DIANNE F NEWBURY
Green College, University of Oxford

A GENOME WIDE SCREEN FOR LOCI INVOLVED IN SPECIFIC LANGUAGE IMPAIRMENT

The work in this thesis was supported by a grant from the Medical Research Council (MRC) and was completed at The Wellcome Trust Centre for Human Genetics, Oxford.

This thesis was submitted for the degree of Doctor of Philosophy in Trinity 2002 and is approximately 48,000 words in length
(excluding appendices, bibliography, figures and tables)

Dianne Newbury
Green College, University of Oxford

A Genome Wide Screen for Loci Involved in Specific Language Impairment (SLI)

ABSTRACT

Approximately 4% of English-speaking children are affected by Specific Language Impairment (SLI); a disorder in the development of language skills despite adequate opportunity and normal intelligence. Several studies have indicated the importance of genetic factors in SLI; a positive family history confers an increased risk of development, and monozygotic concordance consistently exceeds that of dizygotic twins. However, like many behavioural traits, SLI is assumed to be genetically complex with several loci contributing to the overall risk.

This thesis aims to clarify the genetic mechanisms underlying Specific Language Impairment by the exploitation of recent advances in technological, genetic and statistical techniques. This goal is achieved, for the main part, through the completion of the first-ever, systematic genome-wide screen for loci involved in the disorder. A collection of 98 families was drawn from both epidemiological and clinical populations, all with probands who display severe deficits in language skills. Genome-wide linkage analyses were completed for three language-related measures and identified two regions which may harbour susceptibility gene variants for SLI, one on chromosome 16 and a second on chromosome 19. Both of these loci yielded maximum LOD scores of 3.55 and exceeded the threshold for suggestive linkage under all types of analysis performed. Fine mapping of the chromosome 19 locus with a high-density map of microsatellite markers provided further support for the role of this region in SLI but failed to narrow the area of linkage.

The second section of the thesis therefore explores alternative genetic strategies that may facilitate the localisation of susceptibility variants from the genomic regions identified. Mutation screening and association analyses were performed for two candidate genes within a subset of 48 families affected by SLI. The first — numblike (*NBL*), or numb-related (*NUMB-R*) (MIM 604018) — was selected from the region of linkage on chromosome 19q and the second — Forkhead-bOX domain P2 (*FOXP2*) (MIM 605317) — has recently been shown to be mutated in a family with a severe speech and language disorder.

Finally, I describe the mapping of a translocation breakpoint within a child affected by a severe language impairment and orofacial dyspraxia. This breakpoint lies on chromosome 2q and coincides with a putative region of linkage in both language impairment and autism.

In the long-term it is hoped that techniques similar to those described here will allow the identification of the gene variants which underlie SLI allowing to the development of better diagnosis and treatment for those children with language impairments.

Submitted for the degree of Doctor of Philosophy Trinity 2002

This thesis is dedicated to:

Mikey, Lomu and Daisy

For introducing me to the important things in life. Long may they last.

And Mum and Dad

For 28 years of unconditional love, encouragement and belief.

Many thanks go to:

John Armour, Tony Monaco and Simon Fisher

For inspiring me to follow this career path, for being brave enough to provide me with the opportunities I needed and for continuing to support me even after they realised their mistake

Yumiko Ishikawa-Brush

For her friendship and advice

James, Pat and the Monaco lab

For their coffee- and conversation-making abilities. Without which the madness would have taken over

Carol Stott, Melanie Merricks, Gilly Baird, Leila Jannoun and Vicky Slonims

For patiently answering my stupid questions and sending me information over the last few months

All the members of the SLI Consortium

For their efforts in the field, collecting and organising the families

And to all the families who agreed to participate in this project

Without whom none of this would be possible.

ABBREVIATIONS USED IN THIS THESIS

A	Adenine
ABI	Applied Biosystems
AD	Alzheimer's Disease
ADHD	Attention Deficit Hyperactivity Disorder
BAC	Bacterial Artificial Chromosome
BSA	Bovine Serum Albumin
BuDr	5-Bromodeoxyuridine
C	Cytosine
CCD	Charge-Coupled Device
CD	Crohn's Disease
cDNA	Complementary DNA
CELF-R	Clinical Evaluation of Language Fundamentals - Revised
CHLC	Co-operative Human Linkage Center
CHORI	Children's Hospital Oakland Research Institute
CI	Confidence Interval
CLASP	Cambridge Language and Speech Project
CLSA	Collaborative Linkage Study of Autism
CO_2	Carbon Dioxide
CP	Cortical Plate
CpG	CG Dinucleotide
dATP	Deoxyadenine Triphosphate
dCTP	Deoxycytosine Triphosphate
dGTP	Deoxyguanine Triphosphate
dNTP	Deoxynucleotide Triphosphate
dTTP	Deoxythymine Triphosphate
dUTP	Deoxyuracil Triphosphate
ddATP	Dideoxyadenine Triphosphate
ddCTP	Dideoxycytosine Triphosphate
ddGTP	Dideoxyguanine Triphosphate
ddNTP	Dideoxynucleotide Triphosphate
ddTTP	Dideoxythymine Triphosphate
DAPI	4',6-Diamidino-2-Phenylindole
dbSNP	Single Nucleotide Polymorphism Database
DF	DeFries-Fulker
DHPLC	Denaturing High-Performance Liquid Chromatography
DIG	Digoxigenin
DMD	Duchenne Muscular Dystrophy
DNA	Deoxyribonucleic Acid
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
DZ	Dizygotic
E. coli	*Escherichia coli*
EBV	Epstein-Barr Virus
ECACC	European Collection of Cell Cultures
EDTA	Ethylene-diamine-tetra-acetetae
EH	Eps15 Homology Domain
ELS	Expressive Language Score
ELStrans	Expressive Language Score (transformed)
EST	Expressed Sequence Tag
FCS	Foetal Calf Serum
FIQ	Full IQ
FISH	Fluorescence in-situ Hybridisation
FITC	FluoresceinisoThioCyanate
Fox	Forkhead Box Domain
G	Guanine

GAS	Genetic Analysis System
GDB	Geriome Database
GH2.0	Genehunter 2.0
HE	Haseman-Elston
IBD	Identical By Descent
IBD	Inflammatory Bowel Diseases
IBS	Identical By State
ICD-10	International Classification of Diseases
IMGSAC	International Molecular Genetic Study of Autism Consortium
IQ	Intelligence Quotient
IZ	Intermediate Zone
K	Keto (Guanine or Thymine)
KCl	Potassium Chloride
LB	Luria Broth
LD	Linkage Disequilibrium
LMS2-HD5	Linkage Marker Set – High Density 5cM
LMS2-MD10	Linkage Marker Set – Medium Density 10cM
LOD	Logarithm Of Odds
LOD	Logarithm of Odds
M	Amino (Adenine or Cytosine)
Mg^{2+}	Magnesium ions
$MgCl_2$	Magnesium Chloride
MLS	Maximum LOD Score
MZ	Monozygotic
N	Nucleotide base (Guanine, Thymine, Cytosine or Adenine)
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
NF1	Neurofibromatosis Type 1
NWR	Non-Word Repetition
NWRtrans	Non-Word Repetition (transformed)
OD	Optical Density
OD	Orthographic Decoding
OME	Otitis Media with Effusion
p	Probability
PA	Phonological Awareness
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PD	Phonological Decoding
PDD	Pervasive Developmental Disorder
PDDNOS	Pervasive Developmental Disorder Not Otherwise Specified
PEP	Preamplification Extension Protocol
PIC	Polymorphic Information Content
PIQ	Performance (non-verbal) IQ
PIQtrans	Performance (non-verbal) IQ (transformed)
PolI	DNA polymerasel
PTB	Phosphotyrosine Binding Domain
QTDT	Quantitative Transmissiori Disequilibrium Test
QTL	Quantitative Trait Locus
R	Purine (Adenine or Guanine)
RA	Reading Ability
RLS	Receptive Language Score
RLStrans	Receptive Language Score (transformed)
Rpm	Revolutions per minute
S	Strong (Guanine or Cytosine)
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate

SE	Standard Error
SIBMED	Sib-pair Mutation and Error Detection
SLI	Specific Language Impairment
SNP	Single Nucleotide Polymorphism
SSC	Sodium chloride/Sodium Citrate
T	Thymine
TBE	Tris-Borate EDTA
TDT	Transmission Disequilibrium Test
TE	Tris/EDTA
TEAA	Triethyl Ammonium Acetate
TLS	Total Language Score
TOAL	Test of Adolescent Language
TOLDI	Test of Language Development - Intermediate
TOLDP	Test of Language Development - Primary
TROG	Test for the Reception of Grammar
TXRD	Texas Red
UC	Ulcerative Colitis
UCSC	University of California, Santa Cruz
UTR	Untranslated Region
UV	UltraViolet
VC	Variance Components
VIQ	Verbal IQ
VZ	Ventricular Zone
W	Weak (Adenine or Thymine)
WAIS	Wechsler Intelligence Scale for Adults
WIPPSI	Wechsler Pre-school and Primary Scale of Intelligence
WISC-III	Wechsler Intelligence Scale for Children
WS	Waardenburg syndrome
Y	Pyrimidine (Cytosine or Thymine)

SYMBOLS USED IN THIS THESIS

λs	Sibling risk ratio
μ	Mean
σ	Standard Deviation
χ^2	Chi-squared
θ	Recombination Fraction
D	Phenotype Difference
π	Proportion of Alleles Shared IBD.
β	The Slope of a Line of Regression
V_G	Genotypic Variance
V_P	Phenotypic Variance
V_e	Environmental Variance
V_a	Major gene Variance
V_g	Polygenic Variance
v_I	IBD sharing
$\sigma^2 a$	Major Gene Variance Component
$\sigma^2 g$	Polygenic Variance Component
$\sigma^2 e$	Environmental Variance Component
$\sigma^2 I$	Trait Variance of the Sibs who Share i Alleles IBD

MEASUREMENTS USED IN THIS THESIS

Symbol	Measurement	Equivalent to
g	Gram	
mg	Milligram	10^{-3}g
µg	Microgram	10^{-6}g
ng	Nanogram	10^{-9}g
l	Litre	
ml	Millilitre	10^{-3}l
µl	Microlitre	10^{-6}l
M	Molar	1mol/litre
mM	Milli-Molar	10^{-3}mol/litre
µM	Micro-Molar	10^{-6}mol/litre
mol	Mole	
mmol	Millimole	10^{-3}mole
µmol	Micromole	10^{-6}mole
nmol	Nanomole	10^{-9}mole
pmol	Picomole	10^{-12}mole
bp	Base pairs	
Kbp	Kilobase pairs	10^{3}bp
Mb	Megabases	10^{6}bp
°c	Degrees centigrade	
mV	MilliVolts	
M	Morgans	
cM	centiMorgans	10^{-2}M

GENES AND LOCI REFERRED TO IN THIS THESIS

Gene / Locus	Title	Chromosomal Location	Alternative Symbols	OMIM ID
5-HT7	5-hydroxytryptamine (serotonin) receptor 7	10q21-q24		182137
5-HTT	Solute carrier family 6 member 4	17q11.1-q12	*SLC6A4;SERT*	182138
ADAM33	A Disintegrin and Metalloproteinase Domain 33	20p13		607114
APOC2	Apolipoprotein C-II	19q13.2		207750
APOE	Apolipoprotein E	19q13.2		107741
AUTS1	Autism susceptibility 1 locus	7q31		209850
CARD2	Caspase recruitment domain-containing protein-15	16q12	*IBD1;NOD2*	605956
CHRNA7	Neuronal nicotinic acetylcholine receptor α7 subunit	15q14		118511
COMT	Catechol-O-methyltransferase	22q11.21		116790
COPG2	Coatomer protein complex subunit gamma 2	7q32		604355
CORTBP2	Cortactin-binding protein-2	7q		
CPA1	Carboxypeptidase A1 (pancreatic)	7q32		114850
CPA5	Carboxypeptidase A5	7q32		
CUL3	Cullin 3	2q37.1		603136
DAT1	Solute carrier family member 3	5p15.3	*SLC6A3*	126455
DDC	Dopa decarboxylase	7p11	*AADC*	107930
DRD3	Dopamine receptor D3	3q13.3		126451
DRD4	Dopamine receptor D4	11p15.5	*D4DR*	126452
DRD5	Dopamine receptor D5	4p16.1-p15.3	*DRD1B*	126453
DYX1	Dyslexia susceptibility locus 1	15q21		127700
DYX2	Dyslexia susceptibility locus 2	6p		600202
DYX3	Dyslexia susceptibility locus 3	2p16-p15		604254
DYXQTL18	Dyslexia Quantitative Trait Locus on Chromosome 18	18		606616
DβH	Dopamine beta hydroxylase	9q34		22360
FACL3	Fatty acid coenzyme A ligase long chain 3	2q34-q35	*ACS3*	602371
FLJ22724	Hypothetical protein FLJ22724	2q35		
FOXC1	Forkhead box C1	6p25	*FKHL7;FREAC3*	601090

Gene / Locus	Title	Chromosomal Location	Alternative Symbols	OMIM ID
FOXE1	Forkhead box E1	9q22	*FKHL15;TTF2;TITF2*	602617
FOXP2	Forkhead box P2	7q31	*TNRC10;CAGH44*	605317
FOXP2-S	Truncated form of *FOXP2*			
Fy	Drosophila *fuzzy* gene	29C1		
$GABA_A5$	γ-aminobutyric acid receptor, alpha 5	15q11.2-q12		137142
$GABA_B3$	γ-aminobutyric acid receptor, beta 3	15q11.2-q12		137192
$GABA_G3$	γ-aminobutyric acid receptor, gamma 3	15q11.2-q12		600233
HTR2A	5-hydroxytryptamine (serotonin) receptor 2A	13q14-q21		182135
IBD1	Inflammatory Bowel Disease 1	16q12	*NOD2;CARD15*	605956
IBD2	Inflammatory Bowel Disease 2	12p13.2-q24.1		601458
IL-1Ra	Interleukin-1 receptor antagonist	2q14.2	*IL1RN*	147679
KLK	Kallikrein 9	19q13.3-q13.4	*KLK9;KLKL3*	605504
MAOA	Monoamine oxidase A	Xp11.4-p11.3		309850
MAOB	Monoamine oxidase B	Xp11.4-p11.3	*MAO*	309860
MEST	Mesoderm specific transcript homologue	7q32	*PEG1*	601029
Nbl	Mouse numb-like		*Numbl*	
NET1	Solute carrier family 6 member 2	16q12.2	*NAT1;NET;SLC6A2*	163970
NF1	Neurofibromin 1	17q11.2		162200
NOTCH1	Notch homologue 1, translocation-associated (Drosophila)	9q34.3	*TAN1*	190198
NUMB	Numb homologue (Drosophila)	14q24.3		603728
NUMB-R	Numb homologue (Drosophila)-like	19q13.13-q13.2	*NBL;NUMBL;TNRC23*	604018
PAI-1	Plasminogen activator inhibitor-1	7q21.3-q22	*SERPINE1;PLANH1*	173360
PAX3	Paired box gene 3	2q35	WS1	193500
PheHB	Phenylalanine tRNA synthetase-like beta subunit	2q36.3	*FARSLB1;FRSB*	
RAY1	Suppressor of tumourugenicity 7	7q31.1-q31.3	*TSG7;ST7;FAM4A1*	600833
RRM1	Ribonucleotide reductase M1 polypeptide	2q		
SECG2	Secretogranin II	2q		
SNAP-25	Synaptosomal associated protein of 25kDa	20p12-p11.2	*SNAP*	600322
SPCH1	Forkhead box P2	7q31	*FOXP2;TNRC10;CAGH44*	605317

Gene / Locus	Title	Chromosomal Location	Alternative Symbols	OMIM ID
TH	Tyrosine hydroxylase	11p15.5		191290
UBE3A	Ubiquitin protein ligase E3A	15q11-q13	*E6AP*	601623
WNT2	Wingless-type MMTV integration site family member 2	7q31	*INT1L1;IRP*	147870
ZNF1A2	Zinc finger protein, subfamily 1A, 2	2q35	*helios;ZNFN1A2*	606234

CONTENTS

PREFACE ..... **i-xx**

Acknowledgements ..... **i**

Abbreviations used in this thesis ..... **ii-iv**

Symbols used in this thesis ..... **v**

Measurements used in this thesis ..... **vi**

Genes and loci referred to in this thesis ..... **vii-ix**

Contents ..... **x-xv**

List of appendices ..... **xvi**

List of figures ..... **xvii-xviii**

List of tables ..... **xix-xx**

CHAPTER 1 – INTRODUCTION **1-45**

1.1 Introduction **2-8**

1.1i SLI - What's in a name? 2-3

1.1ii The SLI phenotype 3-6

1.1iii The prevalence of SLI 6-8

1.2 Diagnosis and research considerations **9-13**

1.2i Non verbal Intelligence Quotient (IQ) 9

1.2ii Hearing 10

1.2iii Oral function 10

1.2iv Neurological dysfunction 10-11

1.2v Autism and social competence 11

1.2vi Current diagnostic criteria 12-13

1.3 Is language genetic? **13-18**

1.3i The imitation and reinforcement model of language acquisition 13-14

1.3ii Innate language systems 15-16

1.3iii Universal grammar 16-17

1.3iv Empiricists and connectionists 17-18

1.4 Is SLI genetic? **19-31**

1.4i Family studies 19-23

1.4ii Twin studies 24-27

1.4iii Adoption studies 28-29

1.4iv Heritability of language measures 29-31

1.5 Causal theories of SLI **31**

1.6 Innate grammar impairments **31-37**
1.6i Gopnik - the missing feature deficit hypothesis 32-34
1.6ii Rice - the extended optional infinitive account 34
1.6iii Clahsen - the agreement marking deficit hypothesis 34-35
1.6iv Van der Lely - representational deficit for dependent relationships 35-36
1.6v Leonard - auditory perceptual deficit and the surface hypothesis 36-37
1.7 Limitations in processing capacities **37-45**
1.7i Short-term memory deficits in SLI 37-41
1.7ii Auditory perception impairments 41-45
1.8 Aims of this project **45**

