrachromosomal triplication of 15q11-q13. *J Med Genet* **31**, 798-803 (1994).
102. Crolla, J.A., Harvey, J.F., Sitch, F.L. & Dennis, N.R. Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. *Hum Genet* **95**, 161-70 (1995).
103. Hotopf, M. & Bolton, P. A case of autism associated with partial tetrasomy 15. *J Autism Dev Disord* **25**, 41-9 (1995).
104. Flejter, W.L. et al. Cytogenetic and molecular analysis of inv dup(15) chromosomes observed in two patients with autistic disorder and mental retardation. *Am J Med Genet* **61**, 182-7 (1996).
105. Cook, E. et al. Autism or atypical autism in maternally but not paternally derived proximal 15q duplications. *Am J Hum Genet* **60**, 928-934 (1997).

106. Vu, T.H. & Hoffman, A.R. Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. *Nat Genet* **17**, 12-3 (1997).
107. Buxbaum JD et al. Association between a GABRB3 polymorphism and autism. *Mol Psychiatry* **7**, 311-6 (2002).
108. Menold, M.M. et al. Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder. *J Neurogenet* **15**, 245-59 (2001).
109. Martin, E.R. et al. Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. *Am J Med Genet* **96**, 43-8 (2000).
110. McCauley, J.L. et al. A linkage disequilibrium map of the 1-Mb 15q12 GABA(A) receptor subunit cluster and association to autism. *Am J Med Genet B Neuropsychiatr Genet* **131**, 51-9 (2004).
111. Nurmi, E.L. et al. Linkage disequilibrium at the Angelman syndrome gene UBE3A in autism families. *Genomics* **77**, 105-13 (2001).
112. Bass, M.P. et al. Genetic studies in autistic disorder and chromosome 15. *Neurogenetics* **2**, 219-26 (2000).
113. Vincent, J.B. et al. 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 (2000).
114. Sultana, R. et al. Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics* **80**, 129-34 (2002).
115. Casas, K.A. et al. Chromosome 2q terminal deletion: report of 6 new patients and review of phenotype-breakpoint correlations in 66 individuals. *Am J Med Genet A* **130**, 331-9 (2004).
116. Zneimer, S., Schneider, N. & Richards, C. 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 (1993).
117. Kishino, T., Lalande, M. & Wagstaff, J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* **15**, 70-3 (1997).
118. Ozcelik, T. et al. Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. *Nat Genet* **2**, 265-9 (1992).
119. Yan, W.L. et al. 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 (2000).
120. Wassink, T.H. et al. A case of autism and uniparental disomy of chromosome 1. *Hum Genet* **117**, 200-6 (2005).
121. Lukusa, T., Vermeesch, J.R., Holvoet, M., Fryns, J.P. & Devriendt, K. Deletion 2q37.3 and autism: molecular cytogenetic mapping of the candidate region for autistic disorder. *Genet Couns* **15**, 293-301 (2004).
122. Oliveira, G. et al. Partial tetrasomy of chromosome 3q and mosaicism in a child with autism. *J Autism Dev Disord* **33**, 177-85 (2003).

123. Fahsold, R., Rott, H.D., Claussen, U. & Schmalenberger, B. Tuberous sclerosis in a child with de novo translocation t(3;12) (p26.3;q23.3). *Clin Genet* **40**, 326-8 (1991).
124. Ramanathan, S. et al. A case of autism with an interstitial deletion on 4q leading to hemizygoty for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Med Genet* **5**, 10 (2004).
125. Sabaratnam, M., Turk, J. & Vroegop, P. Case report: autistic disorder and chromosomal abnormality 46, XX duplication (4) p12-p13. *Eur Child Adolesc Psychiatry* **9**, 307-11 (2000).
126. Nasr, A. & Roy, M. Association of a balanced chromosomal translocation (4;12)(q21.3; q15), affective disorder and autism. *J Intellect Disabil Res* **44 (Pt 2)**, 170-4 (2000).
127. Liu, Q.J. et al. Detection of chromosome aberrations in Chinese children with autism using G-banding and BAC FISH. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **22**, 254-7 (2005).
128. Harvard, C. et al. A variant Cri du Chat phenotype and autism spectrum disorder in a subject with de novo cryptic microdeletions involving 5p15.2 and 3p24.3-25 detected using whole genomic array CGH. *Clin Genet* **67**, 341-51 (2005).
129. Kroisel, P.M. et al. De novo translocation t(5;18)(q33.1;q12.1) associated with autistic disorder. *Am J Med Genet A* **129**, 98-100 (2004).
130. Tentler, D. et al. A balanced reciprocal translocation t(5;7)(q14;q32) associated with autistic disorder: molecular analysis of the chromosome 7 breakpoint. *Am J Med Genet* **105**, 729-36 (2001).
131. Vostanis, P., Harrington, R., Prendergast, M. & Farndon, P. Case reports of autism with interstitial deletion of chromosome 17 (p11.2 p11.2) and monosomy of chromosome 5 (5pter-->5p15.3). *Psychiatr Genet* **4**, 109-11 (1994).
132. Burd, L., Martsolf, J.T., Kerbeshian, J. & Jalal, S.M. Partial 6p trisomy associated with infantile autism. *Clin Genet* **33**, 356-9 (1988).
133. Sukumar, S. et al. Subtle overlapping deletions in the terminal region of chromosome 6q24.2-q26: three cases studied using FISH. *Am J Med Genet* **87**, 17-22 (1999).
134. Wolpert, C.M. et al. De novo partial duplication of chromosome 7p in a male with autistic disorder. *Am J Med Genet* **105**, 222-5 (2001).
135. Demori, E. et al. Supernumerary ring chromosome 8: clinical and molecular cytogenetic characterization in a case report. *Am J Med Genet A* **130**, 288-94 (2004).
136. Borg, I. et al. A cryptic deletion of 2q35 including part of the PAX3 gene detected by breakpoint mapping in a child with autism and a de novo 2;8 translocation. *J Med Genet* **39**, 391-9 (2002).
137. Ishikawa-Brush, Y. et al. Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3' to the SDC2 gene. *Hum Mol Genet* **6**, 1241-50 (1997).

138. Bolton, P. et al. Autism, mental retardation, multiple exostoses and short stature in a female with 46,X,t(X;8)(p22.13;q22.1). *Psychiatr Genet* **5**, 51-5 (1995).
139. Verri, A. et al. Chromosome 10p deletion in a patient with hypoparathyroidism, severe mental retardation, autism and basal ganglia calcifications. *Ann Genet* **47**, 281-7 (2004).
140. Smith, M. et al. Molecular genetic delineation of a deletion of chromosome 13q12-->q13 in a patient with autism and auditory processing deficits. *Cytogenet Genome Res* **98**, 233-9 (2002).
141. Steele, M.M., Al-Adeimi, M., Siu, V.M. & Fan, Y.S. Brief report: A case of autism with interstitial deletion of chromosome 13. *J Autism Dev Disord* **31**, 231-4 (2001).
142. Merritt, J.L., 2nd, Jalal, S.M., Barbaresi, W.J. & Babovic-Vuksanovic, D. 14q32.3 deletion syndrome with autism. *Am J Med Genet A* **133**, 99-100 (2005).
143. Bolton, P.F. et al. The phenotypic manifestations of interstitial duplications of proximal 15q with special reference to the autistic spectrum disorders. *Am J Med Genet* **105**, 675-85 (2001).
144. Finelli, P. et al. FISH characterisation of an identical (16)(p11.2p12.2) tandem duplication in two unrelated patients with autistic behaviour. *J Med Genet* **41**, e90 (2004).
145. Kerbeshian, J., Severud, R., Burd, L. & Larson, L. Peek-a-boo fragile site at 16d associated with Tourette syndrome, bipolar disorder, autistic disorder, and mental retardation. *Am J Med Genet* **96**, 69-73 (2000).
146. Hebebrand, J. et al. Partial trisomy 16p in an adolescent with autistic disorder and Tourette's syndrome. *Am J Med Genet* **54**, 268-70 (1994).
147. Moog, U. et al. Hereditary motor and sensory neuropathy (HMSN) IA, developmental delay and autism related disorder in a boy with duplication (17)(p11.2p12). *Genet Couns* **15**, 73-80 (2004).
148. Anneren, G., Dahl, N., Uddenfeldt, U. & Janols, L.O. Asperger syndrome in a boy with a balanced de novo translocation: t(17;19)(p13.3;p11). *Am J Med Genet* **56**, 330-1 (1995).
149. Sauter, S. et al. Autistic disorder and chromosomal mosaicism 46,XY[123]/46,XY,del(20)(pter --> p12.2)[10]. *Am J Med Genet A* **120**, 533-6 (2003).
150. Carratala, F. et al. A patient with autistic disorder and a 20/22 chromosomal translocation. *Dev Med Child Neurol* **40**, 492-5 (1998).
151. Michaelis, R.C. et al. Intersitial deletion of 20p: new candidate region for Hirschsprung disease and autism? *Am J Med Genet* **71**, 298-304 (1997).
152. Nair-Miranda, K. et al. An investigation into sub-telomeric deletions of chromosome 22 and pervasive developmental disorders. *Am J Med Genet B Neuropsychiatr Genet* **125**, 99-104 (2004).
153. Goizet, C. et al. Case with autistic syndrome and chromosome 22q13.3 deletion detected by FISH. *Am J Med Genet* **96**, 839-44 (2000).

154. MacLean, J.E., Teshima, I.E., Szatmari, P. & Nowaczyk, M.J. Ring chromosome 22 and autism: report and review. *Am J Med Genet* **90**, 382-5 (2000).
155. Donnelly, S.L. et al. Female with autistic disorder and monosomy X (Turner syndrome): parent-of-origin effect of the X chromosome. *Am J Med Genet* **96**, 312-6 (2000).
156. Thomas, N.S. et al. Xp deletions associated with autism in three females. *Hum Genet* **104**, 43-8 (1999).
157. Bachevalier, J. Medial temporal lobe structures and autism: a review of clinical and experimental findings. *Neuropsychologia* **32**, 627-48 (1994).
158. Bachevalier, J. Brief report: medial temporal lobe and autism: a putative animal model in primates. *J Autism Dev Disord* **26**, 217-20 (1996).
159. Wolterink, G. et al. Early amygdala damage in the rat as a model for neurodevelopmental psychopathological disorders. *Eur Neuropsychopharmacol* **11**, 51-9 (2001).
160. Amaral, D.G. The amygdala, social behavior, and danger detection. *Ann N Y Acad Sci* **1000**, 337-47 (2003).
161. Lawrence, K., Kuntsi, J., Coleman, M., Campbell, R. & Skuse, D. Face and emotion recognition deficits in Turner syndrome: a possible role for X-linked genes in amygdala development. *Neuropsychology* **17**, 39-49 (2003).
162. Edith Frank, J. & Tomaz, C. Lateralized impairment of the emotional enhancement of verbal memory in patients with amygdala-hippocampus lesion. *Brain Cogn* **52**, 223-30 (2003).
163. Good, C.D. et al. Dosage-sensitive X-linked locus influences the development of amygdala and orbitofrontal cortex, and fear recognition in humans. *Brain* (2003).
164. Rodier, P.M., Ingram, J.L., Tisdale, B., Nelson, S. & Romano, J. Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. *J Comp Neurol* **370**, 247-61 (1996).
165. Ingram, J.L., Peckham, S.M., Tisdale, B. & Rodier, P.M. Prenatal exposure of rats to valproic acid reproduces the cerebellar anomalies associated with autism. *Neurotoxicol Teratol* **22**, 319-24 (2000).
166. Andres, C. Molecular genetics and animal models in autistic disorder. *Brain Res Bull* **57**, 109-19 (2002).
167. Narita, N. et al. 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 (2002).
168. Insel, T.R., O'Brien, D.J. & Leckman, J.F. Oxytocin, vasopressin, and autism: is there a connection? *Biol Psychiatry* **45**, 145-57 (1999).
169. Ferguson, J.N. et al. Social amnesia in mice lacking the oxytocin gene. *Nat Genet* **25**, 284-8 (2000).
170. Lim, M.M., Bielsky, I.F. & Young, L.J. Neuropeptides and the social brain: potential rodent models of autism. *Int J Dev Neurosci* **23**, 235-43 (2005).
171. Green, L. et al. Oxytocin and autistic disorder: alterations in peptide forms. *Biol Psychiatry* **50**, 609-13 (2001).