CHAPTER 2 - SUBJECTS & METHODS **46-103**
SUBJECTS **47-63**
2.1 The genome screen cohort **47-53**
2.1i Ascertainment of cohort 47-49
2.1ii The Guys Hospital sample 49-50
2.1iii The Cambridge Language and Speech Project (CLASP) sample 50-53
2.2 The genome screen phenotypes **53-63**
2.2i The Clinical Evaluation of Language Fundamentals - Revised (CELF-R) 53-54
2.2ii Non-word repetition 54-55
2.2iii Intelligence 55-56
2.2iv Descriptive statistics of phenotypes 56-61
2.2v Data transformation 61-63
METHODS **64-103**
2.3 DNA collection and preparation **64-69**
2.3i DNA extraction from whole blood 64-66
2.3ii DNA extraction from buccal swabs 66-68
2.3iii Preamplification Extension Protocol (PEP) 68-69
2.4 The Polymerase Chain Reaction (PCR) **69-72**
2.4i The Polymerase Chain Reaction 69-70
2.4ii PCR conditions and programs 70-72
2.4iii PCR of the genome screen markers 72
2.5 Agarose and polyacrylamide gels **72-76**
2.5i Agarose gels 72-74
2.5ii Polyacrylamide gels 74-76

2.6 Genotyping **76-79**
2.6i Data extraction 76-77
2.6i Data checking 78
2.6ii Data formats and information content maps 78-79
2.7 Linkage Analysis **79-81**
2.7i Haseman Elston and Variance Components analysis 79-80
2.7ii X Chromosome analysis 80
2.7iii Simulations 80-81
2.7iv Fine mapping 81
2.8 Denaturing High Performance Liquid Chromatography (DHPLC) **81-85**
2.8i Using DHPLC to detect polymorphisms 81-83
2.8ii DHPLC analysis of candidate genes 84-85
2.9 Sequencing **85-89**
2.9i Dye terminator sequencing 85-86
2.9ii PCR of template DNA and purification of PCR products 87
2.9iii Sequencing reactions 87-88
2.9iv Ethanol precipitation of sequencing products 88-89
2.10 Single Nucleotide Polymorphisms (SNPs) **89-92**
2.10i SNPs 89
2.10ii SNP typing 90-91
2.10iii Quantitative Transmission Disequilibrium Test (QTDT) 91-92
2.11 Fluorescence In-Situ Hybridisation (FISH) **92-103**
2.11i Epstein-Barr virus transformation of cultured lymphocytes 92
2.11ii Cell culture 92
2.11iii Cryopreservation of cells 92-93
2.11iv Harvesting lymphoblastoid cells 93-95
2.11v Slide creation 95
2.11vi Slide preparation 95
2.11vii Bacterial Artificial Chromosomes as FISH probes 95-96
2.11viii Preparation of FISH probes 96-98
2.11ix Nick translation 98-99
2.11x Probe preparation 100
2.11xi Probe detection 100-102
2.11xii Antibody layers 102-103

CHAPTER 3 – THE SLI GENOME SCREEN **104-147**

3.1 Gene mapping in complex disorders **105-113**

3.1i Functional and positional cloning 105-106

3.1ii Recombination 106-107

3.1iii Recombination fractions 107

3.1iv Linkage analysis 107-108

3.1v Microsatellite markers and Identity By Descent (IBD) 108-110

3.1vi Parametric and non parametric linkage 110

3.1vii Haseman Elston analysis 110-111

3.1viii Variance Components analysis 111-112

3.1ix Simulations 112-113

3.2 Genome screens **113-132**

3.2i Overview of genome screens for complex disorders 113-114

3.2ii Genome screens of disorders related to SLI 114-115

3.2iii Autism 115-121

3.2iv Dyslexia 121-124

3.2v Attention-Deficit Hyperactivity Disorder (ADHD) 124-129

3.2vi Molecular genetic studies of SLI 130-132

3.3 Genome screen results **132-144**

3.3i Genome screen results 132-141

3.3ii Fine mapping of chromosome 19 142-143

3.4 Discussion **144-147**

CHAPTER 4 – CANDIDATE GENE ANALYSES **148-183**

4.1 Introduction **149**

4.2 Association **149-152**

4.2i Association analysis 149-150

4.2ii Approaches to association analysis 150-151

4.2iii Linkage disequilibrium blocks and association studies 151-152

4.2iv Choice of markers for association studies 152

4.3 Numblike (*NUMB-R*) **153-158**

4.3i Human neurogenesis 153-155

4.3ii The *NUMB-R* gene 155-156

4.3iii *NUMB-R* gene expression 156-157

4.3iv *NUMB-R* associated knockouts 157

4.3iii *NUMB-R* analyses 157-158

4.4 Forkhead-bOX domain P2 (*FOXP2*) 158-169

4.4i The KE family 158-161

4.4ii Brain imaging in the KE family 161-162

4.4iii *SPCH1* and chromosome 7q 162-164

4.4iv The *FOXP2* gene 164-166

4.4v *FOXP2* in SLI 167-169

4.5 Candidate gene analysis results 170-177

4.5i *NUMB-R* polyglutamine typing 170

4.5ii DHPLC of *NUMB-R* 170-172

4.5iii DHPLC of *FOXP2* 172-175

4.5iv Association analyses of *NUMB-R* 176

4.5v Association analyses of *FOXP2* 176-177

4.6 Discussion 177-183

4.6i *NUMB-R* 177-178

4.6ii *FOXP2* 178-183

CHAPTER 5 – FISH MAPPING OF A TRANSLOCATION IN A CHILD WITH LANGUAGE IMPAIRMENT 184-203

5.1 Introduction 185

5.2 Chromosome rearrangements and translocations 185-188

5.2i Chromosome rearrangements 185-187

5.2ii Translocations 187-188

5.3 The G family and chromosome 2q 188-191

5.3i Language impairment and a chromosome 2;11 translocation in the G family 188-189

5.3ii Linkage to chromosome 2q in the SLIC genome screen 190

5.3iii Chromosome 2q35 191

5.4 Chromosome 2 translocations 191-196

5.4i Translocations associated with SLI 191

5.4ii Autism and chromosome 2q 193-194

5.4iii Other translocations of chromosome 2q35 194-195

5.5 Characterisation of the translocation within the G family 196

5.6 FISH results **196-201**
5.6i FISH mapping of the G family translocation 196-198
5.6ii Segregation of the 2;11 translocation within the G family 198
5.6iii Sequence analysis of clone RP11-105N14 198-200
5.6iv Association analyses 201
5.7 Discussion **201-203**
CHAPTER 6 - SUMMARY, APPRAISAL AND FUTURE DIRECTIONS **204-211**

ADDENDUM **212-251**
Attributions **212**
Websites **213-214**
Bibliography **215-251**

LIST OF APPENDICES

Appendix A Members of the SLI Consortium (SLIC)

Appendix B Sub-tests of the Clinical Evaluation of Language Fundamentals (CELF-R)

Appendix C The Non-word Repetition Test (20 item) Guys Version

Appendix D The Non-word Repetition Test (28 item) Cambridge Version

Appendix E Sample Deposit Form for SLI Study

Appendix F Markers Excluded from the ABI PRISM LMS2-MD10 Panels

Appendix G Markers Used for Filling in Genome Screen Gaps

Appendix H Fine Mapping Panels for Chromosome 19q

Appendix I Reagents and Suppliers

Appendix J Genome Screen Results by Individual Chromosomes

Appendix K Sequence of *NUMB-R* with SNPs

Appendix L Coding Sequence of the *NUMB-R* Gene

Appendix M *NUMB-R* Primers

Appendix N Sequence of *FOXP2* with SNPs and Microsatellites

Appendix O Coding Sequence of the *FOXP2* Gene

Appendix P *FOXP2* Primers

Appendix Q Restriction Enzymes Used for the Typing of *NUMB-R* and *FOXP2* SNPs

Appendix R Identified Genes Around the Chromosome 2q35 Band

Appendix S SNPs Used for Association Analyses of RP11-105N14

Appendix T Primers and Restriction Enzymes Used for the Analysis of RP11-105N14

Appendix U Publications Arising from this Work

LIST OF TABLES

		PAGE
Table 1.1	Studies of the prevalence of SLI	7-8
Table 1.2	Family studies of SLI	21-23
Table 1.3	Twin studies of SLI	27
Table 2.1	Ascertainment criteria for the SLI genome screen	49
Table 2.2	Numbers of families and sib-pairs referred by Guys Hospital and Cambridge	52
Table 2.3	Numbers of families and sib-pairs excluded from the Guys Hospital and Cambridge samples	52
Table 2.4	The CELF-R sub-tests	54
Table 2.5	Age range and sex distributions of the total genome screen sample and in the proband and sibling groupings	56
Table 2.6	Descriptive statistics for each genome screen phenotype within the total genome screen set and the constituent Guy's Hospital and Cambridge groupings	58
Table 2.7	Descriptive statistics for each genome screen phenotype for the total genome sample split by proband and co-sibs	58
Table 2.8	Correlations and familialities of phenotypes within the total genome screen set and the constituent Guy's Hospital and Cambridge groupings	63
Table 2.9	Proband-Sibling Correlations of phenotypes within the total genome screen set and the constituent Guy's Hospital and Cambridge groupings	63
Table 2.10	SNP validation	90-91
Table 2.11	Antibody layers for FISH	102
Table 2.12	Antibody layers for dual layer FISH	103
Table 3.1	The overlaps between genome screens of complex disorders	114
Table 3.2	Genome screens of autism	116
Table 3.3	Evidence for a chromosome 7q locus in autism	117
Table 3.4	Candidate gene studies for autism at the *AUTS1* locus	119-120
Table 3.5	Major loci detected by dyslexia genome screen	123
Table 3.6	Candidate gene studies of ADHD	125-126

		PAGE
Table 3.7	Genomic regions yielding LOD Scores > 1.5 in ADHD genome screen	127
Table 3.8	Loci excluded by the ADHD screen	129
Table 3.9	LOD scores >1.0	137
Table 3.10	Singlepoint LOD scores for chromosome 16	140
Table 3.11	Singlepoint LOD scores for chromosome 19	140
Table 3.12	Singlepoint LOD scores for chromosome 19q fine mapping	142
Table 4.1	Size of polyglutamine repeat tracts within the genome screen sample	170
Table 4.2	*NUMB-R* sequence variants detected by DHPLC	170
Table 4.3	*FOXP2* sequence variants detected by DHPLC	172
Table 4.4	Association analyses of *NUMB-R* SNPs	176
Table 4.5	Association analyses of *FOXP2* SNPs and microsatellites	177
Table 5.1	Evidence for an autism susceptibility locus on chromosome 2q	193
Table 5.2	Other reported translocations involving chromosome 2q35	195
Table 5.3	Association analyses of RP11-105N14 SNPs	201

LIST OF FIGURES

		PAGE
Figure 1.1	Short-term memory	38
Figure 1.2	The phonological loop	39
Figure 1.3	The auditory perception test	42
Figure 2.1	The Cambridge Language and Speech Project (CLASP) - Study outline	51
Figure 2.2	CELF-R Expressive Language Scores (ELS) and Receptive Language Scores (RLS) in Guys, Cambridge, proband and sibling cohorts	59
Figure 2.3	Non-Word Repetition (NWR) and Performance IQ (PIQ) in Guys, Cambridge, proband and sibling cohorts	60
Figure 2.4	The Polymerase Chain Reaction (PCR)	70
Figure 2.5	1Kb ladder and λ*Hind*III standards	74
Figure 2.6	Fluorescence wavelength and excitation efficiency of dyes	75
Figure 2.7	Genotyping gel	75
Figure 2.8	Allele calling within Genotyper	77
Figure 2.9	DHPLC polymorphism identification	82
Figure 2.10	Denaturing High-Performance Liquid Chromatography (DHPLC)	83
Figure 2.11	Dye terminator sequencing	86
Figure 2.12	Sequencing plots	89
Figure 2.13	The pBACe3.6 vector	96
Figure 2.14	Antibody layers for biotin and dioxigenin probes	101
Figure 3.1	Identity By Descent (IBD)	109
Figure 3.2	LOD significance distributions for each measure used in the genome screen	134
Figure 3.3	Genome-Wide multipoint linkage to SLI (organised by analysis type)	135
Figure 3.4	Genome-wide multipoint linkage to SLI (organised by phenotype)	136
Figure 3.5	Multipoint linkages to chromosomes 16 and 19	139
Figure 3.6	HE and VC multipoint linkage to PIQ on chromosomes 16 and 19	141
Figure 3.7	Linkage to chromosomes 16 and 19 in the Guys and CLASP samples	141
Figure 3.8	Multipoint LOD scores for chromosome 19q fine mapping	143
Figure 4.1	The position of *NUMB-R* in relation to the chromosome 19 peak	153
Figure 4.2	The development of the neocortex	154

		PAGE
Figure 4.3	*NUMB, NOTCH1 and NUMB-R* expression in the developing neocortex	155
Figure 4.4	*NUMB-R* gene structure	156
Figure 4.5	SNPs analysed for *NUMB-R* association	158
Figure 4.6	The KE pedigree	159
Figure 4.7	Critical region of *SPCH1*	163
Figure 4.8	The *FOXP2* coding sequence	166
Figure 4.9	SNPs and microsatellites analysed for *FOXP2* association	169
Figure 4.10	The exon 4 polymorphism in family 27	171
Figure 4.11	The exon 6 polymorphism in family 43	173
Figure 4.12	Possible positions of the CAGCAG insertion in *FOXP2* exon 6	174
Figure 4.13	SNPs detected in *NUMB-R* and *FOXP2*	175
Figure 4.14	Sequence alignment of the *NUMB* and *NUMB-R* PTB regions across human, mouse and drosophila	178
Figure 4.15	The revised genomic structure of *FOXP2*	182
Figure 5.1	Mechanisms of chromosome rearrangements	187
Figure 5.2	Segregation of the translocation and language impairment in the G family	189
Figure 5.3	Linkage to chromosome 2 within the SLIC genome screen	190
Figure 5.4	Gene density and identified transcripts around chromosome 2q35	192
Figure 5.5	Rounds of FISH performed on chromosome 2q35	197
Figure 5.6	Identification of the breakpoint within the G family	198
Figure 5.7	RP11-105N14 and associated clones and transcripts	200

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

At around the age of one year, the average child speaks their first word. By the age of six they have a vocabulary of approximately 14,000 words and, for the most part, have mastered their native tongue. In order to achieve this rate of vocabulary growth alone, it is estimated that the average child must learn nine new words every day for the first six years of their life and this can almost be considered as the easy part of the task. In addition to the rote learning of vocabulary, children must also derive and apply the abstract grammatical rules that underlie their native tongue. But yet, this seemingly impossible undertaking is routinely accomplished by children all around the world, with apparently little effort and minimal specific instruction from the caregiver. The way in which this is achieved is still largely a mystery, despite years of work and research into the area.

For some children, however, language acquisition is not so straightforward and the route to linguistic superiority is a hard fought path. Difficulties with language acquisition may arise for a number of reasons; children may have hearing problems, general learning difficulties or may suffer from neurological deficits that impair their grasp of the human language (e.g. cerebral palsy, down's syndrome). However, for a significant proportion of children none of the above clauses apply. These children appear to be developing normally in non-verbal areas but yet, for reasons that remain unknown, show specific problems with the acquisition of language. These children are said to have Specific Language Impairment or SLI.

1.1i - SLI – What's in a Name?

Gall was perhaps the first to describe a group of children with language limitations that could not be attributed to any recognised medical condition:

> *'There are children....who do not speak to the same degree as other children although they understand well or are far from being idiotic. In these cases the trouble lies not in the vocal organs, ... and still less in the apathetic state of the subject'*
>
> ***Gall (1822)***

Following Gall's account many such children were described in the literature. These reports tended to refer to the disorder as 'congenital aphasia' (Väisse 1866) or 'hearing mutism'

(Coën 1886) and concentrated on children with severe articulation deficits. Because of the distinct lack of neurological deficits in these children, it was hypothesised that their poor language performance may be caused by limitations in memory and attention (Treitel 1893).

By the early 1900's the condition was accepted as a valid diagnostic category and the terms 'congenital aphasia', 'infantile aphasia' (Gesell & Amatruda 1947) and 'developmental aphasia' (Benton 1964) were widely applied. The condition was expanded to include those children with more subtle grammatical and comprehension deficits and was recognised as being consistent with a neurodevelopmental delay (Ewing 1930). The first twin study by Ley (1929) provided evidence for a genetic aetiology behind the disorder.

Liebmann (1898) was one of the first to discuss subtypes of language impairments. He suggested a classification system that categorised children according to whether their deficit was motoric, expressive or comprehensive in nature. However this work was not applied until much later in the mid-twentieth century when the terms 'expressive developmental aphasia' and 'receptive-expressive developmental aphasia' became more commonly used (Lea 1968, Sato & Dreifuss 1973, Duffy 1975).

The later end of the 20th century saw a shift away from the use of the words 'aphasia' and 'dysphasia' — which imply a neurological cause — to more neutral phrases such as 'infantile speech' (Menyuk 1964), 'delayed speech' (Lovell et al 1968), 'deviant language' (Leonard 1972), 'developmental language disorder' (Aram & Nation 1975), 'developmental language impairment' (Wolfus et al 1980), 'specific language deficit' (Stark & Tallal 1981) and 'specific language impairment' (Tallal et al 1989). For ease of reading, throughout this thesis I will refer to the disorder as SLI, even though the original author may have used some other term.

1.1ii - The SLI Phenotype

Most children with SLI are first noticed to be 'slow to talk' but, for the experienced linguist, signs of impairment may be apparent long before the age of one year. At around the age of 6-10 months most children begin to 'babble'. This characteristic mode of preverbal communication involves the combination of consonant and vowel sounds (e.g. mama, dada) and provides the infant with a chance to practice the production of speech sounds. It has

been shown that children who go on to develop SLI babble less than their normally developing age-peers (Stoel-Gammon 1998).