172. Modahl, C. et al. Plasma oxytocin levels in autistic children. *Biol Psychiatry* **43**, 270-7 (1998).
173. McCracken, J.T. et al. Risperidone in children with autism and serious behavioral problems. *N Engl J Med* **347**, 314-21 (2002).
174. Anderson, L.T. et al. The effects of haloperidol on discrimination learning and behavioral symptoms in autistic children. *J Autism Dev Disord* **19**, 227-39 (1989).
175. DiLavore, P.C., Lord, C. & Rutter, M. The pre-linguistic autism diagnostic observation schedule. *J Autism Dev Disord* **25**, 355-79 (1995).
176. Tadevosyan-Leyfer, O. et al. A principal components analysis of the Autism Diagnostic Interview-Revised. *J Am Acad Child Adolesc Psychiatry* **42**, 864-72 (2003).
177. Shao Y et al. Fine Mapping of Autistic Disorder to Chromosome 15q11-q13 by Use of Phenotypic Subtypes. *Am J Hum Genet* **72**, 539-48 (2003).
178. Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-56 (2000).
179. Ardlie KG , Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3**, 299-309 (2002).
180. Abecasis, G.R. et al. Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* **68**, 191-197 (2001).
181. Moffatt, M.F., Traherne, J.A., Abecasis, G.R. & Cookson, W.O. Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. *Hum Mol Genet* **9**, 1011-9 (2000).
182. Ardlie, K.G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3**, 299-309 (2002).
183. Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* **69**, 1-14 (2001).
184. Weiss, K.M. & Clark, A.G. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* **18**, 19-24 (2002).
185. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44 (1999).
186. Cardon, L.R. & Palmer, L.J. Population stratification and spurious allelic association. *Lancet* **361**, 598-604 (2003).
187. Spielman, R., McGinnis, R. & Ewens, W. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**, 506-16 (1993).
188. Falk, C. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Annals of Human Genetics* **51**, 227-233 (1987).
189. Hey, J. What's so hot about recombination hotspots? *PLoS Biol* **2**, e190 (2004).
190. Cook EH Jr et al. Evidence of linkage between the serotonin transporter and autistic disorder. *Mol Psychiatry* **2**, 247-50 (1997).
191. Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L.M. & Millonig, J.H. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Mol Psychiatry* **9**, 474-84 (2004).

192. Ramoz, N. et al. Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. *Am J Psychiatry* **161**, 662-9 (2004).
193. Lujan, R., Shigemoto, R. & Lopez-Bendito, G. Glutamate and GABA receptor signalling in the developing brain. *Neuroscience* **130**, 567-80 (2005).
194. Sutcliffe, J.S., Nurmi, E.L. & Lombroso, P.J. Genetics of childhood disorders: XLVII. Autism, part 6: duplication and inherited susceptibility of chromosome 15q11-q13 genes in autism. *J Am Acad Child Adolesc Psychiatry* **42**, 253-6 (2003).
195. Aldred, S., Moore, K.M., Fitzgerald, M. & Waring, R.H. Plasma amino acid levels in children with autism and their families. *J Autism Dev Disord* **33**, 93-7 (2003).
196. Moreno-Fuenmayor, H., Borjas, L., Arrieta, A., Valera, V. & Socorro-Candanoza, L. Plasma excitatory amino acids in autism. *Invest Clin* **37**, 113-28 (1996).
197. Dhossche, D. et al. Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism. *Med Sci Monit* **8**, PR1-6 (2002).
198. Blatt, G.J. et al. Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. *J Autism Dev Disord* **31**, 537-43 (2001).
199. Buxbaum, J.D. et al. Association between a GABRB3 polymorphism and autism. *Mol Psychiatry* **7**, 311-6 (2002).
200. Ma, D.Q. et al. Identification of Significant Association and Gene-Gene Interaction of GABA Receptor Subunit Genes in Autism. *Am J Hum Genet* **77**, 377-88 (2005).
201. D'Arcangelo, G. et al. A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374**, 719-23 (1995).
202. DeSilva, U. et al. The human reelin gene: isolation, sequencing, and mapping on chromosome 7. *Genome Res* **7**, 157-64 (1997).
203. D'Arcangelo, G. et al. Reelin is a ligand for lipoprotein receptors. *Neuron* **24**, 471-9 (1999).
204. Persico, A. et al. Collaborative Linkage Study of Autism (). Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. *Mol Psychiatry* **6**, 150-9 (2001).
205. Krebs, M.O. et al. Absence of association between a polymorphic GGC repeat in the 5' untranslated region of the reelin gene and autism. *Mol Psychiatry* **7**, 801-4 (2002).
206. Zhang, H. et al. Reelin gene alleles and susceptibility to autism spectrum disorders. *Mol Psychiatry* **7**, 1012-7 (2002).
207. Fatemi, S.H., Stary, J.M. & Egan, E.A. Reduced blood levels of reelin as a vulnerability factor in pathophysiology of autistic disorder. *Cell Mol Neurobiol* **22**, 139-52 (2002).
208. Fatemi, S.H. et al. Reelin signaling is impaired in autism. *Biol Psychiatry* **57**, 777-87 (2005).

209. Ingram, J.L. et al. Discovery of allelic variants of HOXA1 and HOXB1: genetic susceptibility to autism spectrum disorders. *Teratology* **62**, 393-405 (2000).
210. Conroy, J. et al. Serotonin transporter gene and autism: a haplotype analysis in an Irish autistic population. *Mol Psychiatry* **9**, 587-93 (2004).
211. Mulder, E.J. et al. Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. *Am J Med Genet B Neuropsychiatr Genet* **133**, 93-6 (2005).
212. Tordjman, S. et al. Role of the serotonin transporter gene in the behavioral expression of autism. *Mol Psychiatry* **6**, 434-9 (2001).
213. Yirmiya, N. et al. Evidence for an association with the serotonin transporter promoter region polymorphism and autism. *Am J Med Genet* **105**, 381-6 (2001).
214. Barnby, G. et al. Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *Am J Hum Genet* **76**, 950-66 (2005).
215. Bottini, N. et al. Autism: evidence of association with adenosine deaminase genetic polymorphism. *Neurogenetics* **3**, 111-3 (2001).
216. Persico, A.M. et al. Adenosine deaminase alleles and autistic disorder: case-control and family-based association studies. *Am J Med Genet* **96**, 784-90 (2000).
217. Persico, A.M. et al. Enhanced APOE2 transmission rates in families with autistic probands. *Psychiatr Genet* **14**, 73-82 (2004).
218. Petit, E. et al. Association study with two markers of a human homeogene in infantile autism. *J Med Genet* **32**, 269-74 (1995).
219. Gong, X. et al. Association between the FOXP2 gene and autistic disorder in Chinese population. *Am J Med Genet B Neuropsychiatr Genet* **127**, 113-6 (2004).
220. Shuang, M. et al. Family-based association study between autism and glutamate receptor 6 gene in Chinese Han trios. *Am J Med Genet B Neuropsychiatr Genet* **131**, 48-50 (2004).
221. Jamain, S. et al. Linkage and association of the glutamate receptor 6 gene with autism. *Mol Psychiatry* **7**, 302-10 (2002).
222. Beyer, K.S., Klauck, S.M., Benner, A., Poustka, F. & Poustka, A. Association studies of the HOPA dodecamer duplication variant in different subtypes of autism. *Am J Med Genet* **114**, 110-5 (2002).
223. Conciatori, M. et al. Association between the HOXA1 A218G polymorphism and increased head circumference in patients with autism. *Biol Psychiatry* **55**, 413-9 (2004).
224. Comings, D.E., Wu, S., Chiu, C., Muhleman, D. & Sverd, J. Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Res* **63**, 25-32 (1996).
225. Heralut J et al. Possible association of c-Harvey-Ras-1 (HRAS-1) marker with autism. *Psychiatry Res* **46**, 261-7 (1993).
226. Hutcheson, H.B. et al. Examination of NRCAM, LRRN3, KIAA0716, and LAMB1 as autism candidate genes. *BMC Med Genet* **5**, 12 (2004).

227. Cohen, I.L. et al. Association of autism severity with a monoamine oxidase A functional polymorphism. *Clin Genet* **64**, 190-7 (2003).
228. Marui, T. et al. Association between the neurofibromatosis-1 (NF1) locus and autism in the Japanese population. *Am J Med Genet B Neuropsychiatr Genet* **131**, 43-7 (2004).
229. Ylisaukko-Oja, T. et al. Analysis of four neuroligin genes as candidates for autism. *Eur J Hum Genet* (2005).
230. Wu, S. et al. Positive association of the oxytocin receptor gene (OXTR) with autism in the Chinese Han population. *Biol Psychiatry* **58**, 74-7 (2005).
231. D'Amelio, M. et al. Paraoxonase gene variants are associated with autism in North America, but not in Italy: possible regional specificity in gene-environment interactions. *Mol Psychiatry* (2005).
232. Philippi, A. et al. Haplotypes in the gene encoding protein kinase c-beta (PRKCB1) on chromosome 16 are associated with autism. *Mol Psychiatry* (2005).
233. Persico, A.M. et al. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. *Mol Psychiatry* **6**, 150-9 (2001).
234. Skaar, D.A. et al. Analysis of the RELN gene as a genetic risk factor for autism. *Mol Psychiatry* **10**, 563-71 (2005).
235. Nabi, R., Serajee, F.J., Chugani, D.C., Zhong, H. & Huq, A.H. Association of tryptophan 2,3 dioxygenase gene polymorphism with autism. *Am J Med Genet B Neuropsychiatr Genet* **125**, 63-8 (2004).
236. Vourc'h, P. et al. Mutation screening and association study of the UBE2H gene on chromosome 7q32 in autistic disorder. *Psychiatr Genet* **13**, 221-5 (2003).
237. Prosser, J. Detecting single-base mutations. *Trends Biotechnol* **11**, 238-46 (1993).
238. Kleymenova, E., Muga, S., Fischer, S. & Walker, C.L. Application of high-performance liquid chromatography-based analysis of DNA fragments to molecular carcinogenesis. *Mol Carcinog* **29**, 51-8 (2000).
239. Xiao, W. & Oefner, P.J. Denaturing high-performance liquid chromatography: A review. *Hum Mutat* **17**, 439-74 (2001).
240. Fisher, S.E., Vargha-Khadem, F., Watkins, K.E., Monaco, A.P. & Pembrey, M.E. Localisation of a gene implicated in a severe speech and language disorder. *Nat Genet* **18**, 168-70 (1998).
241. Macdermot, K.D. et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* **76**, 1074-80 (2005).
242. Wassink, T.H. et al. Evaluation of FOXP2 as an autism susceptibility gene. *Am J Med Genet* **114**, 566-9 (2002).
243. Newbury, D.F. & Monaco, A.P. Molecular genetics of speech and language disorders. *Curr Opin Pediatr* **14**, 696-701 (2002).
244. O'Brien, E.K., Zhang, X., Nishimura, C., Tomblin, J.B. & Murray, J.C. Association of specific language impairment (SLI) to the region of 7q31. *Am J Hum Genet* **72**, 1536-43 (2003).

245. Gauthier, J. et al. Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. *Am J Med Genet A* **118**, 172-5 (2003).
246. Marui, T. et al. No association of FOXP2 and PTPRZ1 on 7q31 with autism from the Japanese population. *Neurosci Res* (2005).
247. Jamain, S. et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* **34**, 27-9 (2003).
248. Chih, B., Afridi, S.K., Clark, L. & Scheiffele, P. Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum Mol Genet* **13**, 1471-7 (2004).
249. Yan, J. et al. Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry* **10**, 329-32 (2005).
250. Vincent, J.B. et al. Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. *Am J Med Genet B Neuropsychiatr Genet* **129**, 82-4 (2004).
251. Gauthier, J. et al. NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *Am J Med Genet B Neuropsychiatr Genet* **132**, 74-5 (2005).
252. Vincent, J.B. et al. The RAY1/ST7 tumor-suppressor locus on chromosome 7q31 represents a complex multi-transcript system. *Genomics* **80**, 283-94 (2002).
253. Bacchelli, E. et al. Screening of nine candidate genes for autism on chromosome 2q reveals rare nonsynonymous variants in the cAMP-GEFII gene. *Mol Psychiatry* **8**, 916-24 (2003).
254. Wassink, T.H. et al. Evaluation of the chromosome 2q37.3 gene CENTG2 as an autism susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet* **136**, 36-44 (2005).
255. Shinahara, K., Saijo, T., Mori, K. & Kuroda, Y. Single-strand conformation polymorphism analysis of the FMR1 gene in autistic and mentally retarded children in Japan. *J Med Invest* **51**, 52-8 (2004).
256. Vincent, J.B. et al. Association and transmission analysis of the FMR1 IVS10 + 14C-T variant in autism. *Am J Med Genet B Neuropsychiatr Genet* **125**, 54-6 (2004).
257. Bonora, E. et al. Mutation screening and association analysis of six candidate genes for autism on chromosome 7q. *Eur J Hum Genet* **13**, 198-207 (2005).
258. Lam, C.W. et al. Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. *J Med Genet* **37**, E41 (2000).
259. Li, H., Yamagata, T., Mori, M., Yasuhara, A. & Momoi, M.Y. Mutation analysis of methyl-CpG binding protein family genes in autistic patients. *Brain Dev* **27**, 321-5 (2005).
260. Shibayama, A. et al. MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. *Am J Med Genet B Neuropsychiatr Genet* **128**, 50-3 (2004).
261. Castermans, D. et al. The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *J Med Genet* **40**, 352-6 (2003).

262. Laumonnier, F. et al. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* **74**, 552-7 (2004).
263. Wassink, T.H. et al. Evidence supporting WNT2 as an autism susceptibility gene. *Am J Med Genet* **105**, 406-13 (2001).
264. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nat Genet* **37 Suppl**, S38-45 (2005).
265. Vargas, D.L., Nascimbene, C., Krishnan, C., Zimmerman, A.W. & Pardo, C.A. Neuroglial activation and neuroinflammation in the brain of patients with autism. *Ann Neurol* **57**, 67-81 (2005).
266. Auld, D.S. & Robitaille, R. Glial cells and neurotransmission: an inclusive view of synaptic function. *Neuron* **40**, 389-400 (2003).
267. Fields, R.D. & Stevens-Graham, B. New insights into neuron-glia communication. *Science* **298**, 556-62 (2002).
268. Nedergaard, M., Ransom, B. & Goldman, S.A. New roles for astrocytes: redefining the functional architecture of the brain. *Trends Neurosci* **26**, 523-30 (2003).
269. Martin, E.R., Monks, S.A., Warren, L.L. & Kaplan, N.L. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* **67**, 146-54 (2000).
270. Martin, E.R., Bass, M.P. & Kaplan, N.L. Correcting for a potential bias in the pedigree disequilibrium test. *Am J Hum Genet* **68**, 1065-7 (2001).
271. Clayton, D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* **65**, 1170-7 (1999).
272. Dudbridge, F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* **25**, 115-21 (2003).
273. Cordell, H.J. & Clayton, D.G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* **70**, 124-41 (2002).
274. Sham, P. & Curtis, D. An extended transmission/disequilibrium test (TDT) for multi-allelic marker loci. *Ann Hum Genet* **59**, 323-336 (1995).
275. Abecasis, G. & Cookson, W. GOLD--graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182-3 (2000).
276. Russel, P. *Genetics*, Glossary (Tje Bemjamin/Cummings Publishing Company, Inc, California, 1998).
277. Zabetian, C.P. et al. The structure of linkage disequilibrium at the DBH locus strongly influences the magnitude of association between diallelic markers and plasma dopamine beta-hydroxylase activity. *Am J Hum Genet* **72**, 1389-400 (2003).
278. Consortium, T.I.H. The International HapMap Project. *Nature* **426**, 789-96 (2003).
279. Carlson, N. *Physiology of behavior*. (ed. Boston) 96-129 (Boston: Allyn and Bacon, 2001).

280. Simiand, J. et al. Neuropsychopharmacological profile in rodents of SR 57746A, a new, potent 5-HT_{1A} receptor agonist. *Fundam Clin Pharmacol* **7**, 413-27 (1993).
281. Krakowski, M. Violence and serotonin: influence of impulse control, affect regulation, and social functioning. *J Neuropsychiatry Clin Neurosci* **15**, 294-305 (2003).
282. Goodman, M., New, A. & Siever, L. Trauma, genes, and the neurobiology of personality disorders. *Ann N Y Acad Sci* **1032**, 104-16 (2004).
283. Siegel, J.M. The neurotransmitters of sleep. *J Clin Psychiatry* **65 Suppl 16**, 4-7 (2004).
284. Shively, C.A. & Bethea, C.L. Cognition, mood disorders, and sex hormones. *Ilar J* **45**, 189-99 (2004).
285. Cook EH Jr et al. Platelet serotonin studies in hyperserotonemic relatives of children with autistic disorder. *Life Sci.* **52**, 2005-15 (1993).
286. Schain, R. & Freedman, D. Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *J Pediatr* **58**, 315-320 (1961).
287. Mulder, E.J. et al. Platelet serotonin levels in pervasive developmental disorders and mental retardation: diagnostic group differences, within-group distribution, and behavioral correlates. *J Am Acad Child Adolesc Psychiatry* **43**, 491-9 (2004).
288. Ritvo ER, Y.A., Geller E, Ornitz EM, Saeger K, Plotkin S. Increased blood serotonin and platelets in early infantile autism. *Arch Gen Psychiatry* **23**, 566-72 (1970).
289. Minderaa, R.B. et al. Whole blood serotonin and tryptophan in autism: temporal stability and the effects of medication. *J Autism Dev Disord* **19**, 129-36 (1989).
290. Campbell, M., EFriedman, Devito, E., Greenspan, L. & Collins, P. Blood serotonin in psychotic and brain-damaged children. *J Autism Child Schizophr* **4**(1974).
291. Hoshino, Y. et al. Blood serotonin and free tryptophan concentration in autistic children. *Neuropsychobiology* **11**, 22-7 (1984).
292. McBride, P.A. et al. Effects of diagnosis, race, and puberty on platelet serotonin levels in autism and mental retardation. *J Am Acad Child Adolesc Psychiatry* **37**, 767-76 (1998).
293. Naffah-Mazzacoratti, M.G. et al. Serum serotonin levels of normal and autistic children. *Braz J Med Biol Res* **26**, 309-17 (1993).
294. Cook, E.H., Jr. et al. Autistic children and their first-degree relatives: relationships between serotonin and norepinephrine levels and intelligence. *J Neuropsychiatry Clin Neurosci* **2**, 268-74 (1990).
295. Leventhal, B.L., Cook, E.H., Jr., Morford, M., Ravitz, A. & Freedman, D.X. Relationships of whole blood serotonin and plasma norepinephrine within families. *J Autism Dev Disord* **20**, 499-511 (1990).
296. Hoyer, D. et al. International Union of Pharmacology classification of receptors for 5-hydroxytryptamine (Serotonin). *Pharmacol Rev* **46**, 157-203 (1994).

297. McBride, P.A. et al. Serotonergic responsivity in male young adults with autistic disorder. Results of a pilot study. *Arch Gen Psychiatry* **46**, 213-21 (1989).
298. Wapnir, R. & Stevenson, J. Estimation of free tryptophan in plasma: A simplified spectrofluorometric micromethod. *Clin Chim Acta* **26**, 203-206 (1969).
299. Herault, J. et al. Serotonin and autism: biochemical and molecular biology features. *Psychiatry Res* **65**, 33-43 (1996).
300. Stahl, S.M., Woo, D.J., Mefford, I.N., Berger, P.A. & Ciaranello, R.D. Hyperserotonemia and platelet serotonin uptake and release in schizophrenia and affective disorders. *Am J Psychiatry* **140**, 26-30 (1983).
301. Irwin, M., Belendiuk, K., McCloskey, K. & Freedman, D.X. Tryptophan metabolism in children with attentional deficit disorder. *Am J Psychiatry* **138**, 1082-5 (1981).
302. Hanley, H.G., Stahl, S.M. & Freedman, D.X. Hyperserotonemia and amine metabolites in autistic and retarded children. *Arch Gen Psychiatry* **34**, 521-31 (1977).
303. Comings, D.E. & Comings, B.G. Clinical and genetic relationships between autism-pervasive developmental disorder and Tourette syndrome: a study of 19 cases. *Am J Med Genet* **39**, 180-91 (1991).
304. Lam, K., Aman, M. & Arnold, L. Neurochemical correlates of autistic disorder: A review of the literature. *Research in Developmental Disabilities*, In Press (2005).
305. Cohen, D.J., Shaywitz, B.A., Johnson, W.T. & Bowers, M., Jr. Biogenic amines in autistic and atypical children. Cerebrospinal fluid measures of homovanillic acid and 5-hydroxyindoleacetic acid. *Arch Gen Psychiatry* **31**, 845-53 (1974).
306. Cohen, D.J., Caparulo, B.K., Shaywitz, B.A. & Bowers, M.B., Jr. Dopamine and serotonin metabolism in neuropsychiatrically disturbed children. CSF homovanillic acid and 5-hydroxyindoleacetic acid. *Arch Gen Psychiatry* **34**, 545-50 (1977).
307. Winsberg, B.G., Sverd, J., Castells, S., Hurwic, M. & Perel, J.M. Estimation of monoamine and cyclic-AMP turnover and amino acid concentrations of spinal fluid in autistic children. *Neuropediatrics* **11**, 250-5 (1980).
308. Gillberg, C., Svennerholm, L. & Hamilton-Helberg, C. Childhood psychosis and monoamine metabolites in spinal fluid. *Journal of Autism and Developmental Disorders* **13**, 383-396 (1983).
309. Gillberg, C. & Svennerholm, L. CSF monoamines in autistic syndromes and other pervasive developmental disorders of early childhood. *Br J Psychiatry* **151**, 89-94 (1987).
310. Ross, D., Kylvlyo, W. & Anderson, G. Cerebrospinal fluid levels of homovanillic acid and 5-hydroxyindolacetic acid in autism. *Annals of Neurology* **18**, 394 (1985).
311. Narayan M, Srinath S, Anderson GM & DB, M. Cerebrospinal fluid levels of homovanillic acid and 5-hydroxyindoleacetic acid in autism. *Biol Psychiatry* **33**, 630-5 (1993).