By the age of around nine months most infants begin to indicate what they want by pointing. This is an important stage in communicative development and indicates that the child is able to recognise the focus of their caregiver's attention and manipulate this to achieve a goal. Joint attention episodes have been shown to be important for later language development (Tomasello & Farrar 1986, Tomasello et al 1986) and children with social-emotional disturbances are found to partake in these exchanges less often than other infants their own age (Cohen 2001). Although more typically associated with the autism spectrum (Loveland & Landry 1986, McArthur & Adamson 1996), mild delays in the use of joint attention and gesture may be indicative of less severe phenotypes involving language and communication impairments (Cohen 2001).

As with all children, the language profile of children with SLI varies considerably both between individuals and across different time points. Some may have problems in articulating speech sounds, whereas others experience difficulties in both producing and understanding language. Others still, show impairments only in the expressive domain. However, it is very rare to find a child who has poor receptive skills in the absence of any expressive deficit (Bishop et al 1995). Although this may be partly due to the insensitivity of tests to detect receptive impairments, there is a growing body of evidence to suggest that receptive SLI may represent a distinct disorder separate from other forms of language deficits. Unlike the more common types of SLI, receptive language impairments show only low levels of heritability and are found at equal frequencies in both males and females (Bishop et al 1995). Accordingly, current research broadly classifies SLI into three subtypes; articulation disorder (previously known as phonological disorder), expressive language disorder and mixed expressive and receptive language disorder. However, the validity of this subtyping remains questionable and significant overlaps are still found to exist between these classes.

In general, the language of a child with SLI is comparable to that of a younger child. They tend to have limited vocabularies and use immature speech sounds and grammar. As a consequence they often rely upon short, simple sentences and have difficulties following long

or complex instructions. Language-impaired children tend to struggle with tasks which involve abstract reasoning or require a lot of information to be retained over a short period of time (Bishop1997a, Bishop 2001a).

Although SLI is labelled as a specific language problem, many non-verbal deficits have been described which are characteristically associated with the SLI phenotype. Children with SLI tend to have poor memory skills, below average motor skills and limited attention spans (Johnston et al 1981, Bishop & Edmundson 1987, Powell & Bishop 1992, Hill 2001, Bishop 2002). This has lead to the theory that SLI might represent a global learning deficit. However, it remains unclear whether these non-verbal deficits occur in addition to, or as a consequence of, the language impairments characteristic of the SLI phenotype.

The memory limitations associated with SLI play an important role in later educational life. Children with SLI often find it hard to follow what is going on in class and as a consequence suffer from attentional problems. In particular, a language-impaired child is at a distinct disadvantage when learning to read. Reading places high demands on the memory and requires the association of new information with that which has gone before. It is not surprising, therefore, that most studies indicate a high prevalence of dyslexia (Silva et al 1983, Silva et al 1987, Snowling et al 2000, Bishop 2001b) and Attention Deficit Hyperactivity Disorder (ADHD) in children with SLI (Tirosh & Cohen 1998, Cohen et al 2000, Kovac et al 2001). Cohen et al (1993) studied a group of children who had been referred to a specialist clinic for a social or emotional disturbance alone. They found that 34% of these 4-12 year olds showed signs of language impairments but had not previously been diagnosed as such. In his twin study of SLI Tomblin (1997) found that approximately 66% of the language-impaired twins also met the diagnostic criteria for dyslexia.

Although children with SLI do learn to compensate for many of their language problems, the emotional disruptions caused by their early language failings can be extensive and last well into adulthood (Beitchman et al 1994, Beitchman et al 1996a, Johnson et al 1999). Because language-impaired children can feel inadequate in their verbal communication, they are less likely to engage in peer interaction and in turn, other children are less likely to play with them (Rice et al 1991, Gertner et al 1994). These early inadequacies in peer relations have been

shown to undermine the social competence of the child (Beitchman et al 1996b) and have been linked to social-emotional disturbances and low self-esteem in later life (Parker & Asher 1987). It has been shown that children with SLI remain in lower social, economical and educational classes than their peers (Beitchman et al 1996c, Snowling et al 2001) and experience an increased risk of developing psychiatric disorders (e.g. anxiety disorder) (Cantwell & Baker 1987, Beitchman et al 1990, Beitchman et al 1996d, Beitchman et al 2001a), substance abuse (Beitchman et al 2001b), and criminality (Bountress & Richards 1979, Myers & Mutch 1992, Stattin & Klackenberg-Larsson 1993). As such, SLI represents a major health issue.

1.1iii The Prevalence of SLI

Although there have been many epidemiological studies of SLI, differences in methodological approaches, diagnostic criteria and categorical thresholds often render direct comparisons between investigations impractical. The majority of mainstream studies estimate the prevalence of SLI amongst English-speaking pre-school children to be between two and seven percent with two to three times as many boys as girls being affected (Law et al 1998).

Sample size	Factors used to define affection	Prevalence	m:f	Exclusionary criteria
Randell, Reynell & Curwen (1974) Population sample of three-year-olds in London area				
160	>2SD below expected on any one of three language tests (verbal comprehension, expressive language or articulation)	5.6%	5:4	Families in which both parents had a foreign mother tongue were excluded
	>2SD below expected on articulation test	3.1%	3:2	
	>2SD below expected on either verbal comprehension or expressive language test	2.5%	1:1	
Stevenson & Richman (1976) Population sample of three-year-olds in London area				
705	Expressive language age ≤2/3 of that expected for chronological age	2.3%	~2:1	No specific criteria given. Excluded two deaf children & one with Downs syndrome.
	Expressive language age > 6 months below mean for chronological age	3.1%	~2:1	
	Expressive language age ≤2/3 non-verbal mental age	1.4%	~4:1	
Silva (1980) Population sample of three-year-olds from New Zealand				
937	>2SD below expected on verbal comprehension or expressive language tests	8.4%	~2:1	None given
	>2SD below expected on verbal comprehension tests	2.9%	NA	
	>2SD below expected on expressive language tests	2.5%	NA	
	>2SD below expected on verbal comprehension and expressive language tests	2.9%	NA	
Silva, McGee & Williams (1983) Longitudinal study following a population sample from New Zealand between the ages of three and seven				
1027 (age 3)	**Specific Comprehension Delay** - Score at or below the fifth percentile only on the comprehension language scale	2.6%	~2:1*	None given
	Specific Expression Delay - Score at or below the fifth percentile only on the expressive language scale	2.3%	~3:1*	
	General Language Delay - Score at or below the fifth percentile on both the comprehension and expressive language scales	2.6%	~2:1*	
936 (age 5)	Specific comprehension delay	4.0%	NA	
	Specific expression delay	4.3%		
	General language delay	2.1%		
891 (age 7)	Specific comprehension delay	3.6%		
	Specific expression delay	2.8%		
	General language delay	2.0%		
168 (those with delay at any one age)	Transitory language delay (delayed for any one test at only one time point)	74%		
	Moderately stable delay (delayed for any one test at two time points)	16%		
	Stable delay (delayed for any one test across all three time points)	10%		

Table 1.1 – Studies of the Prevalence of SLI

Sample size	Factors used to define affection	Prevalence	m:f	Exclusionary criteria
Beitchman, Nair, Clegg & Patel (1986) Population sample of five-year-olds in the Ottawa-Carleton region				
1655	**Speech only** - >2SD below expected on word articulation subtest **OR** word discrimination subtest **OR** evidence of voice disorder, stuttering or dysarthria	6.4%	1:1	English not first language
	Language only - >2SD below expected on any other subtest from battery (i.e. except articulation and word discrimination) **OR** >1SD below expected on vocabulary test or on auditory memory test **OR** evidence of voice disorder, stuttering or dysarthria	8.0%	1:1	
	Speech and Language - any combination of speech only or language only	4.6%	2:1	
Tomblin, Records, Buckwalter, Zhang, Smith & O'Brien (1997a) Kindergarten (age 5-6) children in Iowa & Illinois				
2009	Language >2SD below expected on any two of five composite language scores taken	1.16%	~3:2	Bilingual or non-English background, mental retardation, autism, neurological problems, blind and hearing impaired children
	Language >1.25SD below expected on any two of five composite language scores taken	7.4%		
	Two composite language scores >1SE below that predicted from non-verbal IQ	13.9%		

Table 1.1 – Studies of the Prevalence of SLI (cont.)

m:f = male:female ratio
NA : Not available
** : These figures were taken from Silva, Williams & McGee (1987)*

1.2 DIAGNOSIS OF SLI AND RESEARCH CONSIDERATIONS

The heterogeneity of the SLI phenotype means that a diagnosis often focuses as much on exclusionary criteria as it does on the quantification of the language deficit itself. Considerations of common exclusionary criteria are discussed below.

1.2i - Non-Verbal Intelligence Quotient (IQ)

Many diagnostic scales require a substantial discrepancy to exist between a child's verbal and non-verbal IQ before a diagnosis of SLI can be made. Whilst the enforcement of these discrepancy scores act to aid the elimination of general IQ effects, they are generally felt to result in an over-restrictive phenotype which is susceptible to compound errors (Stark & Tallal 1981, Vargha-Khadem et al 1995). In a twin study Bishop found that whilst one twin would meet the criteria for discrepancy, the co-twin, who may have scored similar language scores, would fail the discrepancy test. She found that by adopting a less stringent definition, with no account for non-verbal IQ, she received a much more clear-cut pattern of results than under a highly restrictive one (Bishop 1994a). Also relevant to the current study is the finding that discrepancy scores show reduced levels of heritability compared to that of absolute language scores, and hence may not reflect the underlying genetic influences involved in SLI (Bishop et al 1995; Bishop et al 1996; Bishop et al 1999a).

Some researchers have argued that, because a low non-verbal IQ would have no reason to protect children against the development of SLI, IQ scores should be completely disregarded. In support of this view, children who suffer from Williams Syndrome often perform poorly on tests of non-verbal IQ but yet still develop acceptable verbal skills (Jarrold et al 1996, Karmiloff-Smith 1997, Jarrold et al 2001). However, concerns remain about the exact relationship between verbal and non-verbal IQs and many researchers continue to enforce a non-verbal criterion. Thus most modern studies require children to have a deviant language score(s) alongside a non-verbal IQ which is within the 'normal' range (e.g. Tallal et al 1989). This is normally taken to be a non-verbal IQ no greater than 1 or 2 standard deviations below the mean for their chronological age (i.e. above 70 or 85).

1.2ii - Hearing

In 1969 Holm & Kunze suggested that Otitis Media with Effusion (OME), a common childhood infection of the middle ear, which often goes undetected, may underlie the development of SLI. Their theory proposed that OME in young infants might affect the development of neural interconnectivity in the language areas of the brain and thus lead to SLI in later life. Although subsequent studies generally do not support this hypothesis (Bishop & Edmundson 1986, Roberts et al 1991, Roberts et al 1998, Feldman et al 1999, Roberts et al 2000), hearing loss remains an obvious candidate for language impairments (Butler & MacMillan 2001, Psarommatis et al 2001). Prior to a diagnosis of SLI, children must pass a standard hearing screen and report no persistent medical history of OME.

1.2iii - Oral Function

Any children with oral abnormalities which might impede language production (e.g. cleft lip) are excluded from a diagnosis of SLI. Oral function is usually tested by non-verbal lip and tongue movements.

1.2iv - Neurological Dysfunction

The definition of SLI dictates that any language dysfunction is not caused by gross neurological deficits. Thus any children with brain lesions, brain injury, cerebral palsy or seizure disorders are excluded from a diagnosis of SLI. An important condition included in these exclusions is Landau-Kleffner Syndrome (MIM 245570) (Landau & Kleffner 1957). This acquired epileptic disorder develops in the early years of a child's life and is characterised by nocturnal seizures and progressive language loss, especially in the receptive domain. Since the seizures associated with this disorder are nocturnal in nature they often remain undetected and the child is first referred for their language problems. It is thought that these language deficits are caused by the neurological disorder of sound perception and hence it is important that these children are not included in the SLI diagnosis group since their deficits are not strictly limited to verbal material.

Despite the need to exclude children with gross neurological deficits from a diagnosis of SLI, it remains possible that subtle neuro-anatomical abnormalities may underlie the disorder.

Recent brain imaging studies have indicated that an asymmetry in the caudate nucleus may play a role in the aetiology of SLI (Plante et al 1991, Vargha-Khadem et al 1998, Watkins et al 1999, Shafer et al 2001, Watkins et al 2002a). However, many of these studies were carried out on members of the KE family, who display a severe and distinctive language phenotype (see section 4.4i 'The KE family') and others included only small numbers of individuals. Thus further work of a similar manner will be required before any conclusions can be drawn.

1.2v - Autism and Social Competence

An important diagnostic criterion for SLI is that children should not meet the diagnostic criteria for autism.

Autism (MIM 209850) is a neuro-developmental disorder characterised by severe deficits in social and communicative functions, accompanied by repetitive and stereotyped behaviours alongside rigid and obsessive interests (Kanner 1943). Language deficits form a major component of the autism diagnostic criteria and, in general, are found to be more profound than those associated with SLI alone (Lord et al 1994). Autistic children typically make few spontaneous remarks, produce stereotyped utterances, and make only minimal use of gesture (Rapin 1997, Tager-Flusberg 2001). Whilst some autistic children may attain acceptable levels in terms of vocabulary, grammar and phonology, they invariably retain fundamental difficulties with the use of language in a social context (pragmatics) (Rapin 1997, Mawhood et al 2000). A substantial proportion of autistic children completely fail to develop language at all (Rapin 1998, Tager-Flusberg 2001).

Although SLI and autism are generally accepted to be clinically distinct, the boundaries between the two conditions are not always clear and there remains a group of children who, although affected, fail to meet strict diagnostic criteria for either disorder. Several researchers present an argument for the formation of a 'semantic-pragmatic' classification to capture these 'borderline autistic/language-impaired' cases (Rapin & Allen 1983, Bishop & Rosenbloom 1987).

1.2vi - Current Diagnostic Criteria

The current ICD-10 and DSM-IV diagnostic criteria for SLI are outlined below:

ICD-10 Research Diagnostic Criteria for Specific Developmental Disorders of Speech and Language (World Health Organisation 1993)

1) Language skills which fall outside the 2 standard deviation limit for the child's age (although this may not be the case in older subjects)

2) The language delay should not be directly attributable to neurological or speech mechanism abnormalities, sensory impairments, mental retardation or environmental factors

3) The language delay should not be directly attributable to hearing loss in childhood. Cases of partial hearing loss may be included if the hearing loss is considered to be a complicating factor but not a sufficient direct cause for the language delay

4) The language delay should not form part of a pervasive mental retardation or global developmental delay

5) The language delay may be accompanied by associated problems and is often followed by difficulties in reading and spelling, abnormalities in interpersonal relationships, and emotional and behavioural disorders

NOTE : A distinction is made between expressive language disorder (in which a child's ability to use expressive spoken language is delayed, but language comprehension is within the normal limits) and receptive language disorder (in which the child's understanding of language is below the appropriate level for their mental age).

DSMIV Criteria for Mixed Expressive-Receptive Language Disorder (American Psychiatric Association 1994)

1) Receptive and expressive language skills which are substantially below those obtained from standardised measures of nonverbal intellectual capacity

2) Language difficulties interfere with academic or occupational achievement or with social communication

3) The language delay should not be directly attributable to mental retardation, a speech-motor sensory deficit, or environmental deprivation

4) The language delay should not form part of a pervasive mental retardation

5) The language delay may co-occur with Attention-Deficit/Hyperactivity Disorder, developmental co-ordination disorder and enuresis

NOTE : A distinction is made between expressive language disorder (where no impairment in receptive language skills is apparent) and mixed expressive-receptive language disorders (where the impairment is in both expressive and receptive language)

1.3 IS LANGUAGE INNATE?

1.3i - The Imitation-Reinforcement Model of Language Acquisition

Until quite recently, it was widely believed that language acquisition was governed by the same conditioning reflexes as all other human behaviour (Skinner 1957). Under this behaviourist theory children acquire language through a process of imitation and reinforcement which is lead by their caregivers. Positive reinforcement is received as praise from adults and negative reinforcement as misunderstanding or correction. However, it is now felt that such a model is too simplistic to account for all aspects of language acquisition.

Under the behaviourist theory, language is essentially 'taught' to children by their caregivers. However, studies show that the amount of formal instruction a child receives cannot account for the level of linguistic performance they achieve. Providing the meaning is intact, adults rarely correct a child's syntax or grammar. Instead they tend to concentrate on the positive aspects of the infant's speech (Hirsh-Pasek et al 1984, Demetras et al 1986, Bohannon & Stanowicz 1988). It is unclear whether such a sporadic level of feedback would be strong enough to reinforce the language learning process. Furthermore, it has been demonstrated that shortfalls in the linguistic environment do not have the impact on language development that one would predict under behaviourist theory. Research with children cared for by severely depressed mothers or deaf parents with limited oral skills indicate that only a minimal level of linguistic input is required for normal language development to occur (Schiff-Myers et

al 1988, Murray et al 1993, Murray et al 1996, Skuse 1998). It would appear, however, that feedback does play an important role in the development of language. In a few known cases, where the linguistic input to a child has been limited to the television or radio, language has never been found to develop normally (e.g. Sachs et al 1981).

Additional inadequacies of behaviourist theory are highlighted by investigations of the pattern of language acquisition and the mistakes that children make whilst learning language. Studies show that children follow a surprisingly uniform pattern when learning language. The majority of children pass through stages of babbling, single word speech, two word utterances, morphemes, complex constructions etc. until they reach a state of linguistic fluency. Even deaf children are seen to go through manual stages equivalent to these verbal milestones. Moreover children follow these stages along a particularly robust time-scale and sequence (Pinker 1994, Petitto 1997). For example, in English, when a child is learning the negation process they will typically begin by putting a negative at the beginning of a sentence 'No, want juice'. They then pass through a stage where the negative precedes the verb 'I no want juice' and then they finally move on to the complex use of negative auxiliaries 'I don't want juice'. Under a theory of imitation children would not be expected to produce utterances such as 'no want juice' because they would never have heard their caregivers use such phrases. Under a theory of reinforcement, children would not be expected to pass through recognised set stages of learning but instead would be predicted to follow their own unique pattern of development, determined by their environment.

One aspect of all languages is the ability to produce and understand utterances which use unique combinations of words and express ideas which are novel to us. So although children may never have heard the words 'shoe' and 'furry' used in combination before they have the ability to combine them if they feel it expresses their thoughts. A reinforcement mechanism cannot account for how children do this. The infinite number of combinations of words dictates that we must learn rules underlying language structure rather than rote learning language itself.

1.3ii - Innate Language Systems

The main problem facing any theory of language acquisition is explaining how children, with zero knowledge of language and a minimal level of instruction, manage to pick up and apply the abstract rules of a language, within the short time frame that they do. Thus most modern theories of language acquisition, agree that at least some aspect of language acquisition must be innate. The question is how much do we know about language before we are born and what form does this pre-natal knowledge take?

The rationalists (e.g. Japersen, Chomsky, Pinker) claim that language is unique to the human species and forms part of our biological endowment i.e. it is genetically determined. They argue that, at the time when language is acquired, children are incapable of logic, organised thought and therefore we must possess an innate specific language faculty. This faculty forms an abstract set of rules that facilitate language acquisition and are independent of intelligence, cognition or experience (Harley 2001). Under such a theory, SLI would be caused by a defective language module which hampers subsequent language learning processes.

The existence of such a language module could explain why children follow a uniform schedule of language acquisition and only require a minimal level of instruction. In addition, an innate language instinct would account for the human drive to learn language. Studies have shown that infants are more sensitive to the features of speech than to those of any other sounds. Very young infants are able to recognise the pitch and rhythm (prosody) of their native language and babies as young as three days old can identify the voice of their mother (DeCasper & Fifer 1980, Pettito 1997). Furthermore, when brought up in the absence of a structured language, children have been documented to invent their own languages. In early American slave societies, adults from many different cultures were forced together and so communicated in simplified, hybrid languages, or pidgins, which lacked grammatical structure and syntax. Children born into these societies adopted the pidgins as their language and generated rules and constraints that formed the basis of a new structured language known as a Creole. Creolised speech involves embedded and relative clauses, aspectual distinctions and consistent word order despite the absence of these features in the input pidgin language

(Pinker 1994). Creoles have also been seen to develop in deaf communities where no sign language education is given (Pinker 1994, Pettito 1997).

Language is considered to be a universal feature of the human species. No human civilisation has ever been documented which lacks a complex language as a means of communication. This is true even for highly isolated tribes in countries such as New Guinea who were not discovered until the late 1960's (Pinker 1994).

1.3iii - Universal Grammar

In 1965 Chomsky proposed that the innate language system must take the form of a universal grammar (Chomsky 1965). The Universal Grammar is defined as an innate and invariant set of abstract rules which can be applied to all human languages and enable a child to learn their native tongue.

Chomsky splits language into two levels; I-language and E-language. I-Language is the innate system of linguistic knowledge that is highly abstract and remote from ordinary behaviour and mechanisms. I-language may differ between individuals and alter over time but is close enough between individuals to allow them to communicate. E-Language is any concept of language that must be learnt and is not innate (Chomsky 1986).

I-language takes the form of a universal grammar. Universal grammar is not actually a grammar, or a theory of knowledge. It is a computational system of innate principles that determine what can and cannot happen. The bases of these principles are universal to all languages but at the higher level there exist specific parameters that may vary between languages (Chomsky 1981). Your native language determines the exact nature of each of these parameters, and changes in a single parameter may result in complex consequences across many other parts of the grammar. In the presence of universal grammar, language acquisition proceeds according to it's own internal, predetermined course. 'Learning' a language involves learning how to apply the principles of universal grammar to your native tongue, and this involves the 'setting' of the parameters attached to your universal grammar. It is proposed that each parameter within the innate language module has a default setting that will apply unless evidence to the contrary is obtained. Thus, we may think of universal grammar as a partially wired circuit. The circuit is associated with a finite number of switches,

each of which has a finite number of positions. Experience is required to set the switches and it is the position of these switches that determine which language you learn. Once all the switches are set the system functions.

The rationalists account of innate language modules and universal grammar infers that children possess some level of linguistic knowledge from the moment they are born. However, all attempts to demonstrate the existence of such knowledge have been unsuccessful (Bishop 1997a). Rationalists argue that although the principles and parameters of universal grammar are present from birth, they require certain 'trigger' events before they become apparent. Some propose that the expression of any innate linguistic knowledge is limited by a child's intelligence and memory and so is not apparent from birth. Others claim that the innate knowledge is not accessible from birth but becomes gradually available as the child matures (Harley 2001). The main difficulty with the innate theory of language acquisition is that the abstract nature of the hypothesis itself precludes the demonstration of its existence.

1.3iv - Empiricists and Connectionists

Empiricists (e.g. Locke, Hume) maintain that the inadequacies of the Behaviourist theory do not necessitate the existence of an innate language module. They argue that all knowledge must be derived from experience and that complex behaviour can emerge from the interaction of many simple processes without the need for a specific innate ability. Thus children can use the same intellectual capacities to solve the language problem as they would for any other cognitive challenge. According to the empiricists, SLI is caused by a subtle general cognitive deficit that is more prominent during language processes because of the high demands that these tasks place upon the cognitive system.

The parameter setting model predicted by an innate language theory does not allow for the adjustment of knowledge over time and cannot account for periods of incomplete knowledge during language acquisition. Empiricists, and particularly Connectionists, propose that a much better account of language acquisition can be formed from a probabilistic learning model where modularity is a gradually emerging property rather than an innate one. Thus language is learned by forming relationships between elements and meanings through processes of trial and error and the strengthening of meaningful relationships (Elman 1990, 1993).

Unlike the behaviourist imitation and reinforcement model, a probabilistic model does not require linguistic input to be grammatically correct all the time. In addition it does not demand the presence of negative feedback for reinforcement. Instead it postulates that an absence of positive feedback would be sufficient to reinforce language learning. Studies show that although negative feedback is minimal between the caregiver and child, adults tend to repeat sentences, ask questions or change the topic when a child's utterances are not grammatically correct (Hirsh-Pasek et al 1984, Demetras et al 1986, Bohannon & Stanowicz 1988, Saxton 1997).

Under the empiricist model, environmental aids such as child-directed speech and contextual clues act to ease the process of language acquisition whilst minimising the need for structured positive and negative feedback. When speaking to children adults tend to talk slower, use simpler sentences and an exaggerated innotation and refer to things in the immediate environment. They are more likely to use nouns and words that refer to whole objects rather than parts of objects or attributes and the most important words receive the greatest stress. Studies show that these characteristics act to gain and hold a child's attention and that children show a strong preference for listening to child-directed speech over that of normal adult conversation (Fernald 1991). However, child directed speech is not a universal phenomenon and there is no evidence to suggest that children whose parents use child-directed speech acquire language faster than those who do not (Ochs & Schieffelin 1995, Crago et al 1997).

Under the empiricist theory the development of language relies heavily upon the development of logic and general cognition skills. Whilst a complete dissociation of these two processes is unlikely, studies of aphasic adults demonstrate that the loss of language in isolation from general cognition is a common phenomenon (Schwartz 1984). In addition, children with spina bifida (Pinker 1994) and Williams syndrome (Karmiloff-Smith, et al 1997, Jarrold et al 2001) often suffer from significant retardation, but yet still develop acceptable levels of language skills.

1.4 IS SLI GENETIC?

It was initially assumed that SLI must result from a lack of linguistic stimulation during infancy. However, the high prevalence of the disorder, the disparity in language development between siblings, and the low level of verbal stimulation required to trigger normal language development (as discussed above) are all inconsistent with such a hypothesis. Furthermore, studies of prenatal or birth events (e.g. perinatal brain damage, toxaemia and other pregnancy complications) have also failed to identify any consistent factors that influence language outcomes (Bishop 1997b, Tomblin et al 1997b). In contrast, a large body of evidence has now gathered supporting the role of genes in the aetiology of speech and language disorders.

1.4i - Family Studies

A series of studies have investigated the prevalence of language impairments both in relatives of individuals with SLI and in the general population. However, the heterogeneity of SLI and a lack of consistency in diagnostic criteria often render the direct comparison of family studies impractical. Probands may be ascertained from a parental or school questionnaire, from a clinical referral or via the screening of random children with standardised language batteries. Studies may classify family members as affected if they have a history of dyslexia or poor school performance whereas others include only those who have received specialist speech therapy. Some discriminate between language impairment and language delays or select those with specific phonological or grammatical disorders. Given these discrepancies in study design, it is striking that most investigations have demonstrated a strong familial aggregation of cases of language impairment, and many agree upon the importance of genetic factors in the development of SLI (Bishop & Edmundson 1986, Neils & Aram 1986, Tallal et al 1989).

In a recent report, Stromswold (1998) reviewed eighteen studies, all of which investigated language impairment within groups of children and their relatives. Of the seven studies that compared the family history rates in proband and control cases, all reported an increased family history within the proband grouping (Bishop & Edmundson 1986, Neils & Aram 1986,

Tallal 1989, Tomblin et al 1991, Beitchman 1992, Tomblin 1996, Van der Lely & Stollwerck 1996). Eleven of the eighteen studies evaluated the rate of SLI in family members of language-impaired children (Neils & Aram 1986, Lewis et al 1989, Tallal et al 1989, Tomblin 1989, Haynes & Naidoo 1991, Whitehurst 1991, Lewis 1992, Tomblin & Buckwalter 1994, Lahey & Edwards 1995, Van der Lely & Stollwerck 1996, Rice et al 1998). Across these studies, the incidence of SLI within the relatives of probands ranged from 20% (Neils & Aram 1986) to 42% (Tallal 1989) with a mean of 28%. In contrast, the percentage of control family members who were language-impaired ranged between 3% (Neils & Aram 1986) and 19% (Tallal et al 1989) with a mean of 9%. Stromswold reported that the majority of variability in the prevalence rates between the reviewed studies could be explained by the differences in diagnostic criteria used. In general, those studies which applied a restrictive definition of SLI (e.g. Bishop & Edmundson 1986, Tomblin 1989) found lower prevalence rates than those which relied upon more relaxed conditions (e.g. Tallal et al 1989).

Neils & Aram (1986)

Proband affection status	A diagnosis of SLI, as determined by a certified speech-language pathologist and a score of < 85 on a language screening test
Family affection status	Report questionnaire. Classified as affected if they have a history of speech or language delay, stuttering or a reading problem
Control group	Matched for age and SES, considered by their teachers to have age-appropriate language abilities and a score of > 85 on a language screening test
Exclusion criteria	Abnormal hearing, performance IQ< 85, neurological disorders, structural or oral deviation, non-English speaking families

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
1st degree relative	63.5%	8.3%	7.7	<0.001

Tallal, Ross & Curtiss (1989)

Proband affection status	Clinically tested. Classified as affected if they had a mean language age >1yr below both mental and chronological age
Parent affection status	Self report questionnaire. Classified as affected if they had a history of language, reading or writing problems, or were ever held back a grade
Sibling affection status	Parental report questionnaire. Classified as affected if any problems with reading, writing, language or other learning difficulties
Control group	Selected from same longitudinal study as probands, matched for age, race and SES
Exclusion criteria	PIQ<85, hearing loss, oral, structural or motor impairments, multilingual (or not English) background, autism, seizures

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
Mothers	36.8%	18.5%	1.97	<0.05
Fathers	43.4%	20.4%	2.13	<0.02
Sibs	37.4%	19.3%	1.93	<0.05
1st degree relative	77.0%	46.0%	1.67	<0.002

Table 1.2 - Family Studies of SLI

Tomblin (1989)

Proband affection status	Clinically diagnosed as language-impaired, including phonological disorders
Family member affection status	Report questionnaire. Classified as affected if they reported any articulation or language problems, or stuttering which had been treated by a speech therapist
Control group	Age matched children
Exclusion criteria	Mental retardation, hearing impairments reported in questionnaire

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
Mothers	19.57%	4.51%	4.3	<0.001
Fathers	19.05%	3.31%	5.8	<0.0001
Brothers	39.9%	2.15%	18.5	<0.001
Sisters	16.67%	1.22%	13.6	<0.0001
1st degree relative	22.93%	2.98%	7.7	<0.0001

Rice, Haney & Wexler (1996)

Proband affection status	Receptive language measure >1SD below age-expected and a shorter mean length utterance than age-expected. Probands had specific deficits on a test of tense-marking
Family member affection status	Mostly collected by phone interview - classified as **speech/language-impaired** if they had received speech or language treatment, or reported a history of speech, language difficulties. Classified as **reading/learning impaired** if they reported a history of reading/learning problems
Control group	Used both age-equivalent and language equivalent controls. Both groups were combined for frequency estimates
Exclusion criteria	Abnormal non-verbal skills, articulation problems, limited intelligibility, hearing problems, known neurological, social-emotional or behavioural disorders

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
Speech/Language-Impaired				
+ve family history	58.1%	19.4%	3.0	NA
Nuclear family	22%	7%	3.1	<0.001
Extended family	14%	6%	2.3	<0.001
Reading/Learning-Impaired				
+ve family history	35.5%	22.4%	1.6	NA
Nuclear family	12%	9%	1.3	NS
Extended family	5%	4%	1.25	NS

Table 1.2 - Family Studies of SLI (cont.)

NA = Not Available

NS = Not Significant

Van der Lely & Stollwerck (1996)

Proband affection status	Language scores >1.5SD below expected for chronological age with non-verbal IQ in the 'normal' range (>85). Greater impairment in comprehension and expression grammatical abilities than in vocabulary abilities
Positive family history	Parental questionnaire or case history. Positive family history recorded if they had received specialist help for a speech problem or a reading or writing problem
Control group	Age-matched normally developing children
Exclusion criteria	Social or emotional abnormalities, autistic-like symptoms, semantic-pragmatic disorder, severe-articulatory speech disorder

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
+ve family history	77.8%	28.57%	2.7	$5x10^{-6}$
Mothers	33.3%	2.04%	16.1	<0.001
Fathers	37.5%	8.33%	4.5	<0.025
Brothers	44.0%	18.92%	2.3	>0.05
Sisters	40.0%	7.5%	5.3	<0.05

Tallal, Hirsch, Realpe-Bonilla, Miller, Brzustowicz, Bartlett & Flax (2001)

Proband affection status	Total language score, or average score across two sub-tests < 85 with a performance IQ>80
Family affection status	Two affection criteria were applied; **(a) questionnaire** - classified as affected if they had received treatment for a speech or articulation disorder or reported a history of language problems **(b) clinical testing** - same criteria as those applied to probands
Control group	Selected from local schools, clinically tested and required to have language scores within the normal range (>90)
Exclusion criteria	Incomplete families (<2 biological parents or no siblings), abnormal hearing, motor handicaps, oral structural impairments, autism, emotional difficulties, neurological disorders

	% Affected			
	Proband group	**Control group**	**Relative risk**	**p**
Questionnaire				
Parents	13.2%	9.6%	1.4	0.54
Siblings	51.4%	6.1%	8.4	<0.001
1st degree relative	32.0%	8.2%	3.9	NA
Clinical testing				
Parents	29.5%	7.7%	3.8	0.004
Siblings	31.1%	6.1%	5.1	0.006
1st degree relative	30.3%	7.1%	4.3	<0.001

Table 1.2 - Family Studies of SLI (cont.)

NA = Not Available

1.4ii - Twin Studies

Whilst family studies indicate that language impairments tend to run in families, it remains possible that this may be caused by environmental factors which are shared among family members (i.e. the biological and cultural environment). Twin studies provide a framework within which we are able to separate out genetic and environmental factors. Identical or monozygotic (MZ) twins share all of their genes whereas dizygotic (DZ) or non-identical twins share only half their genes — as do any other sibling pair. However, unlike sib-pairs, DZ twins can be assumed to share a high level of pre-natal and post-natal environmental experiences. Thus if a trait is genetically determined, the fraction of twin pairs who share that trait compared to those where only one twin of the pair is affected (i.e. concordance) can be expected to be greater for MZ twins than for DZ twin pairs. Such an approach however, does require certain assumptions to be made.

The equal environment assumption - we assume that since both DZ and MZ twins are both reared together, then they must share a comparable level of environmental similarities. However, it has been suggested that, because of their appearance, MZ twins are often treated more similarly than DZ twins (Lewontin et al 1984). Whilst this has been shown to be true for some aspects of development, no-one has yet identified any environmental factors which are associated with language competence and are more similar for MZ than DZ twins (Lytton 1980). Thus the equal environment assumption is presumed to be valid for the study of language development.

The twinning effect - If twin data is to be applied to language in a non-twin environment we must assume that twinning itself has no effect upon language development. However, there is ample evidence that the early language development of twins is somewhat slower than that of singletons (Lytton 1980, Clark & Dickman 1984, Tomasello et al 1986, Hay et al 1987, Stafford 1987, Bishop & Bishop 1998). These language delays have been hypothesised to be due to unfavourable pre-natal conditions, increased prematurity and reduced caregiver attention in the twinning situation (Lytton 1980). Nevertheless, it should be noted that the typical twin delay persists only in the early years, and that when twins reach school age their

language abilities are generally comparable to that of their singleton peers (Wilson 1975). Thus the twinning effect should not be valid in samples of school-age twins.

Twin studies of SLI consistently indicate a significant increase in monozygotic concordance rates over those of dizygotic twins (Lewis & Thompson 1992; Bishop et al 1995; Tomblin & Buckwalter 1998) suggesting that much of the reported familial aggregation can be attributed to genetic influences (see table 1.3).