312. Croonenberghs, J., Verkerk, R., Scharpe, S., Deboutte, D. & Maes, M. Serotonergic disturbances in autistic disorder: L-5-hydroxytryptophan administration to autistic youngsters increases the blood concentrations of serotonin in patients but not in controls. *Life Sci* **76**, 2171-83 (2005).
313. McDougle CJ et al. A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder. *Arch Gen Psychiatry*. **53**, 1001-8 (1996).
314. Varley, C.K. & Holm, V.A. A two-year follow-up of autistic children treated with fenfluramine. *J Am Acad Child Adolesc Psychiatry* **29**, 137-40 (1990).
315. Hollander, E. et al. The relationship between repetitive behaviors and growth hormone response to sumatriptan challenge in adult autistic disorder. *Neuropsychopharmacology* **22**, 163-7 (2000).
316. Chugani, D.C. et al. Altered serotonin synthesis in the dentatohalamocortical pathway in autistic boys. *Ann Neurol* **42**, 666-9 (1997).
317. Chugani DC et al. Developmental changes in brain serotonin synthesis capacity in autistic and nonautistic children. *Ann Neurol* **45**, 287-95 (1999).
318. Parks, C.L., Robinson, P.S., Sibille, E., Shenk, T. & Toth, M. Increased anxiety of mice lacking the serotonin1A receptor. *Proc Natl Acad Sci U S A* **95**, 10734-9 (1998).
319. Gross, C. et al. Serotonin1A receptor acts during development to establish normal anxiety-like behaviour in the adult. *Nature* **416**, 396-400 (2002).
320. Moiseiwitsch JR & JM, L. Serotonin regulates mouse cranial neural crest migration. *Proc Natl Acad Sci U S A* **92**, 7182-6 (1995).
321. Lavdas AA, Blue ME, Lincoln J & JG, P. Serotonin promotes the differentiation of glutamate neurons in organotypic slice cultures of the developing cerebral cortex. *J Neurosci* **17**, 7872-80 (1997).
322. Deckert, J. et al. Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder. *Hum Mol Genet* **8**, 621-4 (1999).
323. Jones, M.B., Palmour, R.M., Zwaigenbaum, L. & Szatmari, P. Modifier effects in autism at the MAO-A and DBH loci. *Am J Med Genet B Neuropsychiatr Genet* **126**, 58-65 (2004).
324. Filinger, E.J., Garcia-Cotto, M.A., Vila, S., Gerbaldo, H. & Jerez, D. Possible relationship between pervasive developmental disorders and platelet monoamine oxidase activity. *Braz J Med Biol Res* **20**, 161-4 (1987).
325. Takahashi, S., Kanai, H. & Miyamoto, Y. Monoamine oxidase activity in blood platelets from autistic children. *Folia Psychiatr Neurol Jpn* **31**, 597-603 (1977).
326. Cohen, D.J., Young, J.G. & Roth, J.A. Platelet monoamine oxidase in early childhood autism. *Arch Gen Psychiatry* **34**, 534-7 (1977).
327. Veenstra-VanderWeele, J. et al. Transmission disequilibrium studies of the serotonin 5-HT2A receptor gene (HTR2A) in autism. *Am J Med Genet* **114**, 277-83 (2002).
328. Lassig JP et al. Physical mapping of the serotonin 5-HT(7) receptor gene (HTR7) to chromosome 10 and pseudogene (HTR7P) to chromosome 12, and testing of linkage disequilibrium between HTR7 and autistic disorder. *Am J Med Genet* **88**, 472-5 (1999).

329. Fernstrom, J.D. Role of precursor availability in control of monoamine biosynthesis in brain. *Physiol Rev* **63**, 484-546 (1983).
330. Comings, D.E. et al. Exon and intron variants in the human tryptophan 2,3-dioxygenase gene: potential association with Tourette syndrome, substance abuse and other disorders. *Pharmacogenetics* **6**, 307-18 (1996).
331. Comings, D.E. Clinical and molecular genetics of ADHD and Tourette syndrome. Two related polygenic disorders. *Ann N Y Acad Sci* **931**, 50-83 (2001).
332. Ramamoorthy S, B.A., Moore KR, Han H, Yang-Feng T, Chang AS, Ganapathy V, Blakely RD. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *Proc Natl Acad Sci U S A* **90**, 2542-6 (1993).
333. Heils A et al. Allelic variation of human serotonin transporter gene expression. *J Neurochem* **66**, 2621-4 (1996).
334. MacKenzie A & J, Q. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proc Natl Acad Sci U S A* **96**, 15251-5 (1999).
335. Persico AM et al. Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples. *Am J Med Genet* **96**, 123-7 (2000).
336. Walther, D.J. et al. Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* **299**, 76 (2003).
337. Kumer, S., Mockus, S., Rucker, P. & Vrana, K. Amino-Terminal Analysis of Tryptophan Hydroxylase: Protein Kinase Phosphorylation Occurs at Serine-58. *J Neurochem* **69**, 1738-1745 (1997).
338. Zhang, X., Beaulieu, J.M., Sotnikova, T.D., Gainetdinov, R.R. & Caron, M.G. Tryptophan hydroxylase-2 controls brain serotonin synthesis. *Science* **305**, 217 (2004).
339. Zhang, X. et al. Loss-of-function mutation in tryptophan hydroxylase-2 identified in unipolar major depression. *Neuron* **45**, 11-6 (2005).
340. Sheehan, K. et al. Tryptophan hydroxylase 2 (TPH2) gene variants associated with ADHD. *Mol Psychiatry* (2005).
341. Sham, P.C., Cherny, S.S., Purcell, S. & Hewitt, J.K. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* **66**, 1616-30 (2000).
342. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Bmj* **315**, 629-34 (1997).
343. Wu, S. et al. Lack of evidence for association between the serotonin transporter gene (SLC6A4) polymorphisms and autism in the Chinese trios. *Neurosci Lett* **381**, 1-5 (2005).
344. Lowe, N. et al. Joint analysis of the DRD5 marker concludes association with attention-deficit/hyperactivity disorder confined to the predominantly inattentive and combined subtypes. *Am J Hum Genet* **74**, 348-56 (2004).
345. Lotrich, F.E. & Pollock, B.G. Meta-analysis of serotonin transporter polymorphisms and affective disorders. *Psychiatr Genet* **14**, 121-9 (2004).

346. Cook, E.H., Jr. et al. Platelet serotonin studies in hyperserotonemic relatives of children with autistic disorder. *Life Sci* **52**, 2005-15 (1993).
347. Ernst, M., Zametkin, A.J., Matochik, J.A., Pascualvaca, D. & Cohen, R.M. Low medial prefrontal dopaminergic activity in autistic children. *Lancet* **350**, 638 (1997).
348. Anderson, L.T. et al. Haloperidol in the treatment of infantile autism: effects on learning and behavioral symptoms. *Am J Psychiatry* **141**, 1195-202 (1984).
349. Lake CR, Ziegler MG & DL, M. Increased norepinephrine levels and decreased dopamine-beta-hydroxylase activity in primary autism. *Arch Gen Psychiatry*. **34**(553-6 (1977).
350. Goldstein, M. Dopamine-beta-hydroxylase studies: Dopamine-beta hydroxylase and endogenous total 5-hydroxyindoles levels in autistic patients and controls. in *The Autistic Syndromes* (ed. M, C.) 57-59 (North Holland Publishing Co., 1976).
351. Garnier, C. et al. Dopamine-beta-hydroxylase (DBH) and homovanillic acid (HVA) in autistic children. *J Autism Dev Disord* **16**, 23-9 (1986).
352. Coleman, M., M, C., LS, F. & al, e. Serum Dopamine Beta hydroxylase levels in Down's syndrome. *Clin Genet* **5**, 312-315 (1974).
353. Gurguis, G.N. & Uhde, T.W. Plasma 3-methoxy-4-hydroxyphenylethylene glycol (MHPG) and growth hormone responses to yohimbine in panic disorder patients and normal controls. *Psychoneuroendocrinology* **15**, 217-24 (1990).
354. Charney, D.S. et al. Adrenergic receptor sensitivity in depression. Effects of clonidine in depressed patients and healthy subjects. *Arch Gen Psychiatry* **39**, 290-4 (1982).
355. Sullivan, G.M., Coplan, J.D., Kent, J.M. & Gorman, J.M. The noradrenergic system in pathological anxiety: a focus on panic with relevance to generalized anxiety and phobias. *Biol Psychiatry* **46**, 1205-18 (1999).
356. Palmen, S.J., van Engeland, H., Hof, P.R. & Schmitz, C. Neuropathological findings in autism. *Brain* **127**, 2572-83 (2004).
357. Roman, T. et al. Further evidence for the association between attention-deficit/hyperactivity disorder and the dopamine-beta-hydroxylase gene. *Am J Med Genet* **114**, 154-8 (2002).
358. Cubells, J.F. et al. A haplotype at the DBH locus, associated with low plasma dopamine beta-hydroxylase activity, also associates with cocaine-induced paranoia. *Mol Psychiatry* **5**, 56-63 (2000).
359. Wood, J.G., Joyce, P.R., Miller, A.L., Mulder, R.T. & Kennedy, M.A. A polymorphism in the dopamine beta-hydroxylase gene is associated with "paranoid ideation" in patients with major depression. *Biol Psychiatry* **51**, 365-9 (2002).
360. Yirmiya N et al. Family-based and population study of a functional promoter-region monoamine oxidase A polymorphism in autism: possible association with IQ. *Am J Med Genet* **114**, 284-7 (2002).
361. Lauritsen MB et al. Investigation of two variants in the DOPA decarboxylase gene in patients with autism. *Am J Med Genet* **114**, 466-70 (2002).

362. O'Brien, R.A., Semenuk, G., Coleman, M. & Spector, S. Catechol-O-methyltransferase activity in erythrocytes of children with autism. *Clin Exp Pharmacol Physiol* **3**, 9-14 (1976).
363. Gilbert, J.R. et al. Physical and cDNA mapping in the DBH region of human chromosome 9q34. *Hum Hered* **50**, 151-7 (2000).
364. GeneCards. GeneCard for gene DBH GC09P131843. (2005).
365. Kobayashi, K., Kurosawa Y, Fujita K & T., N. Human dopamine beta-hydroxylase gene: two mRNA types having different 3'-terminal regions are produced through alternative polyadenylation. *Nucleic Acids Res.*, 1089-102 (1989).
366. Zaret, K.S. & Sherman, F. Mutationally altered 3' ends of yeast CYC1 mRNA affect transcript stability and translational efficiency. *J Mol Biol* **177**, 107-35 (1984).
367. Shaw, G. & Kamen, R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-67 (1986).
368. Treisman, R. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. *Cell* **42**, 889-902 (1985).
369. Dahlstrom, A. & K, F. Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiol Scand.* **62**, 1-55 (1964).
370. Hokfelt, T., R, M., S, B., S, K. & M, G. Distribution maps of tyrosine-hydroxylase-immunoreactive neurons in the rat brain. in *Handbook of Chemical Neuroanatomy*, Vol. 2, Classical Transmitters in the CNS (eds. Bjorklund, A. & Hokfelt, T.) 277-379 (Elsevier, New York, 1984).
371. Swanson, L.W. & Hartman, B.K. The central adrenergic system. An immunofluorescence study of the location of cell bodies and their efferent connections in the rat utilizing dopamine-B-hydroxylase as a marker. *J. Comp Neurol* **163**, 467-505 (1975).
372. Dunnette, J. & Weinshilboum, R. Human serum dopamine beta-hydroxylase: correlation of enzymatic activity with immunoreactive protein in genetically defined samples. *Am J Hum Genet* **28**, 155-66 (1976).
373. Weinshilboum, R.M., Raymond, F.A., Elveback, L.R. & Weidman, W.H. Serum dopamine-beta-hydroxylase activity: sibling-sibling correlation. *Science* **181**, 943-5 (1973).
374. RM, W. Serum dopamine beta-hydroxylase. *Pharmacol Rev* **30**, 133-166 (1978).
375. Hoyle, G.W., Mercer, E.H., Palmiter, R.D. & Brinster, R.L. Cell-specific expression from the human dopamine beta-hydroxylase promoter in transgenic mice is controlled via a combination of positive and negative regulatory elements. *J Neurosci* **14**, 2455-63 (1994).
376. Tuchman, R. & Rapin, I. Epilepsy in autism. *Lancet Neurol* **1**, 352-8 (2002).
377. Wong, V. Epilepsy in children with autistic spectrum disorder. *J Child Neurol* **8**, 316-22 (1993).

378. Giovanardi Rossi, P., Posar, A. & Parmeggiani, A. Epilepsy in adolescents and young adults with autistic disorder. *Brain Dev* **22**, 102-6 (2000).
379. Rutter, M.L. Psycho-social disorders in childhood, and their outcome in adult life. *J R Coll Physicians Lond* **4**, 211-8 (1970).
380. Olsson, I., Steffenburg, S. & Gillberg, C. Epilepsy in autism and autisticlike conditions. A population-based study. *Arch Neurol* **45**, 666-8 (1988).
381. Rutter, M. Psycho-social disorders in childhood, and their outcome in adult life. *J R Coll Physicians Lond* **4**, 211-8 (1970).
382. Jobe, P.C., Ko, K.H. & Dailey, J.W. Abnormalities in norepinephrine turnover rate in the central nervous system of the genetically epilepsy-prone rat. *Brain Res* **290**, 357-60 (1984).
383. Dailey, J.W. & Jobe, P.C. Indices of noradrenergic function in the central nervous system of seizure-naive genetically epilepsy-prone rats. *Epilepsia* **27**, 665-70 (1986).
384. Browning, R.A., Wade, D.R., Marcinczyk, M., Long, G.L. & Jobe, P.C. Regional brain abnormalities in norepinephrine uptake and dopamine beta-hydroxylase activity in the genetically epilepsy-prone rat. *J Pharmacol Exp Ther* **249**, 229-35 (1989).
385. Lauterborn, J.C. & Ribak, C.E. Differences in dopamine beta-hydroxylase immunoreactivity between the brains of genetically epilepsy-prone and Sprague-Dawley rats. *Epilepsy Res* **4**, 161-76 (1989).
386. Dailey, J.W., Mishra, P.K., Ko, K.H., Penny, J.E. & Jobe, P.C. Noradrenergic abnormalities in the central nervous system of seizure-naive genetically epilepsy-prone rats. *Epilepsia* **32**, 168-73 (1991).
387. Szot, P. et al. Norepinephrine-deficient mice have increased susceptibility to seizure-inducing stimuli. *J Neurosci* **19**, 10985-92 (1999).
388. Thomas, S., AM, M. & RD, P. Noradrenaline is essential for mouse fetal development. *Nature* **374**, 643 - 646 (1995).
389. Thomas, S.A. & Palmiter, R.D. Impaired maternal behavior in mice lacking norepinephrine and epinephrine. *Cell* **91**, 583-92 (1997).
390. Munoz, A., Serrano, C., Garcia-Estan, J., Quesada, T. & Miras Portugal, M.T. Effect of diabetic hyperglycemia and other sugars on plasma dopamine-beta-hydroxylase activity. *Diabetes* **33**, 1127-32 (1984).
391. Simonyi, A., Kanyicska, B., Szentendrei, T. & Fekete, M.I. Effects of beta-adrenergic inhibitors on noradrenaline and adrenaline metabolism in brain regions of the rat. *Eur J Pharmacol* **98**, 285-8 (1984).
392. Rockson, S.G. et al. Plasma dopamine-beta-hydroxylase activity in oral contraceptive hypertension. *Circulation* **51**, 916-23 (1975).
393. Robinson PD, Schutz CK, Macciardi F, White BN & JJ., H. Genetically determined low maternal serum dopamine beta-hydroxylase levels and the etiology of autism spectrum disorders. *Am J Med Genet.* **100**, 30-6 (2001).
394. Hawi, Z. et al. Linkage disequilibrium mapping at DAT1, DRD5 and DBH narrows the search for ADHD susceptibility alleles at these loci. *Mol Psychiatry* **8**, 299-308 (2003).

395. Spielman RS, McGinnis RE & WJ, E. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**, 506-16 (1993).
396. Freedman, L.S. et al. Changes in human serum dopamine-β-hydroxylase activity with age. *Nature* **236**, 310-1 (1972).
397. Goldin, L.R. et al. Segregation and linkage studies of plasma dopamine-beta-hydroxylase (DBH), erythrocyte catechol-O-methyltransferase (COMT), and platelet monoamine oxidase (MAO): possible linkage between the ABO locus and a gene controlling DBH activity. *Am J Hum Genet* **34**, 250-62 (1982).
398. Weinshilboum, R.M., Schorott, H.G., Raymond, F.A., Weidman, W.H. & Elveback, L.R. Inheritance of very low serum dopamine-beta-hydroxylase activity. *Am J Hum Genet* **27**, 573-85 (1975).
399. Zabetian, C. et al. A quantitative-trait analysis of human plasma-dopamine beta-hydroxylase activity: evidence for a major functional polymorphism at the DBH locus. *Am J Hum Genet.* **68**, 515-22 (2001).
400. Kim, H.S., Seo, H., Yang, C., Brunet, J.F. & Kim, K.S. Noradrenergic-specific transcription of the dopamine beta-hydroxylase gene requires synergy of multiple cis-acting elements including at least two Phox2a-binding sites. *J Neurosci* **18**, 8247-60 (1998).
401. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. & Tjian, R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079-90 (1987).
402. Thomas, S.A. & Palmiter, R.D. Disruption of the dopamine beta-hydroxylase gene in mice suggests roles for norepinephrine in motor function, learning, and memory. *Behav Neurosci* **111**, 579-89 (1997).
403. Buxbaum, J. et al. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* **68**, 1514-20 (2001).
404. Shao, Y. et al. Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am J Hum Genet* **70**, 1058-61 (2002).
405. Roberts, S.B., MacLean, C.J., Neale, M.C., Eaves, L.J. & Kendler, K.S. Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* **65**, 876-84 (1999).
406. Darvasi, A., Weinreb, A., Minke, V., Weller, J.I. & Soller, M. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943-51 (1993).
407. Gallagher, L. et al. Brief report: A case of autism associated with del(2)(q32.1q32.2) or (q32.2q32.3). *J Autism Dev Disord* **33**, 105-8 (2003).
408. Ghaziuddin, M. & Burmeister, M. Deletion of chromosome 2q37 and autism: a distinct subtype? *J Autism Dev Disord* **29**, 259 -63 (1999).
409. Taysi, K., Dengler, D.R., Jones, L.A. & Heersma, J.R. Interstitial deletion of the long arm of chromosome 2: case report and review of literature. *Ann Genet* **24**, 245-7 (1981).