Lewis & Thompson (1992) carried out the first large twin study of SLI using 57 same-sex twin pairs. In this sample, 86% of the MZ twins had both received specialist treatment for a speech and language impairment, whereas only 48% of the DZ twins presented with a concordant phenotype. In addition, it was reported that first-degree relatives of concordant twin pairs experienced a higher risk of language problems than those of disconcordant twins. Analysis of twin subgroups, stratified according to impairment type, indicated that MZ twins were more likely to have similar patterns of language impairments than DZ twins. It should be noted, however, that the sizes of these subclasses were often small and that approximately 80% of the Lewis and Thompson sample were found to have simple articulation disorders. In addition, no exclusion criteria were described for the study.

Bishop et al (1995) studied a set of 90 same-sex twin pairs, all with at least one twin affected by a developmental speech or language disorder. Using a strict definition of language impairment (i.e. a discrepancy of at least 20 points between verbal and non-verbal abilities) they found a male-male MZ concordance of 70% and a DZ rate of 46%. Relaxation of the diagnostic criteria to include those co-twins with a non-verbal IQ below 70 or a past history of speech and language problems resulted in heightened concordance rates of 92:62 (MZ:DZ) for male-male pairs and 100:56 for female-female twins. In an extension of this twin pair study, individuals were sub-classified according to the type of disorder they displayed. Of the four groups formed (articulation with or without receptive (A), articulation and expressive with or without receptive (AE), expressive with or without receptive (E) and pure receptive (R)) those which included children with expressive impairments (i.e. the AE and E subgroups) showed probandwise concordance rates close to 100:50 (MZ:DZ). In contrast, the pure

receptive disorders showed little evidence of genetic influence (MZ:DZ = 71:75) (Bishop et al 1995).

Furthermore, it appeared that the composition of the receptive only (R) group was different from that of the other subtypes. Those twins with pure receptive disorder showed a high level of disconcordance from their co-twins and, unlike the other subtypes, were represented by equal numbers of males and females.

Tomblin and Buckwalter (1998) studied 120 twin pairs where affected individuals were defined as having a composite language score (computed from 4 receptive measures, and 4 expressive measures) one standard deviation below that expected for their age and a non-verbal IQ of above 70. Using this sample they demonstrated an MZ concordance rate of 96% and a DZ concordance rate of 69% (Tomblin & Buckwalter 1998).

Affection criteria		No. of twin pairs	Age	Probandwise concordance		Exclusions
Lewis & Thompson (1992)						
Phone interview identified twins as affected if they had received speech and language treatment. Follow up phone interviews sub-classified language problems as articulation, language, voice, delayed onset of speech, tongue thrust, lisp, stuttering, hearing loss and learning disorders	Speech and language disorder	32	6-12	MZ	0.86	None given
		25		DZ	0.48	
	Articulation disorder	21		MZ	0.95	
		18		DZ	0.22	
	Delayed speech	15		MZ	0.53	
		9		DZ	0.33	
	Learning disorder	7		MZ	0.71	
		2		DZ	0.00	
Bishop, North & Donlan (1995)						
Strict Criteria - Both twins showed a 20 point discrepancy between non-verbal IQ and at least one of 4 language tests used (expressive language skills, vocabulary, grammar and comprehension)	m	52	7-25	DZ	0.46	Mental retardation, sensorineural hearing loss, structural abnormalities of the articulators, serious visual impairments, medical syndrome (e.g. myotonic dystrophy), mild or transient language problems (e.g. lisping), English not first language
		20		MZ	0.70	
	f	11		DZ	0.44	
		7		MZ	0.43	
Relaxed Criteria - One twin met strict diagnostic criteria, the other had non-verbal IQ<70 OR at least one language score below 81 but no discrepancy with non-verbal IQ OR speech therapy history	m	52		DZ	0.62	
		20		MZ	0.92	
	f	11		DZ	1.00	
		7		MZ	0.56	
Tomblin & Buckwalter (1998)						
Composite language score > 1SD below expected		85	4-16	MZ	0.96	Non-verbal IQ <70, hearing problems, additional developmental and behavioural disorders
		35		DZ	0.69	

Table 1.3 – Twin Studies of SLI

m=male
f=female

1.4iii - Adoption Studies

Another method by which it is possible to separate out cultural and biological factors is an adoption study. By comparing the occurrence of SLI in adopted children and their biological (i.e. genetic) and adoptive (i.e. environmental) parents we are able to quantify genetic and environmental effects. Unlike twin studies, adoption studies make no assumptions about the environment. However, they do entail a complex study design and require large numbers of adoptive children with consenting and available biological and adoptive parents. For these reasons twin studies remain a more commonly employed strategy and, to date, there is only one adoption study of SLI in the literature.

Felsenfeld & Plomin (1997) used pre-existing data collected by an adoption agency to undertake a retrospective study of SLI. From this pool they selected 66 children whose parents were affected by language impairments • 16 with biologically affected parents, 19 whose adoptive parents were affected, and 31 who lived with their natural, affected parents. In addition 90 control children, whose parents showed no signs of language delays, were selected at random. The speech status of these children was assessed from a pre-recorded sample of speech, a parental phone interview, and the child's history of speech and language therapy. Of the 156 children 22 (13 male) were considered affected. This included 5 (31.3%) from the biological group, 2 (22.6%) from the adoptive group, 7 (22.6%) from the non-adoptive group, and 8 (8.9%) from the control group.

Using the speech history data, Felsenfeld and Plomin performed a logistic regression analysis to identify the best predictors of speech and language problems. They examined five variables (child sex, biological parent affected, rearing parent affected, score on an environmental test, and non-verbal IQ), of which the only significant risk factor was having an affected biological parent. It was estimated that this attribute alone elevates a child's risk of speech disorder by between 1.16 and 7.47 times, even if that child has never had contact with their affected biological parent. Furthermore, their studies indicated that being reared by a speech-affected parent (biological or otherwise) does not increase a child's risk of developing speech problems above that conferred by genes. Felsenfeld and Plomin concluded that if environmental factors are important in speech disorder development they are likely to be non-

shared factors such as prenatal environment, peer group effects and random environmental influences.

1.4iv - Heritability of Language Measures

Further support for a genetic aetiology in language disorders is provided by estimates of heritabilities of quantitative measures of language-related components. These heritability values are usually calculated using the DeFries-Fulker (DF) method - a regression-based method designed for the analysis of twin data (DeFries & Fulker 1988). In DF analysis the affected twin is designated as a proband, and therefore, by definition, has a language score which falls significantly below the population mean. By comparing the difference in scores between these probands, their co-twins and the population mean, it is possible to estimate the level of genetic effects upon the trait outcome. Just as in twin studies, the differences between DZ twin pairs can be considered to be a function of genetic and non-shared environmental effects, whereas differences between MZ twins will mainly be due to the environmental factors that are unique to each twin. Thus an estimation of the genetic effects upon the trait (the heritability) can be derived by doubling the difference between MZ and DZ co-twins. Additionally, it is possible to investigate the impact of the non-shared environment by looking at the difference in mean score between MZ probands and their co-twins.

In a series of investigations Bishop and colleagues used the DeFries-Fulker method to demonstrate significant heritabilities in several psychometric language measures within families affected by SLI. These included tests of receptive syntactic language abilities (e.g. the Test for the Reception of Grammar (TROG) (h^2_g = 1.09±0.83)), and expressive language skills (e.g. the Clinical Evaluation of Language Fundamentals (CELF) repeating sentences sub-test (h^2_g = 0.56±0.43)), as well as tests which examine specific cognitive processes thought to be important in language acquisition (e.g. non-word repetition (h^2_g = 1.17±0.32)). Interestingly, whilst many language-related traits were shown to be highly heritable, when the discrepancy score between these traits and non-verbal IQ were considered, no significant heritability was seen (Bishop et al 1995, Bishop et al 1996, Bishop et al 1999a).

Tomblin & Buckwalter (1998) also used De-Fries Fulker method to demonstrate significant heritabilities (ranging from 0.56 to 1.35) for the language scores assessed in their twin study.

Again they demonstrated that these positive levels of heritabilities do not extend to the IQ discrepancy scores. In addition, they demonstrated that the heritabilites for the group of children with non-verbal IQs of below 70 was not significantly different from those with IQs in the 'normal' range.

It should be noted that the heritability measures (h^2_g) provided by each of the above studies are calculated from extreme proband scores and therefore can only be extended to represent the trait heritability in the general population (h^2) if the genetic factors acting in the two populations are identical. Dale, Siminoff, Bishop, Eley, Oliver, Price, Purcell, Stevenson & Plomin (1998) attempted to quantify the deviations between h^2_g and h^2 in a language-related measure using a large sample of twins. They assessed 3039 two-year-old twin-pairs (1044 MZ and 1006 same-sex DZ and 989 different sex DZ) and compared the trait heritability of a productive vocabulary measure in children performing below the 95th percentile with that of the remainder of the population. A maximum likelihood model-fitting analysis (Neale & Cardon 1992) was used to partition trait variance into genetic (h^2) and environmental factors within the normal population and the DF method to estimate trait heritability (h^2_g) within the extreme ('language-impaired') range. h^2_g was estimated at 73% (95% Confidence Interval (CI) = 0.38-1.00) whereas h^2 was estimated at 25% (95% CI=0-0.78), thus indicating that whilst genetic factors are important within the language-impaired sample, they have little relevance within the language-normal sample. When progressively relaxed cut-off values were imposed to select the language-impaired sample the extreme heritability was seen to regress back to the trait heritability within the normal population. The remaining variability within the language-impaired sample could be explained by shared environment (18%) and measurement error (9%). This study indicates that the SLI population do not merely represent the bottom 5% of the continuous language performance distribution within the normal population, rather they can be considered as a qualitatively distinct population with different genetic influences from those associated with language variation within the normal range. It should be noted that a study of this design requires large sample sizes and although thousands of twins were assessed within the present study the confidence intervals attached to the heritability estimates still remain quite large. In addition, as discussed above, caution should be taken

with studies of language which use young twin pairs. Given that vocabulary level at age two is not a good predictor of language outcome in later infancy and that twins tend to be delayed at very young ages compared to the general population it remains unclear how this study specifically applies to the general SLI population.

In conclusion, the last few decades have witnessed a shift in the theory of the aetiology of SLI. Not unlike those theories governing general linguistics, this shift has been away from environmental causes and towards a more genetic and biological point of view. However, it should be noted that although there is now ample evidence to support a genetic role in the determination of absolute language abilities, family studies of SLI have failed to detect any clear co-segregation between phenotype and genotype and most conclude that the genetic basis of the disorder is likely to be complex (Bishop et al 1995).

1.5 CAUSAL THEORIES OF SLI

Over the years many theories as to the underlying causes of SLI have been suggested. Although each of these is a distinct theory with differing details, most can be classified under three headings. There are those which propose SLI is caused by an impairment in auditory perception, others suggest that an innate grammar impairment underlies SLI and others suggest that SLI is caused by limitations in phonological short-term memory.

1.6 INNATE GRAMMAR IMPAIRMENTS

As an extension of Chomskyan theory, many researchers propose that SLI may arise from the absence of, or deficits in, innate grammatical modules resulting in an inadequate universal grammar. Under such a hypothesis we would expect children with SLI to develop selective impairments which affect specific components of grammar and this, to some extent, appears to be the case. Most children with SLI display some level of grammatical deficits, and these typically include problems with the use of inflections (e.g. I jump, he jumps), tense-marking (I jump, I jumped), auxiliaries (e.g. as, be, may, can, do) and copulas (I am jumping). Studies indicate that up to 50% of the verbs used by language-impaired children lack the obligatory agreements. In addition, their speech often involves a high number of over-generalisation errors and argument omissions, even at later ages when other deficits are not so apparent

(Clahsen 1989, Leonard 1989, Gopnik 1990, Van der Lely & Stollwerck 1997). In contrast, there are other areas of grammar in which children with SLI appear to develop normally. These typically include plural nouns, past tense forms of irregular verbs and the progressive marking of verbs (-ing). In these areas, language-impaired children often perform at levels comparable to their age-peers (Clahsen 1989, Leonard 1989, Gopnik 1990, Gopnik & Crago 1991, Clahsen et al 1992).

The apparent specificity of this deficit has encouraged investigation of the grammatical phenotype in language-impaired children and promoted the development of many theories of innate grammar impairment theory. Despite the number of hypotheses in this field, there is little agreement the exact nature of the grammatical impairments underlying the aetiology of SLI.

1.6i - Gopnik - The Missing Feature Deficit Hypothesis

Gopnik and colleagues propose that children with SLI have an inability to form the implicit grammatical rules that govern number, tense and person (i.e. morphological rules) (reviewed in Gopnik 1997). It is believed that normal children initially learn vocabulary in a rote fashion. However, as they acquire more and more words, their memory becomes saturated, and this system becomes less efficient. They therefore look for an alternative strategy to word storage and, around this stage, begin to recognise associations between related words. The derivation of the rules which relate structures such as present and past tense (e.g. walk and walked) enable children to minimise the number of words which they need to store and thus the memory load of the lexicon. At this time, the vocabulary of children is reorganised and they begin to store stems of words (i.e. walk) from which they are able to apply their new found rules to derive any form of the word which is required (e.g. he walks, they walked, I'm walking etc.). For words that take irregular forms, however, separate lexical entries are still required. Gopnik proposes that although children with SLI are able to recognise that grammatical features, such as the past tense, require marking, they do not realise that there is a generalised rule that governs these markings. They therefore approach all words as if they were irregular and continue to rote learn vocabulary which places a great burden on the processes of speech understanding and production.

Goad & Rebellati (1994) showed that children with SLI may benefit from explicit education in the grammatical rules that govern inflection (e.g. add -s for plural). However, once taught these rules, children with SLI often make over-generalisations and may fail to make voicing assimilations for inflected forms. For example, the -ed ending in the word 'jumped' has a different sound from that in the word 'jogged' (jump[t] and jog[d]). A normal child is able to recognise that although these sounds are different the principle governing the inflection is the same. In contrast, children with SLI do not acknowledge this system and therefore do not automatically produce the correct forms (Goad & Rebellati 1994, Gopnik 1994, Ullman & Gopnik 1994).

Much of the evidence to support Gopnik's theory arises from comparisons of tense marking abilities between children with and without SLI. For example, if a normally developing child is shown a picture of an animal and told it is a 'wug', more often than not, they are able to generate the word 'wugs' to describe another picture that shows two such animals. In contrast, a language-impaired child will struggle to produce the plural form of this nonsense word (Gopnik 1997). However, if a child with SLI is asked to point to the picture of 'books', they are often successful in choosing a picture showing a pile of books over one that shows a single book. Thus a child with SLI is able to understand that 'book' refers a singular book whereas 'books' refers to more than one. Nevertheless, they are unable to generalise this concept to novel words such as 'wug' (Gopnik 1997). Additional support is provided by comparisons of lexical retrieval times between normal and language-impaired children. It has been noted that, following the lexicon rearrangement in young children, the access of inflected forms of words takes longer than those of stem words. Thus in normal children, the decision that 'plip' is a false word can be made quicker than the same decision for the word 'plipped'. This is because they must take the word 'plipped' and shorten it to its stem form before they are able to analyse whether it is a real word or not. In children with SLI this time delay is not seen. This is thought to indicate that children with SLI treat the words plip and plipped as separate entities, rather than different forms of the same word (Paradis & Gopnik 1994).

Adversaries of the morphological rule theory, also rely upon evidence from the speech of language-impaired children. Researchers have demonstrated that over-generalisations are common in the speech of children with SLI (Leonard et al 1992a, Vargha-Khadem et al 1995) and that language-impaired children occasionally do manage to correctly inflect a novel word (Connell & Stone 1992, Bellaire et al 1994). These two observations are difficult to incorporate into a theory that presupposes a total absence of morphological rules.

1.6ii - Rice - The Extended Optional Infinitive Account

Rice proposes that children with SLI do recognise the existence of morphological rules, but regard their application as optional (Rice et al 1995).

Very young normally developing children typically use the infinite form of a verb for all situations (e.g. 'Daddy like me'). As they begin to recognise and apply morphological paradigms they use both the grammatically correct finite form (e.g. 'Daddy likes me') and the immature infinite form intermittently. After they have correctly deduced the grammatical rules governing the use of finite and infinite forms, they will move into a linguistically mature stage where they are able to apply the finite form and always produce the correct agreement (Wexler 1994). Rice proposes that children with SLI remain at a stage of optional infinitive agreement for an undefined period of time that may even last until adulthood. Under her hypothesis the application of grammatical rules suffers as the demands of a linguistic situation increase (Rice et al 1995, Rice & Wexler 1996).

Rice's theory allows for the intermittent production of correct and incorrect grammatical forms that are typical of children with SLI.

1.6iii - Clahsen - Agreement Marking Deficit Hypothesis

Clahsen observed that whilst children with SLI show some level of understanding with regards to morphological paradigms, they do not always apply this knowledge in the correct way (Clahsen 1989). He noted that children with SLI often experience disproportionate problems in the marking of grammatical agreement e.g. subject-verb agreement, auxiliaries and definite and indefinite articles. In contrast, they may encounter few problems in areas such as noun plurals and past tense forms of irregular words. In light of these observations, Clahsen (1989) proposed that SLI is not caused by a deficit in morphological rule formation per se, but rather

by a problem in establishing relationships between sentence constituents which makes it hard to recognise when it is appropriate to mark agreement. For example, in the sentences 'the girl drinks' and 'they drink', the differences in verb ending rely upon the relationship between the noun-phrase ('the girl' or 'they') and the verb ('drink'), which must agree in number. Thus a child has to first establish the nature of the relationships between the noun-phrases and the verb before they can decide upon which inflectional agreement is appropriate.

Clahsen developed and applied the missing agreement account to his observations in German children, and whilst many features can be transferred to other languages, a few suffer from cross-linguistic discrepancies. For example, Clahsen's theory can make no provision for why, in languages such as Hebrew and Italian, the use of verb agreement inflections are comparable between language-impaired and normal children. In addition, English-speaking children with SLI often experience problems with regular past verb inflections (-ed) which require no assumptions about inflectional agreement (Leonard et al 1987, Leonard 1998).