410. Al-Awadi, S.A. et al. Interstitial deletion of the long arm of chromosome 2: del(2)(q31q33). *J Med Genet* **20**, 464-5 (1983).
411. Buchanan, P., Rhodes, R. & Stevenson, C. Interstitial deletion 2q31 leads to q33. *Am J Med Genet* **15**, 121-6 (1983).
412. Franceschini, P., Cirillo Silengo, M., Davi, G., Bianco, R. & Biagioli, M. Interstitial deletion of the long arm of chromosome 2 (q31q33) in a girl with multiple anomalies and mental retardation. *Hum Genet* **64**, 98 (1983).
413. Young, R. et al. Deletion 2q: two new cases with karyotypes 46, XY, del(2)q(31q33) and 46,XY,del(2)(q36). *J Med Genet* **20**(1983).
414. Ramer, J.C., Ladda, R.L., Frankel, C.A. & Beckford, A. A review of phenotype-karyotype correlations in individuals with interstitial deletions of the long arm of chromosome 2. *Am J Med Genet* **32**, 359-63 (1989).
415. Pinkstaff, J.K., Detterich, J., Lynch, G. & Gall, C. Integrin subunit gene expression is regionally differentiated in adult brain. *J Neurosci* **19**, 1541-56 (1999).
416. Bader, B.L., Rayburn, H., Crowley, D. & Hynes, R.O. Extensive vasculogenesis, angiogenesis, and organogenesis precede lethality in mice lacking all alpha v integrins. *Cell* **95**, 507-19 (1998).
417. Yang, J.T., Rayburn, H. & Hynes, R.O. Cell adhesion events mediated by alpha 4 integrins are essential in placental and cardiac development. *Development* **121**, 549-60 (1995).
418. LeDouarin, N. *The Neural Crest*, (Cambridge Univ. Press, 1982).
419. Kil, S.H., Krull, C.E., Cann, G., Clegg, D. & Bronner-Fraser, M. The alpha4 subunit of integrin is important for neural crest cell migration. *Dev Biol* **202**, 29-42 (1998).
420. Werner, A. et al. Impaired axonal regeneration in alpha7 integrin-deficient mice. *J Neurosci* **20**, 1822-30 (2000).
421. Dulabon, L. et al. Reelin binds alpha3beta1 integrin and inhibits neuronal migration. *Neuron* **27**, 33-44 (2000).
422. Forster, E. et al. Reelin, Disabled 1, and beta 1 integrins are required for the formation of the radial glial scaffold in the hippocampus. *Proc Natl Acad Sci USA* **99**, 13178-83 (2002).
423. Rohrbough, J., Grotewiel, M.S., Davis, R.L. & Broadie, K. Integrin-mediated regulation of synaptic morphology, transmission, and plasticity. *J Neurosci* **20**, 6868-78 (2000).
424. Hemler, M.E., Elices, M.J., Parker, C. & Takada, Y. Structure of the integrin VLA-4 and its cell-cell and cell-matrix adhesion functions. *Immunol Rev* **114**, 45-65 (1990).
425. Baron, J.L., Madri, J.A., Ruddle, N.H., Hashim, G. & Janeway, C.A., Jr. Surface expression of alpha 4 integrin by CD4 T cells is required for their entry into brain parenchyma. *J Exp Med* **177**, 57-68 (1993).
426. Myers, K.J. et al. Antisense oligonucleotide blockade of alpha 4 integrin prevents and reverses clinical symptoms in murine experimental autoimmune encephalomyelitis. *J Neuroimmunol* **160**, 12-24 (2005).
427. Relton, J. Inhibition of alpha4 integrin to protect the brain against ischemic injury. *Drug News Perspect* **14**, 346-52 (2001).

428. Baird, G., Cass, H. & Slonims, V. Diagnosis of autism. *Bmj* **327**, 488-93 (2003).
429. Fombonne, E., Simmons, H., Ford, T., Meltzer, H. & Goodman, R. Prevalence of pervasive developmental disorders in the British nationwide survey of child mental health. *J Am Acad Child Adolesc Psychiatry* **40**, 820-7 (2001).
430. Ashwood, P. & Van de Water, J. Is autism an autoimmune disease? *Autoimmun Rev* **3**, 557-62 (2004).
431. Karp, G. *Cell and Molecular Biology*, (John Wiley & Sons, New York, 2002).
432. Heymann, G.A., Kiesewetter, H. & Salama, A. Frequencies of alpha4 A3061G variants and identification of three new variants of the human integrin alpha4-subunit. *Mol Immunol* **39**, 855-60 (2003).
433. Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. & Contopoulos-Ioannidis, D.G. Replication validity of genetic association studies. *Nat Genet* **29**, 306-9 (2001).
434. Williams, N.M. et al. Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry* **61**, 336-44 (2004).
435. Bray, J. et al. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. *Hum Mol Genet* **14**, 1947-1954 (2005).
436. Hynes, R.O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25 (1992).
437. Milner, R. & Campbell, I.L. The integrin family of cell adhesion molecules has multiple functions within the CNS. *J Neurosci Res* **69**, 286-91 (2002).
438. Gall, C.M., Pinkstaff, J.K., Lauterborn, J.C., Xie, Y. & Lynch, G. Integrins regulate neuronal neurotrophin gene expression through effects on voltage-sensitive calcium channels. *Neuroscience* **118**, 925-40 (2003).
439. Vogelezang, M.G. et al. Alpha4 integrin is expressed during peripheral nerve regeneration and enhances neurite outgrowth. *J Neurosci* **21**, 6732-44 (2001).
440. Le Hir, H., Nott, A. & Moore, M.J. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* **28**, 215-20 (2003).
441. Baron-Cohen, S. et al. Social intelligence in the normal and autistic brain: an fMRI study. *Eur J Neurosci* **11**, 1891-8 (1999).
442. Chan, C.S., Weeber, E.J., Kurup, S., Sweatt, J.D. & Davis, R.L. Integrin requirement for hippocampal synaptic plasticity and spatial memory. *J Neurosci* **23**, 7107-16 (2003).
443. Kramar, E.A., Bernard, J.A., Gall, C.M. & Lynch, G. Integrins modulate fast excitatory transmission at hippocampal synapses. *J Biol Chem* **278**, 10722-30 (2003).
444. Grooms, S.Y. & Jones, L.S. RGDS tetrapeptide and hippocampal in vitro kindling in rats: evidence for integrin-mediated physiological stability. *Neurosci Lett* **231**, 139-42 (1997).

445. Becker, K., Kindrick, D., Relton, J., Harlan, J. & Winn, R. Antibody to the alpha4 integrin decreases infarct size in transient focal cerebral ischemia in rats. *Stroke* **32**, 206-11 (2001).
446. James, W.G., Bullard, D.C. & Hickey, M.J. Critical role of the alpha 4 integrin/VCAM-1 pathway in cerebral leukocyte trafficking in lupus-prone MRL/fas(lpr) mice. *J Immunol* **170**, 520-7 (2003).
447. Kielinen, M., Rantala, H., Timonen, E., Linna, S.L. & Moilanen, I. Associated medical disorders and disabilities in children with autistic disorder: a population-based study. *Autism* **8**, 49-60 (2004).
448. Emsley, J.G. & Hagg, T. alpha6beta1 integrin directs migration of neuronal precursors in adult mouse forebrain. *Exp Neurol* **183**, 273-85 (2003).
449. Fasen, K., Elger, C.E. & Lie, A.A. Distribution of alpha and beta integrin subunits in the adult rat hippocampus after pilocarpine-induced neuronal cell loss, axonal reorganization and reactive astrogliosis. *Acta Neuropathol (Berl)* **106**, 319-22 (2003).
450. Georges-Labouesse, E., Mark, M., Messaddeq, N. & Gansmuller, A. Essential role of alpha 6 integrins in cortical and retinal lamination. *Curr Biol* **8**, 983-6 (1998).
451. Mercado, M.L. et al. Neurite outgrowth by the alternatively spliced region of human tenascin-C is mediated by neuronal alpha7beta1 integrin. *J Neurosci* **24**, 238-47 (2004).
452. Ke, X. et al. Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples. *Hum Mol Genet* **13**, 2557-65 (2004).
453. Iles, M.M. The effect of SNP marker density on the efficacy of haplotype tagging SNPs--a warning. *Ann Hum Genet* **69**, 209-15 (2005).
454. Meng, Z., Zaykin, D.V., Xu, C.F., Wagner, M. & Ehm, M.G. Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am J Hum Genet* **73**, 115-30 (2003).
455. Cadigan, K.M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev* **11**, 3286-305 (1997).
456. Logan, C.Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810 (2004).
457. Muller, H., Samanta, R. & Wieschaus, E. Wingless signaling in the *Drosophila* embryo: zygotic requirements and the role of the frizzled genes. *Development* **126**, 577-86 (1999).
458. Cadigan, K.M., Fish, M.P., Rulifson, E.J. & Nusse, R. Wingless repression of *Drosophila* frizzled 2 expression shapes the Wingless morphogen gradient in the wing. *Cell* **93**, 767-77 (1998).
459. Sato, A., Kojima, T., Ui-Tei, K., Miyata, Y. & Saigo, K. Dfrizzled-3, a new *Drosophila* Wnt receptor, acting as an attenuator of Wingless signaling in wingless hypomorphic mutants. *Development* **126**, 4421-30 (1999).
460. Wehrli, M. et al. arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527-30 (2000).
461. Rousset, R. et al. Naked cuticle targets dishevelled to antagonize Wnt signal transduction. *Genes Dev* **15**, 658-71 (2001).

462. Jho, E.H. et al. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol Cell Biol* **22**, 1172-83 (2002).
463. Spiegelman, V.S. et al. Wnt/beta-catenin signaling induces the expression and activity of betaTrCP ubiquitin ligase receptor. *Mol Cell* **5**, 877-82 (2000).
464. Roose, J. et al. Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science* **285**, 1923-6 (1999).
465. Hovanes, K. et al. Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nat Genet* **28**, 53-7 (2001).
466. Zeng, Y.A. & Verheyen, E.M. Nemo is an inducible antagonist of Wnt signaling during Drosophila wing development. *Development* **131**, 2911-20 (2004).
467. Thorpe, C.J. & Moon, R.T. nemo-like kinase is an essential co-activator of Wnt signaling during early zebrafish development. *Development* **131**, 2899-909 (2004).
468. Thomas, K.R. & Capecchi, M.R. Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**, 847-50 (1990).
469. Bailey, A. et al. A clinicopathological study of autism. *Brain* **121** (Pt 5), 889-905 (1998).
470. Altman, J. Postnatal development of the cerebellar cortex in the Rat. II. Phases in the maturation of Purkinje cells and the molecular layer. *J. Comp Neurol* **145**, 399-464 (1972).
471. Liu, P. et al. Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**, 361-5 (1999).
472. Millar, S.E. et al. WNT signaling in the control of hair growth and structure. *Dev Biol* **207**, 133-49 (1999).
473. Olson, D.J., Christian, J.L. & Moon, R.T. Effect of wnt-1 and related proteins on gap junctional communication in Xenopus embryos. *Science* **252**, 1173-6 (1991).
474. Wang, Y., Thekdi, N., Smallwood, P.M., Macke, J.P. & Nathans, J. Frizzled-3 is required for the development of major fiber tracts in the rostral CNS. *J Neurosci* **22**, 8563-73 (2002).
475. Wang, Y., Huso, D., Cahill, H., Ryugo, D. & Nathans, J. Progressive cerebellar, auditory, and esophageal dysfunction caused by targeted disruption of the frizzled-4 gene. *J Neurosci* **21**, 4761-71 (2001).
476. Simon, H.H., Thuret, S. & Alberi, L. Midbrain dopaminergic neurons: control of their cell fate by the engrailed transcription factors. *Cell Tissue Res* **318**, 53-61 (2004).
477. Millen, K.J., Wurst, W., Herrup, K. & Joyner, A.L. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development* **120**, 695-706 (1994).
478. Baader, S.L., Sanlioglu, S., Berrebi, A.S., Parker-Thornburg, J. & Oberdick, J. Ectopic overexpression of engrailed-2 in cerebellar Purkinje cells causes restricted cell loss and retarded external germinal layer development at lobule junctions. *J Neurosci* **18**, 1763-73 (1998).