1.6iv - Van Der Lely - Representational Deficit for Dependent Relationships

In an extension of Clahsen's hypothesis Van der Lely and colleagues proposed that some forms of language impairment might be caused by difficulties with the derivation of thematic roles on the basis of syntactic relationships (i.e. sentence structures) alone (Van der Lely & Stollwerck 1997). Thus children with SLI disregard the sentence structure when interpreting meaning and instead rely upon contextual clues such as gender and number or semantic constraints.

In support of this theory, Van der Lely demonstrated that whilst children with SLI are relatively successful when asked to assign meaning to a word in the presence of additional visual information, their performance levels fall if the contextual clues are removed (Van der Lely 1994). For example, if language-impaired children are shown a picture of a teddy kicking a can and presented with the word 'sloodge' in isolation they are able to create a feasible sentence containing the word 'sloodge' (e.g. 'Teddy sloodges the can'). However if given a

teddy and a can and asked to act out a situation where 'Teddy sloodges the can', they are often unable to do so.

Van der Lely demonstrated that, in tasks which require pictures to be matched to sentences entirely on the basis of syntactic information, the performance of language-impaired children deteriorates as the sentence structures become more complex. Thus whilst children with SLI may attain age-expected levels in tests involving simple active sentences ('the man eats the fish') and short passives ('the fish is being eaten'), they performed less well if the task involved full passives ('the fish is eaten by the man'). Ambiguous passives ('the fish is eaten') and reversible sentences ('the fish is being eaten by the man') were noted to be particularly confusing. When confronted with these arrangements, it appeared that the language-impaired children treated the verb as an adjective (i.e. in both the above examples they would point to a fish that had been eaten), overlooking the more complex interpretation (Van der Lely & Harris 1990, Van der Lely 1994, Van der Lely 1996).

It should be noted, however, that many of Van der Lely's investigations involved a subgroup of language-impaired children selected to have a large discrepancy between their grammatical and semantic-lexical abilities (Grammar-specific SLI).

1.6v - Leonard - Auditory Perceptual Deficit & the Surface Hypothesis

Leonard (1989) proposed that the grammatical deficits observed in SLI could be caused by the transient and indistinct nature of the grammatical morphemes. If children with SLI have a perceptual deficit which makes it difficult to identify brief or rapid stimuli (see section 1.7ii 'Auditory Perception Impairments'), then we may expect them to have particular problems with English inflectional endings which tend to be unstressed, of low prominence and therefore hard to detect (e.g. -s). Unlike other grammatical theories, Leonard's 'surface hypothesis' accounts for the differences observed in the grammatical deficits of SLI between different languages. In fact, it directly predicts that a distinct pattern of deficits will be seen for each language, according to the phonetic characteristics of it's grammar. Thus, since Italian inflections tend to be syllabic and stressed, we would expect fewer Italian children to encounter problems with the inflection of third person singular verbs than their English counterparts. In a comparison of English and Italian children with SLI Leonard demonstrated

that this indeed was true. Furthermore, his studies indicated that Italian children preferred to use phonologically apparent pronouns (la and una) to the more subtle masculine forms (il and un) (Leonard 1989).

A second prediction of the surface hypothesis would be that we would expect to see a straightforward correlation between the ease with which an inflection is perceived, and the rate at which it is incorporated into the morphological paradigm. Unlike the cross-linguistic differences, however, there is little evidence to suggest that this is the case. Investigations of English language-impaired children, demonstrate that whilst the -s ending is readily used to mark the plurality of nouns, it's use in a possessive or third singular context is much less common (Rice & Oetting 1993, Bishop 1994b). Similarly, words such as 'on and 'over' can be understood perfectly when in their preposition form (e.g. sit ON the chair), but pose more of a problem when used as a verb particle (e.g. put ON the hat) (Watkins & Rice 1991). Furthermore, some of the grammatical deficits found in SLI typically involve phonetically strong components (e.g. English irregular forms) (Bishop 1997a).

1.7 LIMITATIONS IN PROCESSING CAPACITIES

Some researchers propose that the impairments found in SLI are not specific to language but instead represent limitations in more general processing capabilities. Although these limitations affect all cognitive capabilities investigators propose that they would be most pronounced in those processes which require a large amount of on-line computation e.g. language processing.

1.7i – Short-Term Memory Deficits in SLI

Gathercole and Baddeley proposed that SLI might be caused by a deficit in short-term memory. This is a specialised memory that allows the storage and manipulation of small amounts of information for short periods of time. This system is essential for tasks which involve the immediate recall of information (e.g. remembering a phone number for a sufficient period of time to be able to write it down) and is central to more complex tasks such as comprehension, learning and reasoning (reviewed in Baddeley 1992, Gathercole 1999).

Note that whilst some researchers treat the terms short-term memory and working memory as synonymous, others make a distinction between the two. Short-term memory refers to the site at which the memories are stored and working memory refers to the processes which are required to hold these memories at that site.

Short-term memory can be considered to consist of four components (Figure 1.1) (Baddeley & Hitch 1974, Baddeley 2000).

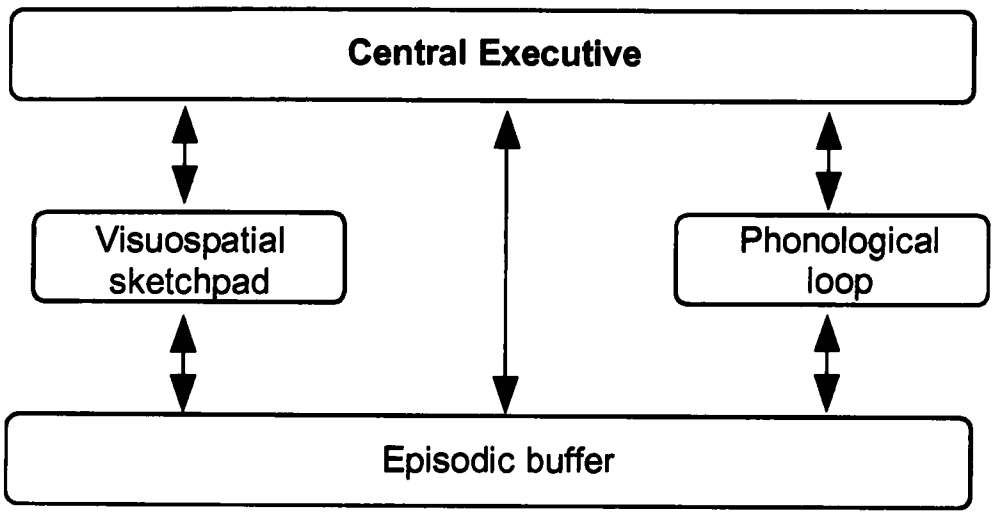


Figure 1.1 – Short-Term Memory : *The short-term memory system consists of a master regulatory central executive that controls three interacting slave systems; the visuospatial sketchpad, the phonological loop and the episodic buffer. Note that the visuospatial sketchpad and the phonological loop act independently of each other and are only able to interact via the episodic buffer*

The central executive manages the short-term memory and is responsible for the division of attentional capacity between the three slave components. It is involved in the storage and retrieval of information, the direction of information through the slaves, and the control of action and planning. Underlying the central executive are the phonological loop, which is responsible for the storage and rehearsal of verbal and acoustic information, and the visuospatial sketchpad, which forms a separate store for visual and spatial information (Baddeley & Hitch 1974). Interactions between the phonological loop and the visuospatial sketchpad are mediated by the episodic buffer, which can also be directly controlled by the central executive (Baddeley 2000). Neuroanatomical studies indicate that the separation of the short-term memory into these functional components may also be reflected at the physiological level. The processes controlled by the central executive occur in the frontal lobe

of the brain (Brodmann areas 9, 10, 44, 45 and 46) whereas the visuospatial sketchpad activities take place in the right hemisphere (Brodmann areas 6, 19, 40 and 47) and the tasks carried out by the phonological loop can be localised to the left hemisphere (Brodmann areas 6, 40 and 44) (Smith & Jonides 1998, O'Reilly et al 1999).

Of the three slaves the phonological loop is responsible for the recollection of linguistic information. The processes controlled by the phonological loop are summarised in the figure below (Figure 1.2).

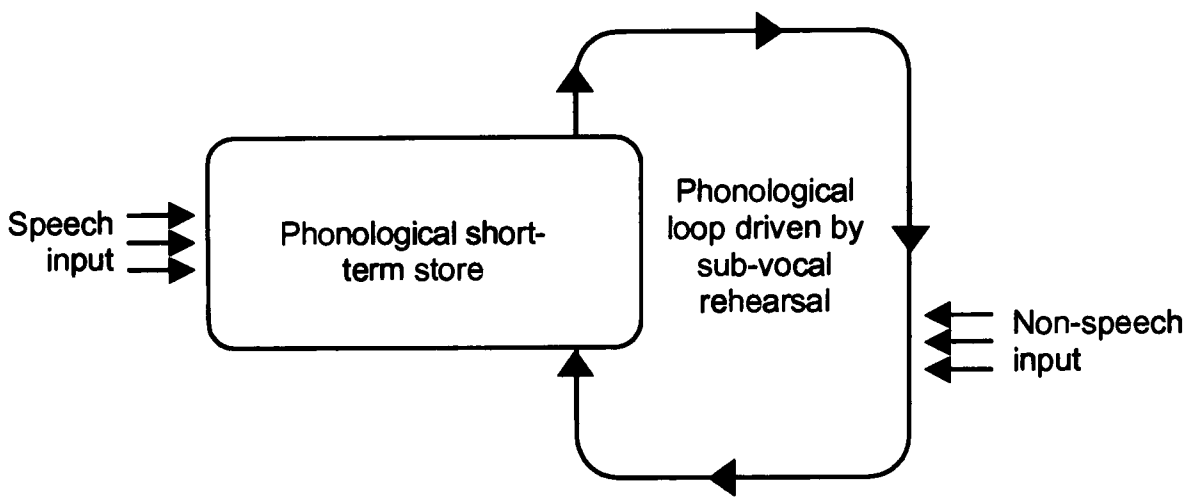


Figure 1.2 – The Phonological Loop : *The phonological loop consists of a temporary store for auditory information. Auditory data is stored as phonological traces that are subject to rapid decay (2-3 secs) unless refreshed by sub-vocal rehearsal. The efficiency of the rehearsal process is limited by the capacity of the phonological store. Thus, even when reinforced, information tends to accumulate inaccuracies over time, as the phonological store becomes saturated. Rehearsal can also act to incorporate non-verbal material into the phonological format via the episodic buffer. Information in the phonological loop that is adequately rehearsed is passed onto the long-term memory for more substantial storage*

The capacity of the phonological short-term memory can be measured using a variety of tests, all of which involve the recollection of verbal information presented in the absence of any visual cues. Examples of such tests include the repetition of nonsensical words (e.g. contramponist) and the immediate recollection of digit spans (e.g. 8, 5, 2, 7, 9) (Gathercole et al 1994, Gathercole 1999, Gathercole & Pickering 2000). The performance of subjects on such tests is subject to characteristic variations, both between individuals, and over time (Isaacs & Vargha-Khadem 1989, Gathercole et al 1994, Siegel 1994, Gathercole & Baddeley 1996, Swanson 1999). The capacity of the phonological short-term memory in children increases exponentially between the ages of four and fourteen. After this point, the memory capacity plateaus and eventually declines in later age. The average adult can be expected to

recall a list of seven digits without error. In general, studies indicate that inter-subject variations in phonological memory abilities are highly correlated with existing vocabulary size and the ability to learn novel words (either in their native tongue or in a foreign language) (Michas & Henry 1994, Service & Kohonen 1995, Baddeley et al 1998, Gathercole et al 1999, Gathercole & Pickering 2000). It is therefore proposed that the phonological loop system is essential for establishing long-term representations of novel phonological forms, and plays a particularly important role in the acquisition of vocabulary. Individuals with Downs syndrome, in which language is severely impaired, perform poorly on tests of non-word repetition, whereas children with Williams syndrome, in which language skills remain relatively intact, usually attain good scores (Grant et al 1997, Jarrold & Baddeley 1997, Laws 1998).

Baddeley suggested that the children with SLI may have difficulties learning language because of a deficit in the phonological loop. He postulated that such a deficit would result in an inability to retain verbal information in the short-term memory for an extended period of time, and thus would prevent the in-depth processing of novel phonological information and it's storage in the long-term memory (Baddeley & Wilson 1993, Baddeley et al 1998). In support of this theory, it has been demonstrated that children with SLI have severe impairments in their phonological short-term memory span. Studies show that individuals with language impairments function poorly in tests of digit span, word recall and non-word repetition (Gathercole & Baddeley 1990, Gathercole et al 1994, Gillam et al 1995, Bishop et al 1999a). Furthermore, the non-word repetition test is capable of identifying adults who experienced language difficulties during childhood, thus indicating that any underlying phonological deficit may persist into adulthood, even after the language difficulties have apparently resolved (Gathercole et al 1994). As predicted by a theory based on phonological memory, children with SLI have a smaller vocabulary store than would be expected for their age, and have been shown to require significantly more exposures to a new word before they incorporate it into their lexicon (Rice et al 1994).

Despite the differences in ability levels between language-impaired and control children on tests of phonological short-term memory, the pattern of performance is strikingly similar across the two groups. In general, lists of phonologically similar words are harder to recall

than lists of phonologically distinct words, and the accuracy of non-word repetition decreases as the length of the words increases (Gathercole & Baddeley 1990, Montgomery 1995). The phonological effect is believed to represent the efficiency with which words are encoded into phonological traces, and the length effect is caused by limitations in the capacity of the phonological store.

In conclusion, there is an accumulation of evidence which support the presence of phonological deficits in SLI. However, it is not clear whether this reduced phonological capacity arises due to a limitation in the available storage space for information or an increased decay rate of information in the phonological loop. Furthermore it has been suggested that a short-term memory deficit alone cannot account for all aspects of the SLI phenotype (e.g. associated spatial deficits) (Van der Lely & Howard 1993). It is therefore generally accepted that, any phonological deficits are likely to be accompanied by additional impairments in associated cognitive domains.

1.7ii - Auditory Perception Impairments

It has been proposed that SLI is caused by a limitation in the processing of brief or rapid stimuli. Although it is not clear whether this impairment would be language-specific it is anticipated that it would be particularly profound in terms of the temporal resolution of speech sounds which involve many brief and rapid stimuli.

Much of this theory is based on observations of the performance of language-impaired children in tests of auditory perception (Figure 1.3).

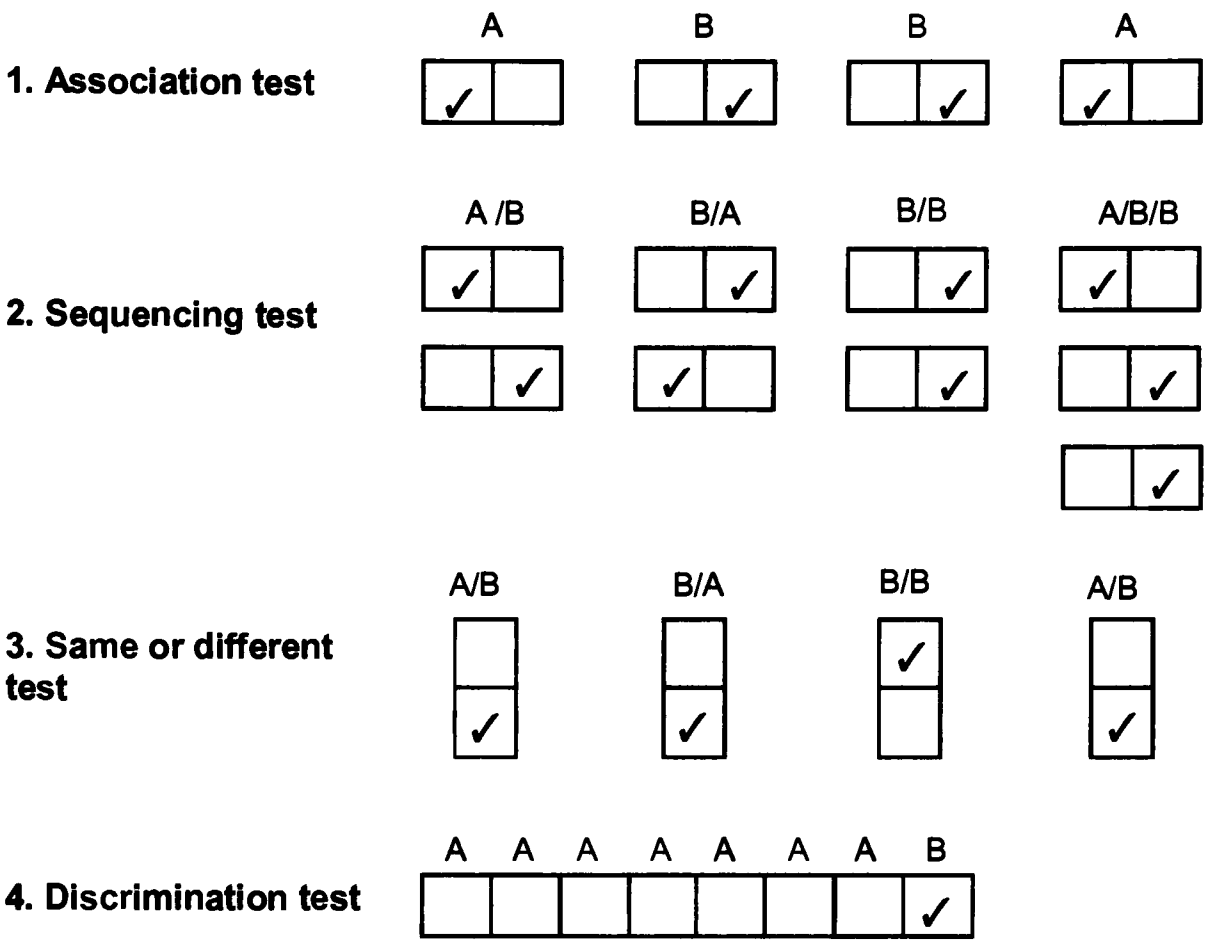


Figure 1.3 - The Auditory Perception Test : *There are four main variations of the auditory perception test (as represented above) (Tallal & Piercy 1973a,b). The letters represent two distinct tones and the boxes signify the responses given by the subject.*
***The association test** usually functions as a training schedule in which the subject is taught to associate tone A with the left side of the response panel, and tone B with the right side of the response panel. Training is given on the association phase until the subject reaches an accepted level of performance.*
***The sequencing test** involves the presentation of pairs of tones with variable inter-stimuli intervals. The subject is required to mark the appropriate response panels to indicate the order of the tones. This test can involve sequences of up to four tones long.*
***The same or different test** entails the presentation of two tones, again with variable inter-stimuli intervals. The subject is required to mark the box according to whether the tones presented are the same or different.*
***The discrimination test** involves a series of identical tones with varying inter-stimulus intervals. The subject is required to respond when they hear a tone that deviates from the sequence.*
Note that different studies may use different variations of tones. Some use pure tones, whilst others use naturally produced or computer simulated speech sounds. It is also possible to produce computer-generated sounds that occur along a continuum between two end point syllables.

Using the auditory perception test, many researchers have demonstrated that children with SLI have problems separating brief sequential tones presented in rapid succession (Lowe & Campbell 1965, Tallal & Piercy 1973a). The introduction of longer intervals between the tones or the lengthening of the tones themselves, has been shown to ameliorate these difficulties

(Tallal & Piercy 1973b). It is believed that this interval effect does not simply represent immature perceptual abilities on the part of language-impaired children, as the performance profile of children with SLI does not match that of normally developing children at any age (Tallal 1976). In sequencing tests that involve three or more tones, children with SLI continue to perform below expected, even after the introduction of long inter-stimulus intervals (Tallal & Piercy 1973b).

If the difficulties of children with SLI are caused by an inability to perceive brief or rapid stimuli, then it follows that their language problems should be more pronounced in tasks which require the discrimination of syllables with rapid frequency changes (e.g. stop consonants) than in those which require the perception of extended sounds, such as vowels. In an investigation of this hypothesis, Tallal and Piercy (1974) demonstrated that whilst language-impaired and control children performed at similar levels in the discrimination and sequencing of vowel sounds ([ɛ] and [æ]), there were significant differences between the two groups when the tasks involved stop consonants ([ba] and [da]) (Tallal & Piercy 1974, Tallal et al 1976). Further studies revealed that the discrimination and sequencing of vowels also became problematic when the vowel was followed by additional acoustic information of a brief duration (Tallal & Stark 1981, Leonard et al 1992b). Furthermore, it was demonstrated that these problems could be ameliorated by altering the sound trace of given syllables (Tallal & Piercy 1975, Tallal & Stark 1981). By doubling the duration of the distinguishing portion of stop constants, the error rates of the language-impaired children could be brought down to a similar level of that in the controls. Similarly, by truncating the distinctive portion of the vowels the performance of children with SLI deteriorated, whereas those of controls remained stable (Tallal & Piercy 1975, Tallal & Stark 1981).

Strong evidence for the involvement of auditory deficits in SLI, comes from studies which demonstrate that intensive training on auditory perception tasks can lead to an improvement in performance (Merzenich et al 1996, Tallal et al 1996). In particular, this improvement in auditory skills has been reported to be accompanied by advances in the child's language function (Tallal et al 1996).

However, controversy still exists as to the exact nature of the perception deficit and the level at which it acts. Whilst some researchers propose the deficit is specific to auditory input, others advocate a generalised deficit which encompasses other cognitive domains. In general, discrimination studies indicate the presence of a visual perception deficit in younger children with language problems (Tallal et al 1981), which are apparently absent from older subjects (Tallal & Piercy 1973b). Thus it is believed that young language-impaired children may suffer from widespread perception problems, the majority of which resolve with age leaving only deficits in the auditory module. These auditory difficulties may also gradually recede over a longer period of time.

Another point of interest is that electrophysiological studies of auditory response do not generally support the findings of behavioural studies. Electrophysiological studies involve the direct measurement of cortical evoked potentials in individuals completing an auditory discrimination task. These studies find no evidence for any differences between the auditory response of language-impaired subjects and language-normal controls (Tomblin et al 1995). This is true, even when the task involves sequences of tones with particularly short inter-stimulus intervals. Thus it is likely that the brains of language-impaired children are actually capable of differentiating between auditory stimuli (Neville et al 1993, Sussman 1993, Kraus et al 1996), and that the auditory problem occurs at a higher level of processing (e.g. in the encoding of sensory trace information). Although a deficit in the encoding process would not be apparent if the interval between stimuli was long, it would be expected to become more pronounced, as the rate of incoming information increased (e.g. in the encoding of incoming speech) (Sussman 1993). Note that deviant cortical responses have been demonstrated for children with receptive SLI, providing further evidence that this may represent a distinct subgroup of language impairment (see section 1.1ii 'The SLI Phenotype') (Stephanos et al 1989).

Also important to the present study is the fact that, although Tallal's auditory repetition task has been demonstrated to reliably identify language-impaired individuals, it shows only low levels of heritability (h^2_0 = 0.109, Standard Error (SE)=0.324 – Bishop et al 1999a). This study concluded that an auditory processing deficit was neither necessary nor sufficient to cause

SLI (see also Bishop et al 1999b). Instead it is thought to act as an environmental factor in SLI, which assumes importance in the presence of a genetic predisposition.

In conclusion, there is ample evidence to support the fact that children with SLI do have problems with auditory processing, particularly when the auditory stimuli are brief or rapid or when there is a large amount of information to be processed. Although this processing impairment may not limited to auditory information it is expected to be most pronounced in the verbal domain as a critical aspect of language learning depends on the ability to detect rapidly changing auditory signals. However, the nature of this impairment is seen to vary over time and there is some disagreement as to whether it exists at the level of perception or auditory trace coding.

1.8 AIMS OF THIS PROJECT

This project aims to clarify the genetic mechanisms underlying Specific Language Impairment. A goal which will be achieved, for the main part, through the completion of a genome screen for loci involved in the disorder. Given the problems associated with the derivation of an affection status for SLI, this genome screen will use quantitative scores of language abilities to assess linkage across the genome.

Subsequent sections will describe the fine mapping of a locus identified by the genome scan and the characterisation of two candidate genes for SLI. In addition I will describe the mapping a 2;11 translocation in a language-impaired family and an association study of the translocation breakpoint on chromosome 2.

In the long-term studies such as the one described here, aim to identify gene variants that underlie language impairment. It is hoped that the identification of these variants will not only clarify the pathways involved in language acquisition but will eventually lead to the development of better diagnosis and treatment for those children with language impairments.

CHAPTER 2

SUBJECTS

&

METHODS

2.1 THE GENOME SCREEN COHORT

2.1i - Ascertainment of Cohort

In total 473 individuals (219 total sib-pairs) were recruited from 98 families, drawn from two independent samples. The Newcomen Centre at Guy's Hospital, London diagnosed and referred a clinically based sample and the Cambridge Language and Speech Project (CLASP) provided families drawn from an ongoing epidemiological study.

The first fifteen families in this study were identified through Guys Hospital and were ascertained on an affected sib-pair basis. It was initially felt that such an approach would enrich the number of affected individuals within the study, and therefore increase the chances of Quantitative Trait Locus (QTL) detection. However, it has been suggested that in the genetic study of common, complex disorders, an alternative selection procedure, which samples individuals from both tails of the distribution, may be more powerful (Cardon & Fulker 1994). Thus the remainder of the Guys sample, and the entire Cambridge sample, were ascertained on the basis of a single affected proband, and information was collected from all siblings regardless of affection status. Of the original fifteen families collected by Guys, ten were excluded (see table 2.1 for exclusion criteria), and thus only five of the affected sib-pair families were included within the final genome screen sample.

As discussed in the introduction (see section 1.1ii 'The SLI phenotype' and section 1.1iii ' The prevalence of SLI'), the heterogeneity of the SLI phenotype often makes the derivation of a consistent qualitative affection status unfeasible. For the genome screen, we therefore chose to employ quantitative measures as indicators of language skills. We used a broad phenotype battery (the Clinical Evaluation of Language Fundamentals (CELF)), which yielded scores of expressive and receptive language ability and a single specific test of non-word repetition (For a full description of these tests please refer to section 2.2 'The genome screen phenotypes'). All three traits have been demonstrated to be significantly heritable and good predictors of language abilities (Semel et al 1992; Bishop et al 1996; Bishop et al 1999a). Within the genome screen sample, all probands, either currently or in the past, had language skills more than 1.5 standard deviations below the normative mean for their chronological age on the

receptive and/or expressive scales of the Clinical Evaluation of Language Fundamentals (CELF-R) battery (Semel et al 1992). Further concerns regarding both the reliability and heritability of IQ discrepancies (see section 1.2i 'Non-verbal IQ') resulted in the use of these language measures as discrete scores of language abilities alongside an IQ cut-off criterion. Thus any proband or sibling found to have a non-verbal IQ of below 80 was excluded from the genome screen.

In line with the DSM-IV and ICD-10 criteria (see section 1.2vi 'Current diagnostic criteria') additional medical exclusion criteria were also enforced. These included deafness, an ICD-10/DSM-IV diagnosis of childhood autism, English being a second language and children with known neurological disorders. The application of the exclusion criteria differed depending on the nature of the exclusionary condition and the family member affected (see table 2.1) (for example, if any family member was diagnosed as autistic, then the entire family was excluded. However, if a sibling had a non-verbal IQ of below 80, then only that individual was excluded). In the clinical (Guy's Hospital) sample those families with chromosome abnormalities, including fragile X, were also excluded by cytogenetic testing. A complete list of ascertainment and exclusion criteria are given in table 2.1

The majority of the families included in the genome screen consisted of four or five individuals, with just five larger families participating.

In the total sample collected, we found that 34.4% of the siblings could be classified as 'affected' under our ascertainment criteria (i.e. a single language score >1.5 Standard Deviations (SD) below that expected for their age). Assuming a population prevalence of 4% this yields a sibling risk ratio (λs) value of 8.6 within our families.

Proband Ascertainment

Language measures >1.5 standard deviations below the mean for their chronological age on the receptive or expressive portion of the CELF-R battery
WISC-III Performance IQ (PIQ) of above 80

Exclusion Criteria

Criterion	Action
Severe mental retardation (IQ<60)	Exclude whole family
Performance IQ<80	If proband - exclude whole family If sibling - exclude individual
Monozygotic twins	If further siblings are available, include only one twin of the pair.
Chronic illness requiring multiple hospital visits	Exclude whole family
Hearing loss or deafness	If proband - exclude whole family If sibling - exclude individual
Diagnosis of autism or pervasive developmental disorder	Exclude whole family
Bi- or multi-lingual background	Exclude whole family
Children under local authority care	Exclude whole family
Children with known neurological disorders	Exclude whole family
Children with known chromosomal abnormalities (Guys families only)	If proband - exclude whole family If sibling - exclude individual

Table 2.1 - Ascertainment Criteria for the SLI Genome Screen

2.1ii - The Guys Hospital Sample

The Guys sample represents a self-referred sample of children with persistent language problems needing special schooling and cannot be considered to be representative of a total population in the community. The cases were identified through three special schools for language disorders (John Horniman school, Moor House School and Meath School), paediatricians and speech therapists in the South Thames region and through Afasic, a support organisation for people with developmental and language impairments. Parental, and where appropriate child, permission was gained by a letter and ethical approval was given by the Guy's and St Thomas' Trust ethics committee 1996 (Ref. No 96/7/11).

The probands and all available siblings were assessed with the language and IQ battery (see section 2.2 'The genome screen phenotypes') either at school or at Guys Hospital outpatients. For some individuals, a historical score on the appropriate tests was already available, and no additional assessment was required. Blood samples were taken from the majority of children

and parents but for a few cases (2 individuals) blood was refused and buccal swabs were taken instead.

In total Guys referred 64 families, of whom 21 were later excluded (see Tables 2.2 and 2.3). Thus 43 families selected by Guys were included in the final genome screen sample. Within this sample, 52 children (37%) were attending a specialist language unit, special school or a mainstream school placement with a statement of special educational needs.

2.1iii - The Cambridge Language and Speech Project (CLASP) Sample

The Cambridge Language and Speech Project (CLASP) is an ongoing epidemiological study investigating speech and language difficulties in children. The parents of all children born between April and December 1989 within the Cambridgeshire area (2950 children in total) were contacted and asked to complete a questionnaire regarding their child's language and behavioural development (Burden et al 1996). The progress of those children considered to be 'at risk' for language difficulties was longitudinally followed alongside a 'language-normal' control group across three specified time points between the ages of three and eight. All children within the CLASP database who were felt to meet the genome screen criteria at age eight were referred into the genome screen. All available siblings were tested with the appropriate language and IQ tests (see section 2.2 'The genome screen phenotypes') and buccal swab samples were taken from all available parents and children. In total Cambridge referred 67 families, of whom 12 were later excluded (see Tables 2.2 and 2.3). Thus, 55 families from Cambridge were included in the genome screen sample. Informed consent was given by all families and ethical approval for the collection of families by the CLASP group was given by the Cambridge Local Research Ethics Committee (Ref. No. LREC96/212).

A description of the CLASP ascertainment procedure can be found in Stott, Merricks, Bolton and Goodyer (2002), and is detailed in figure 2.1.

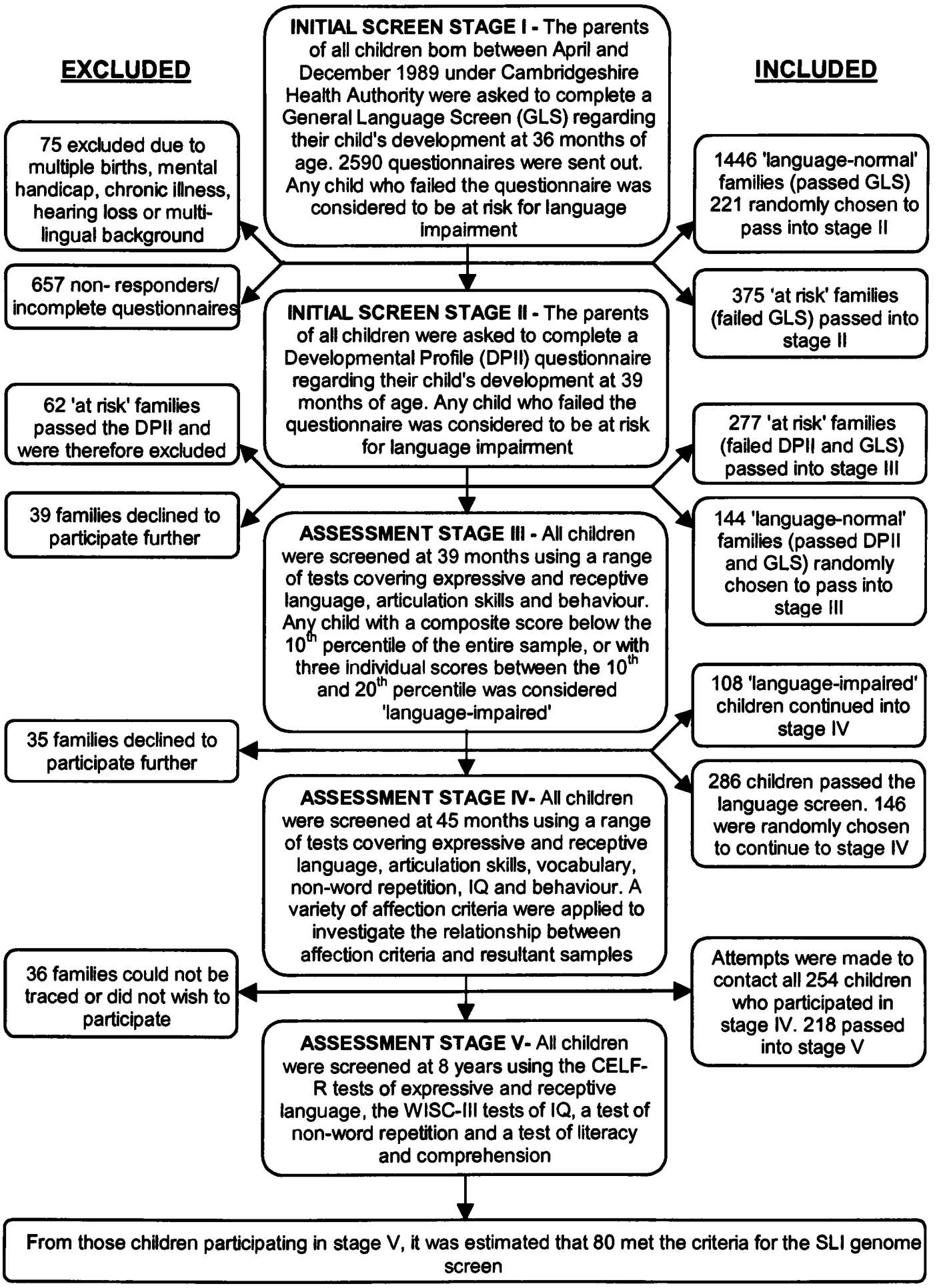


Figure 2.1 - The Cambridge Language and Speech Project (CLASP) - Study Outline

Number of Families Referred	Families		Sib-Pairs (independent)[a]		Sib-Pairs (all)[b]	
	Guys	Cam	Guys	Cam	Guys	Cam
Total Families Collected	64	67	109	99	164	136
Total Families Excluded (see Table 2.3)	21	12	37	17	58	23
Remaining Families (genome screen sample)	**43**	**55**	**72**	**82**	**106**	**113**

Table 2.2 – Numbers of Families and Sib-Pairs Referred by Guys Hospital and Cambridge : ***NOTE:*** *In a family with n sibs:*

***a** : The number of independent sib-pairs is given by (n-1)*

***b** : All sib-pairs comprises all possible pairings of sibs in a sib-ship. For families with more than two sibs this is given by [n(n-1)/2]*

Excluded Families and Individuals	Families		Sib-Pairs (independent)		Sib-Pairs (all)		Additional Individuals	
Reason for Exclusion	Guys	Cam	Guys	Cam	Guys	Cam	Guys	Cam
Autism	6	2	14	3	26	4	—	—
Non-paternity	2	1	3	2	4	3	—	—
No siblings available	4	9	5	12	6	16	—	—
Chromosome abnormality	3	—	4	—	5	—	3	—
DNA problems	4	—	7	—	10	—	—	—
PIQ<80	2	—	4	—	7	—	2	—
Too old to phenotype	—	—	—	—	—	—	3	—
Too young to phenotype	—	—	—	—	—	—	4	2
Singleton (no sibs)	—	—	—	—	—	—	—	14
Total	**21**	**12**	**37**	**17**	**58**	**23**	**12**	**16**

Table 2.3 – Numbers of Families and Sib-Pairs Excluded from the Guys Hospital and Cambridge Samples.