479. Wurst, W., Auerbach, A.B. & Joyner, A.L. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development* **120**, 2065-75 (1994).
480. Marie, B., Cruz-Orengo, L. & Blagburn, J.M. Persistent engrailed expression is required to determine sensory axon trajectory, branching, and target choice. *J Neurosci* **22**, 832-41 (2002).
481. Lijam, N. et al. Social interaction and sensorimotor gating abnormalities in mice lacking Dvl1. *Cell* **90**, 895-905 (1997).
482. Lee, M. et al. Nicotinic receptor abnormalities in the cerebellar cortex in autism. *Brain* **125**, 1483-95 (2002).
483. Kemper, T.L. & Bauman, M.L. The contribution of neuropathologic studies to the understanding of autism. *Neurol Clin* **11**, 175-87 (1993).
484. Miller, J.R. The Wnts. *Genome Biol* **3**, REVIEWS3001 (2002).
485. Wainwright, B.J. et al. 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 (1988).
486. Warburton, P. et al. 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 (2000).
487. Ashley-Koch, A. et al. Genetic studies of autistic disorder and chromosome 7. *Genomics* **61**, 227-36 (1999).
488. Hoang, B., Moos, M., Jr., Vukicevic, S. & Luyten, F.P. Primary structure and tissue distribution of FRZB, a novel protein related to Drosophila frizzled, suggest a role in skeletal morphogenesis. *J Biol Chem* **271**, 26131-7 (1996).
489. Wang, S., Krinks, M., Lin, K., Luyten, F.P. & Moos, M., Jr. Frzb, a secreted protein expressed in the Spemann organizer, binds and inhibits Wnt-8. *Cell* **88**, 757-66 (1997).
490. Lin, K. et al. The cysteine-rich frizzled domain of Frzb-1 is required and sufficient for modulation of Wnt signaling. *Proc Natl Acad Sci U S A* **94**, 11196-200 (1997).
491. Dann, C.E. et al. Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature* **412**, 86-90 (2001).
492. Leyns, L., Bouwmeester, T., Kim, S.H., Piccolo, S. & De Robertis, E.M. Frzb-1 is a secreted antagonist of Wnt signaling expressed in the Spemann organizer. *Cell* **88**, 747-56 (1997).
493. Moon, R.T., Brown, J.D., Yang-Snyder, J.A. & Miller, J.R. Structurally related receptors and antagonists compete for secreted Wnt ligands. *Cell* **88**, 725-8 (1997).
494. Pizzuti, A. et al. cDNA characterization and chromosomal mapping of two human homologues of the Drosophila dishevelled polarity gene. *Hum Mol Genet* **5**, 953-8 (1996).
495. Rosso, S.B., Sussman, D., Wynshaw-Boris, A. & Salinas, P.C. Wnt signaling through Dishevelled, Rac and JNK regulates dendritic development. *Nat Neurosci* **8**, 34-42 (2005).
496. Raymond, G.V., Bauman, M.L. & Kemper, T.L. Hippocampus in autism: a Golgi analysis. *Acta Neuropathol (Berl)* **91**, 117-9 (1996).

497. Poole, S.J., Law, M.L., Kao, F.T. & Lau, Y.F. Isolation and chromosomal localization of the human En-2 gene. *Genomics* **4**, 225-31 (1989).
498. Logan, C., Willard, H.F., Rommens, J.M. & Joyner, A.L. Chromosomal localization of the human homeo box-containing genes, EN1 and EN2. *Genomics* **4**, 206-9 (1989).
499. Hooper, J.E. Distinct pathways for autocrine and paracrine Wingless signalling in *Drosophila* embryos. *Nature* **372**, 461-4 (1994).
500. Joyner, A.L., Herrup, K., Auerbach, B.A., Davis, C.A. & Rossant, J. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science* **251**, 1239-43 (1991).
501. Gabriel SB et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
502. Li, J. et al. Lack of evidence for an association between WNT2 and RELN polymorphisms and autism. *Am J Med Genet B Neuropsychiatr Genet* **126**, 51-7 (2004).
503. McCoy, P.A. et al. No association between the WNT2 gene and autistic disorder. *Am J Med Genet* **114**, 106-9 (2002).
504. Zhong, H., Serajee, F.J., Nabi, R. & Huq, A.H. No association between the EN2 gene and autistic disorder. *J Med Genet* **40**, e4 (2003).
505. Belmaker, R.H., Hattab, J. & Ebstein, R.P. Plasma dopamine-beta-hydroxylase in childhood psychosis. *J Autism Child Schizophr* **8**, 293-8 (1978).
506. Garnier C et al. Dopamine-beta-hydroxylase (DBH) and homovanillic acid (HVA) in autistic children. *J Autism Dev Disord.* **16**, 23-9 (1986).
507. Straub, R.E. et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet* **71**, 337-48 (2002).
508. Schwab, S.G. et al. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* **72**, 185-90 (2003).
509. Van Den Bogaert, A. et al. The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet* **73**, 1438-43 (2003).
510. Morris, D.W. et al. No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. *Schizophr Res* **60**, 167-72 (2003).
511. Talbot, K. et al. Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest* **113**, 1353-63 (2004).
512. Hill, E.W., Jobling, M.A. & Bradley, D.G. Y-chromosome variation and Irish origins. *Nature* **404**, 351-2 (2000).
513. Rosenberg, N.A. et al. Genetic structure of human populations. *Science* **298**, 2381-5 (2002).
514. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-7 (1996).

515. Kleinjan, D.A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* **76**, 8-32 (2005).
516. de Kok, Y.J. et al. Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum Mol Genet* **5**, 1229-35 (1996).
517. Nyholt, D.R. Genetic case-control association studies--correcting for multiple testing. *Hum Genet* **109**, 564-7 (2001).
518. Bastone, L., Reilly, M., Rader, D.J. & Foulkes, A.S. MDR and PRP: a comparison of methods for high-order genotype-phenotype associations. *Hum Hered* **58**, 82-92 (2004).
519. Ritchie, M.D. et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* **69**, 138-47 (2001).
520. Kuemerle B, Zanjani H, Joyner A and Herrup K. Pattern Deformities and Cell Loss in *Engrailed-2* Mutant Mice Suggest Two Separate Patterning Events during Cerebellar Development. *J Neuroscience* **17**, 7881-7889 (1997).
521. Murphy M, McHugh B, Tighe O, Mayne P, O'Neill C, Naughten E and Croke D. The genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers. *Eur J Hum Genet* **7**, 549-554 (1999).
522. Dawson GWP. The frequencies of the ABO and Rh (D) blood groups in Ireland from a sample of 1 in the 18 of the population. *Ann Hum Genet* **28**, 49-59 (1964).
523. O'Donnell K, O'Neill C, Tighe O, Bertorelle G, Naughten E, Mayne P and Croke D. The mutation spectrum of hyperphenylalaninaemia in the Republic of Ireland: the population history of the Irish revisited. *Eur J Hum Genet* **10**, 530-538 (2002).

APPENDIX I

REAGENTS AND SOLUTIONS

1.1 DNA Extraction Buffers (Chapter 2, Section 2.2.2.1.1)

Lysis Buffer (10X)

- Sucrose 547.7g
- 2M Tris (pH 7.5) 25ml
- 1M MgCl₂ 25ml
- Triton X100 50ml
- H₂O up to 1L

Suspension Buffer (10X)

- 2M Tris (pH 7.5) 5ml
- 0.5M EDTA (pH 8.0) 20ml
- 5M NaCl 2ml
- H₂O up to 100ml

TE Buffer

- 2M Tris (pH 7.5) 1ml
- 0.5M EDTA (pH 8.0) 40μl
- H₂O up to 200ml

1.2 PCR Reagents (section 2.2.4)

MgCl₂ Buffer (10X) 10mM/15mM/20mM

- 1M KCl 2.5ml
- 1M Tris (pH 8.9) 0.5ml
- 1M MgCl₂ 50µl/75µl/100µl
- 1% Gelatin 0.5ml
- Triton X100 50µl
- H₂O up to 5ml

Triton and Gelatin Free MgCl₂ Buffer (10X)

10mM/15mM/20mM

- 1M KCl 2.5ml
- 1M Tris (pH 8.9) 0.5ml
- 1M MgCl₂ 50µl/75µl/100µl
- H₂O up to 5ml

1.3 Gel Electrophoresis and Visualisation (Chapter 2, Section 2.2.7)

TAE Buffer (50X)

- Trizma Base 242g
- 0.5M EDTA (pH 8.0) 100ml
- Glacial Acetic Acid 57.1ml
- H₂O up to 1L

40% Polyacrylamide Solution

- Acrylamide 39g
- Bisacrylamide 1g
- H₂O up to 100ml

Non Denaturing Polyacrylamide Gel (6%)

- 40% Polyacrylamide 9ml
- TBE (10X) 6ml
- H₂O 45ml
- 10% APS 0.5ml
- TEMED 80 μ l

Denaturing Polyacrylamide Gels for Use with the ABI PRISM 377 DNA Sequencer (12% and 6% (w/v))

- Long Ranger (ABI) 6ml/3ml
- TBE (10X) 3ml
- H₂O 13ml/16ml
- Urea 11g
- 10% APS 150 μ l
- TEMED 21 μ l

Loading Dye

- Bromophenol Blue 12.5mg
- Xylene Cyanol 12.5mg
- Glycerol 15ml
- H₂O up to 50ml

Formamide Loading Dye

- Formamide 10ml
- Xylene Cyanol 10mg
- Bromophenol Blue 10mg
- 0.5M EDTA (pH 8.0) 200 μ l

SNaPshot Buffer

- 1M Tris (pH 9.0) 8ml
- 2mM MgCl₂ 8ml
- H₂O 34ml

1.3.2 Silver Staining Solutions

Silver Staining Solution 1

- 100% Alcohol 50ml
- Acetic Acid 2.5ml
- H₂O up to 500ml

Silver Staining Solution 2

- 1% Silver Nitrate 25ml
- H₂O 225ml

Silver Staining Solution 3

- NaOH 7.5g
- H₂O up to 497ml
- Formaldehyde 3ml

Silver Staining Solution 4

- NaCO₃ 3.75g
- H₂O up to 250ml

1.4 DHPLC Solutions (Chapter 2, Section 2.2.9)

Buffer A

- 2M TEAA (Transgenomic) 50ml
- Acetonitrile (HPLC grade) 250μl
- H₂O up to 1L

Buffer B

- 2M TEAA (Transgenomic) 50ml
- Acetonitrile (HPLC grade) 250ml
- H₂O up to 1L

Buffer C

- Acetonitrile (HPLC grade) 250ml
- H₂O up to 1L

Buffer D

- Acetonitrile (HPLC grade) 250ml
- H₂O up to 1L

1.5 Solutions For Measurement of DBH Activity

Mix A

- 1M NaOH (pH 5.0) 250μl
- 0.2M Fumarate 62.5μl
- 0.2M Ascorbate 62.5μl
- 0.002M Parglyine 62.5μl
- Catalase 125μl
- H₂O 250μl

Appendix II

Amplification and SNaPShot™ Conditions

This Appendix lists the primers for each of the variants genotyped in this thesis. These include the primers used to genotype the serotonin transporter (Table I), *TPH2* variants (Table II), *DBH* (Table III), *ITGA4* (Table IV) and Wnt pathway (Table V) variants used and the DHPLC amplicon primers used in *ITGA4* mutation screen of exons 16 and 17 (Table VI). It also describes the annealing temperature (T_m), and any additional variations to the standard PCR protocol (section 2.2).

PCR primers are labelled 'F' for the forward primer and 'R' for the reverse primer. All primers are orientated 5' to 3' and were designed using the Primer3 software (www.es.embet.org/cgi-bin/primer3_www.cgi), except where otherwise specified. SNaPshot primers, which were designed by hand based on the sequence surrounding the SNP, are listed in a similar fashion and preceded by the letter 'e'. All primers were ordered from Invitrogen (www.invitrogen.com).

All variants were amplified following the following cycling conditions: an initial denaturing step of 2 min at 94°C followed by 41 cycles of 94 °C for 1 min, Annealing temp for 1 min, 72 °C for 1 min. The amplification ended with a final extension period of 10 min at 72 °C. Any alterations to the above cycling conditions can be found in the respective tables.

Extension primer sequences for allelic discrimination using the SNaPShot™ technique are given below. Extension primer sequences were designed by hand according to the sequence surrounding the SNP. SNaPShot™ reactions were carried out according to Section 2.2.4.3.

Marker Name	Forward Primer	Reverse Primer	Extension Primer	Annealing Temp	Modifications
In/Del	5'-GGG GTT GCC GCT CTG AAT GC-3'	5'-GAG GAC TGA GCT GGA CAA CCA-3'	NA	61	Replacement of dGTP with dGTP: 7-deaza-dGTP (in a ratio of 1:1)
rs2020936	5'-TTA TTT TAA AAG CCT ATC AAG C-3'	5'-CCC CAG GCT CAA GAA TGC-3'	5'-TG AGC AGG GTG AGG TTA TGG AGA-3'	56	
VNTR	5'-TGG ATT TCC TTC TCT CAG TGA TTG G-3'	5'-TCA TGT TCC TAG TCT TAC GCC AGT G-3'	NA	57	35 cycles of 94 ⁰ C (30 sec), 57 ⁰ C (60 sec) and 72 ⁰ C (120 sec)
rs2020942	5'GAC ACT CAT TCC CCA GCG TA-3'	5'-TGT GCA AAT CAG AAA GGT CCA T-3'	5'-GAA CAC ATG GTT TTA TTC TCG AGC C-3'	58	
3'UTR	5'CCG CTT GAA TGC TGT GTA ACA CAC-3'	5'-GTA CCC TTC CAA TAA TAA CCT CC-3'	NA	56	35 cycles of 94 ⁰ C (30 sec), 57 ⁰ C (30 sec) and 72 ⁰ C (45 sec)

Table I Primer Sequences and Conditions used to Genotype Serotonin Transporter Variants. All primers sequences were designed by Kim et al ³, and provided by personal communication with Ed Cook, University of Chicago, Illinois.

Marker Name	Forward Primer	Reverse Primer	Extension Primer	Annealing Temperature	Modifications
rs4570625	5'-TGC ATA GAG GCA TCA CAG GA-3'	5'-TCT TAT CCC TCC CAT CAG CA-3'	5'-CAC ACA TTT GCA TGC ACA AAA TTA-3'	57°C	30 cycles of 94°C (30 sec), 57°C (30 sec) and 72°C (30 sec)
rs4341581	5'-CCA TAG GAT TCA ACG AGG CTA-3'	5'-GAA GTT GCC GTG TCA CTC AT-3'	5'-GAA AGG CAG GAA TTT TTC ATA TAT GAT ATT-3'	57°C	30 cycles of 94°C (30 sec), 57°C (30 sec) and 72°C (30 sec)

Table II Primer Sequences and Conditions used to genotype *TPH2* variants. Both of these variants underwent allele discrimination using the SNaPShot Method of Analysis and the ABI-3100.

Marker Number	Marker Name	Forward Primer	Reverse Primer	Annealing Temp	Modifications
1	In/Del	GCT TGG GTG TGG TCT GGA GC	CCT TTC TGG GTC CTG CTT CTC C	72 – 62 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
2	C-2124T	GGA GGA GAC GCT GGG TTG G	GGG TAT TGA GGC CCC AGG C	70 – 60 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
3	C-1021T	GGA GGG ACA GCT CTC AGT CC	TAC CTC TCC CTC CTG TCC TCT CGC	70 – 60 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
4	Ex2	GCA ATG AAT GCG GAG CTC TG	GGC CCC AGA CTT ATC AGG GAC	72 – 62 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
5	Ex5	GCC CTC TCA GGA CAC ACC	ACA CAG CTG AGT CCT AGG G	70 – 60 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
6	Ex11	CCA GGG ACA GGA CTC GAG TTG	AGC AGT TTG GAG TGC AGA CCC	70 – 60 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)
7	Ex12	CCC ATG GAA CAG CCG TGC	ACA CCC TCT CAG CCA TGC AG	72 – 62 C	35 cycles of 94 ⁰ C (30 sec), A ⁰ C (30 sec) and 72 ⁰ C (30 sec)

Table III Primer Sequences and Conditions used to genotype *DBH* variants. All variants were amplified using the touch-down technique, whereby the initial annealing temperature (A) (highest) decreases by 2⁰C each cycle until it reaches the lower annealing temperature. All variants also had an initial denaturing step of 94⁰C for 10 minutes.

Exon	Forward Primer	Reverse Primer	Annealing Temp	Modifications
16	AGT GTT TGG CCC TTT TCA GG	TCC GAC ATA TAG TTC ACT TCT TC	65 °C	40 cycles of 94°C (30 sec), 65 °C (30 sec) and 72 °C (30 sec), followed by a final extension period of 72 °C for 2 min
17	CAT GCA TCA CAG GTG TCG	GGA ATG CAA ATG CAC ACA	56 °C	40 cycles of 94°C (30 sec), 65 °C (30 sec) and 72 °C (30 sec), followed by a final extension period of 72 °C for 2 min

Table IV Amplification of Exons 16 and 17 regions in the *ITGA4* gene.

Marker Name	Forward Primer	Reverse Primer	Extension Primer	Annealing Temp	Modifications	Genotyping Method	Products
rs12690517	AGT GTT TGG CCC TTT TCA GG	TCC GAC ATA TAG TTC ACT TCT TC	AAT AAA GGA AAA TAT TCC TAC	65 °C	40 cycles of 94 °C (30 sec), 65 °C (30 sec) and 72 °C (30 sec), followed by a final extension period of 72 °C for 2 min	SNaPShot and Analysis on ABI3100	C (21bp) or T (21bp)

Table V Amplification and Genotyping Conditions for rs12690517 in *ITGA4* gene.

Marker Name	Forward Primer	Reverse Primer	Annealing Temp	Modifications	Product Size
Exon 1	CAT GAA GAG TTG ACC TCG G	GCT GAG CGC TTC TGC TCT GGG CAC	64 °C	40 cycles of 94°C (30 sec), 64 °C (30 sec) and 72 °C (30 sec), followed by a final extension period of 72 °C for 2 min	155 bp
Exon 5	GAA GGG AAG GTG GAT GGT GAC	CTT GTC TTC CAG GCT CCC T	59 °C	40 cycles of 94°C (30 sec), 59 °C (30 sec) and 72 °C (30 sec), followed by a final extension period of 72 °C for 2 min	273 bp

Table VI Amplification of *WNT2* Exons 1 and 5 for SSCP Analysis.

Marker Name	Forward Primer	Reverse Primer	Extension Primer	Annealing Temperature
rs307354	5'- CTT TCT GTC CAG GCC ACT GT -3'	5'- GGT CCC ATC ACT ATG CCT GT-3'	5'- CTG GGT GAC TCT GT - 3'	60°C
rs307371	5'- TGC TCC TGC TGT GTA CTT GG-3'	5'- GCC TCC TGC CTC TAC CAT AG -3'	5'- GAT CCA GGA GCC AG -3'	60 °C
rs307356	5'- ACT ACA TTC GCT TGG CAT CC -3'	5'- CTA CGG AGG ACC CTC ATC AA -3'	5'- CAT CTC CAA CTA CAT TCG CTT GGC ATC CAG GG -3'	60°C
rs307359	5'- GAC CCC TAC CTA GCC CTC TG -3'	5'- CAT GGG GTC ATG GTC TTA CC -3'	5'- AGA GCA GAG AAG GTC ACT TCT C -3'	60 °C

Table VII Amplification Conditions for *DVL1* Gene Variants. All products were analysed using the SNaPShot Technique.

Marker Name	Forward Primer	Reverse Primer	Extension Primer	Annealing Temperature	Modifications
rs3757846	5'- CAA TCA GCC CTA GGG TGT CC -3'	5'- CTC CTT CCC ATT TGC TCT TG-3'	5'- CCG CCT CCC CCG CAC TGA GGG TA-3'	60°C	MgCl ₂ concentration in PCR Buffer increased to 2
rs1861972	5'- GCT CCC TAA AGC CGA TTC AT-3'	5'- CTC CCA ACT GGG ATT CAG AC -3'	5'- GGT CAC CAC TCC CTG CCA -3'	60 °C	MgCl ₂ concentration in PCR Buffer increased to 2
rs1861973	5'- GCT CCC TAA AGC CGA TTC AT -3'	5'- CTC CCA ACT GGG ATT CAG AC -3'	5'- CTA GAA GCC TTA CAG CGA CCC TG -3'	60°C	MgCl ₂ concentration in PCR Buffer increased to 2
rs2361689	5'- TGA GCC TCA ACG AGT CAC AG -3'	5'- ATT GTT TAG CGC GGA CTG AG -3'	5'- GCT GGC CGT GCA CCT CAT GGC ACA GGG C -3'	60°C	
rs1861958	5'- GCG AAG TTC AAC AGG TCT CC -3'	5'- CCC CAA TCC TGC TTA GGA A -3'	5'- GTG CTG GAC CGA GGC GCT TTC CCC TGG GGA C -3'	60 °C	MgCl ₂ concentration in PCR Buffer increased to 2

Table VIII Amplification of *EN2* variants. All variants were genotyped using the SNaPShot Technique.

1.6 Genotyping Conditions and Mutation Detection Conditions

Restriction Digestion Conditions are also given for those variants that underwent RFLP analysis. In all cases, unless stated otherwise, incubation was for 3¹/₂ hours at 37°C. Conditions used for mutation screening of the ITGA4 gene (see Section 5.3.5) are also given.

Marker Name	Method of Allele Detection	Restriction Enzyme	Enzyme Incubation Conditions	Gel Type	Conc. Of Gel	Fragment Sizes
In/Del	Size Discrimination	None	None	Agarose	2%	528bp (in) 484bp (del)
rs2020936	SNaPshot	None	None	ABI-377 polyacrylamide	12%	23bp (C) 23bp (T)
VNTR	Size Discrimination	None	None	Agarose	4% (made with "low melt" agarose)	345bp (9) 360bp (10) 390bp (12)
rs2020942	SNaPshot	None	None	ABI-377 polyacrylamide	12%	21bp (G) 21bp (A)
3'UTR	RFLP	<i>MseI</i>	37°C for 31½ hours	Agarose	2%	741bp (G) 689+52 (T)

Table IX Allele Discrimination Method and Conditions for Serotonin Transporter Variants