As may be expected, a greater number of families were excluded from the clinical (Guys Hospital) sample than from the less severely affected epidemiological (CLASP) sample. In general, families were excluded from the clinical sample because of chromosomal

abnormalities and autism spectrum disorders whereas those excluded from the epidemiological sample were due to collection problems and lack of availability of siblings.

2.2 THE GENOME SCREEN PHENOTYPES

As discussed in the introduction, three language measures were assessed for the genome screen: Expressive and receptive language skills were scored using the Clinical Evaluation of Language Fundamentals (CELF-R) and a test of non-word repetition was used as a marker of phonological short-term memory. No parental phenotype data was used as the linkage analysis utilises only information from sib-pair phenotype data.

Trained psychologists within the SLI Consortium (see appendix A) performed all psychometric testing and normalisation of scores.

2.2i - The Clinical Evaluation of Language Fundamentals - Revised (CELF-R)

The CELF-R is a clinical tool widely used for the identification, diagnosis and follow up evaluation of language disorders in school age children. The battery is split into receptive and expressive scales, each of which consists of three sub-tests designated to be primarily receptive or expressive in nature. The exact combination of individual tests used is dependent upon subject age (see table 2.4). Additive raw scores from each segment are transformed to derive a standardised Receptive Language Score (RLS) and an Expressive Language Score (ELS), which can be combined to derive a Total Language Score (TLS). Each score has a mean of 100 and a standard deviation of 15 in the general population calibration sample (Semel et al 1992). The genome screen used only the ELS and RLS scores as TLS merely represents a composite of these two factors.

Receptive tests range from simple comprehension tasks (e.g. 'point to the red line'), to tasks which require a complex interpretation of information presented orally (e.g. chose two correct answers from the following; 'a man is bigger than a house, a penny, a spoon or a plane?'). Expressive tests include sentence formation (e.g. 'make a sentence including the word shoe') and tasks which examine the subjects grasp of word structure (e.g. 'here is one dog, here are two...?'). There are eleven sub-tests in total (see table 2.4).

All available children between the ages of 5 and 17 were tested with the CELF-R battery.

Sub test	Description	Age of child
Receptive Language		**Age 5-7**
Linguistic concepts	Assesses comprehension of concepts relating to inclusion, exclusion, co-ordination, time, condition and quantity	
Sentence structure	Assesses comprehension of structural rules at the sentence level	
Oral directions	Assesses comprehension, recall and execution of oral commands of increasing length and complexity	
Receptive Language		**Age > 8**
Oral directions	As above	
Word classes	Assesses the ability to categorise words that are related (or unrelated) by semantic class, opposites, spatial or temporal features.	
Semantic relationships	Assesses interpretation of semantic relationships (temporal, spatial, passive and comparative) in spoken sentences	
Expressive Language		**Age 5-7**
Word structure	Assesses knowledge of word structure rules in an expressive task	
Formulating sentences	Assesses formulation of simple, compound and complex sentences	
Recalling sentences	Assesses recall and reproduction of surface structure as a function of syntactic complexity	
Expressive Language		**Age > 8**
Formulating sentences	As above	
Recalling sentences	As above	
Sentence assembly	Assesses the ability to assemble syntactic structures into grammatically acceptable and semantically meaningful sentences	

Table 2.4 - The CELF-R Sub-Tests : *For a full description of each sub-test please refer to appendix B. Two additional sub-tests (word association (expressive language) and listening to paragraphs (receptive language)) are provided as supplementary tests that can be used if a child is unable to complete any other task.*

2.2ii - Non-Word Repetition

The non-word repetition test (NWR) provides a measure of phonological short-term memory (see section 1.7i 'Short-term memory deficits in SLI') (Gathercole et al 1994) and involves the repetition of tape-recorded nonsensical words of increasing length and complexity (e.g. brufid, contramponist). Studies show that individuals with current language impairments, as well as those who experienced language difficulties in childhood which have since resolved, perform poorly on this test (Gathercole et al 1994; Bishop et al 1999a) (see section 1.7i 'Short-term memory deficits in SLI'). All available children between the ages of 7.5 and 18 were tested using the NWR test.

All Guy's Hospital subjects completed a published version of the children's test of non-word repetition (Gathercole et al 1994) (see appendix C). However, all Cambridge subjects were examined using a pre-publication revision of this test (see appendix D). Although both tasks are similar in administration, and some words are common to both tests, it was evident that the published standardisation introduced flooring effects, which resulted in an undesirable skewing of the distribution of scores. For this reason, and in order to allow combination of the NWR scores across groups, both versions of the NWR tests were administered to 111 subjects from both cohorts (aged between 4.8 and 53.6) and a between test regression calibration coefficient was determined. Raw scores correlated 0.89 ($p<0.001$) and were linearly related across the whole range, the relationship being the same for both adults and children. A linear regression calibration equation gave raw scores from the pre-published form of the test as 0.658 (SE = 0.009) times the raw score from the published test. Raw scores from the Guy's Hospital sample were therefore multiplied by this factor to make them comparable with the raw scores from the Cambridge sample. Standard scores for a British population were then obtained using norms extended for older children by Gathercole (pers comm). The NWR transformation was performed by Carol Stott, Melanie Merricks and Andrew Pickles (see appendix A).

2.2iii - Intelligence

IQ was assessed using the Wechsler Intelligence Scale for Children (WISC-III-UK) (Wechsler 1992). This is a battery of tests that yield measures of verbal and non-verbal (performance) intelligence quotients (IQ). The verbal scale comprises tests of comprehension, vocabulary and abstract reasoning whereas completion of the performance tasks relies primarily upon visual and constructional clues (e.g. mazes, symbol arrangement and abstract visual problem solving). The verbal (VIQ) and performance (PIQ) IQs can then be combined to give a full scale IQ (FIQ). The WISC-III requires no reading or writing of words.

Within the genome screen strategy, the WISC measure of PIQ was primarily used to allow the exclusion of those children felt to have an 'abnormal' non-verbal IQ (i.e. below 80), as it was felt that the language problems experienced by these individuals may be due to more general learning difficulties. In addition, PIQ was analysed as a quantitative trait in those areas of the

genome where suggestive linkage was found for a language measure. It was hoped that this approach would allow the identification of language specific loci as opposed to those influenced by more general cognitive factors. The WISC verbal and full IQs (VIQ and FIQ respectively) were not used for the genome screen except to exclude those individuals with severe mental retardation (see table 2.1).

2.2iv - Descriptive Statistics of Phenotypes

In total 252 children (153 males and 99 females), between the ages of 5 and 19 (mean age =9.4, SD=3.04) were assessed for CELF-R expressive and receptive language (ELS and RLS respectively), non-word repetition (NWR) and non-verbal IQ (PIQ) as described above. Table 2.5 provides the descriptive statistics for the genome screen cohort split both by group and by proband/sibling.

The design of the CLASP study meant that the Cambridge probands had a much narrower age range than those selected by Guys. However, the average age and the age range of the group as a whole (i.e. probands and siblings) were not significantly different between sites (Table 2.5).

The probands from both groups showed an excess of males to female (~2.5:1), which is consistent with the male predominance reported in previous studies (see section 1.1iii 'The prevalence of SLI'). Interestingly, this male excess was not reflected within the sibling groups, which were found to contain approximately equal numbers of boys and girls (see table 2.5).

	Age Range			Mean Age			Male:Female		
	All	Prob	Sibs	All	Prob	Sibs	All	Prob	Sibs
Total	5-19	5-16	5-19	9.46	8.56	10.04	1.6:1	2.5:1	1.1:1
CLASP	5-17	7-9	5-17	9.38	7.95	10.34	1.6:1	2.2:1	1.2:1
Guys	5-19	5-16	5-19	9.56	9.35	9.69	1.6:1	2.9:1	1.1:1

Table 2.5 – Age Range and Sex Distributions of the Total Genome Screen Sample and In the Proband and Sibling Groupings : *Prob=Probands*

Descriptive statistics for each genome screen phenotype within the total genome screen sample, the Guys and Cambridge cohorts, and the proband and sibling groupings can be

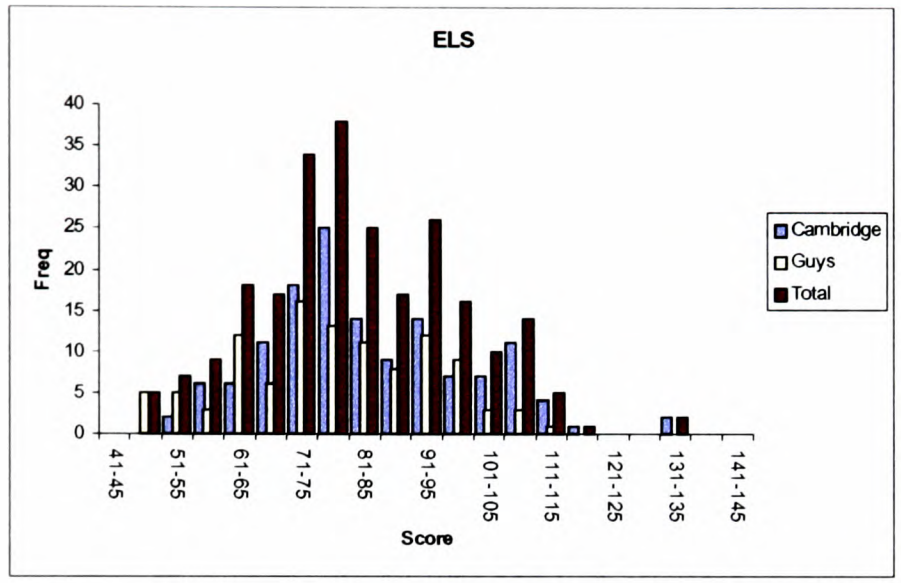
found in tables 2.6 and 2.7. Histograms showing the distribution of each of the phenotypes within these subsets can be found in figures 2.2 and 2.3.

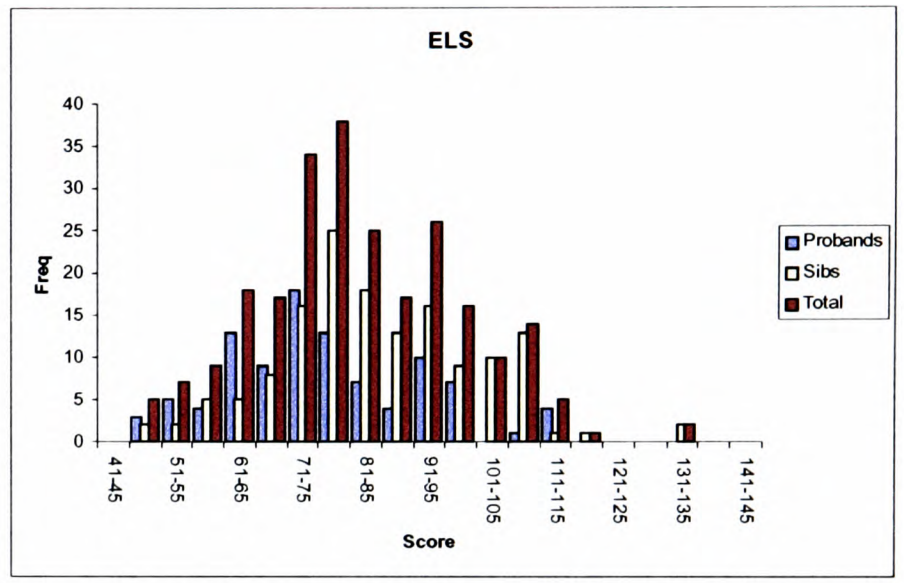
	Total				Cambridge			Guy's Hospital		
	ELS	RLS	NWR	PIQ	ELS	RLS	NWR	ELS	RLS	NWR
Mean	81.68	91.11	96.62	100.24	84.13	94.22	102.42	78.55	87.14	89.12
Median	80	91	99	97	82	93	105	78	87	93.5
Std Error	1.03	1.17	1.18	1.13	1.38	1.61	1.45	1.49	1.61	1.68
Std Deviation	16.05	18.23	18.33	17.39	16.13	18.85	16.97	15.46	16.66	17.34
Interquartile Range	21	27	25	24	22	25	21	24	25	26
Skewness	0.351	0.160	0.404	0.170	0.538	0.173	0.493	0.070	0.040	0.424
Kurtosis	-0.162	-0.104	-0.384	-0.418	-0.130	-0.174	-0.162	-0.649	-0.313	-0.749
Count	244	244	243	239	137	137	137	107	107	106
					Mann Whitney U test (p)			**0.023**	**0.005**	**0.000**

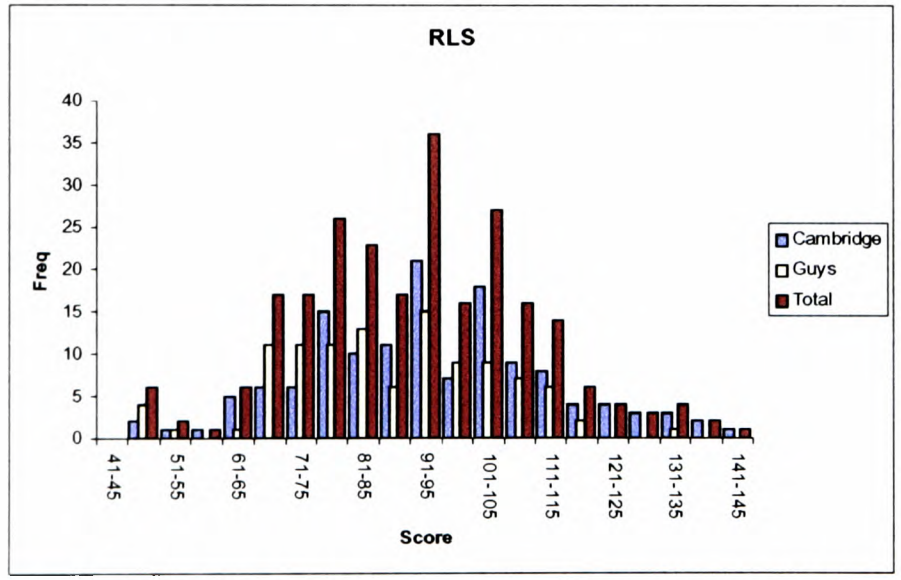
Table 2.6 - Descriptive Statistics for Each Genome Screen Phenotype Within the Total Genome Screen Set and the Constituent Guy's Hospital and Cambridge Groupings : *Note that the mean performance IQ is consistent with the population average. The Mann Whitney U test demonstrates that the mean of each phenotype varies significantly between the clinical (Guys Hospital) and epidemiological (Cambridge) groups. Note that in some cases only a selection of the phenotypes available were collected from all siblings. 8 siblings from Guy's Hospital were typed for CELF but not NWR, and another 9 were typed for NWR but not CELF. The samples given for Guy's Hospital therefore form overlapping subsets of the total sample available for genotyping.*

	Probands			Co-Sibs		
	ELS	RLS	NWR	ELS	RLS	NWR
Mean	76.44	86.22	91.86	85.08	94.29	99.62
Median	73	85	92	82	95	102
Std Deviation	15.41	17.62	20.69	15.59	17.97	16.02
Interquartile Range	22.5	23.5	31.75	20.5	22	20
Skewness	0.56	0.31	-0.02	0.28	0.07	-0.58
Kurtosis	-0.12	-0.28	-0.76	0.04	0.19	0.09
Count	98	98	98	146	146	145

Table 2.7 - Descriptive Statistics for Each Genome Screen Phenotype for the Total Genome Sample Split by Proband and Co-Sibs







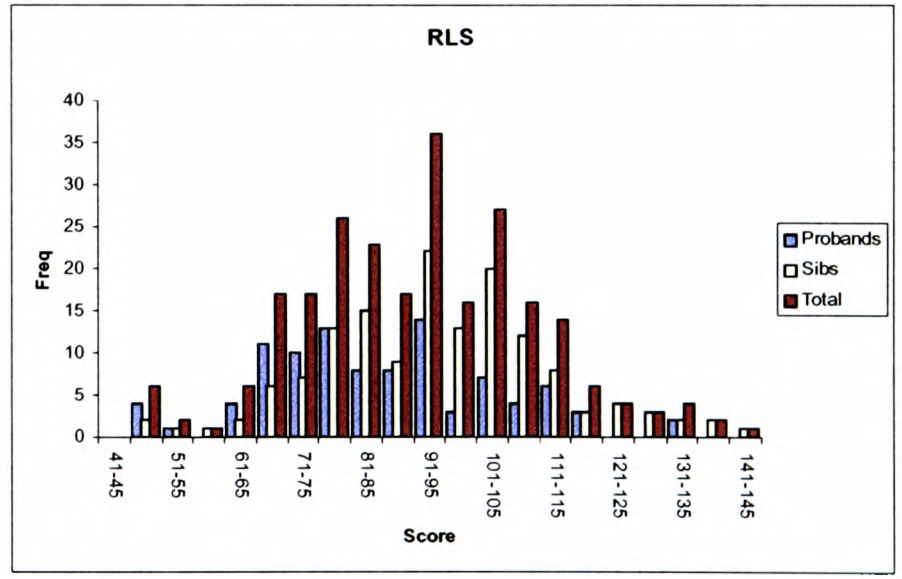
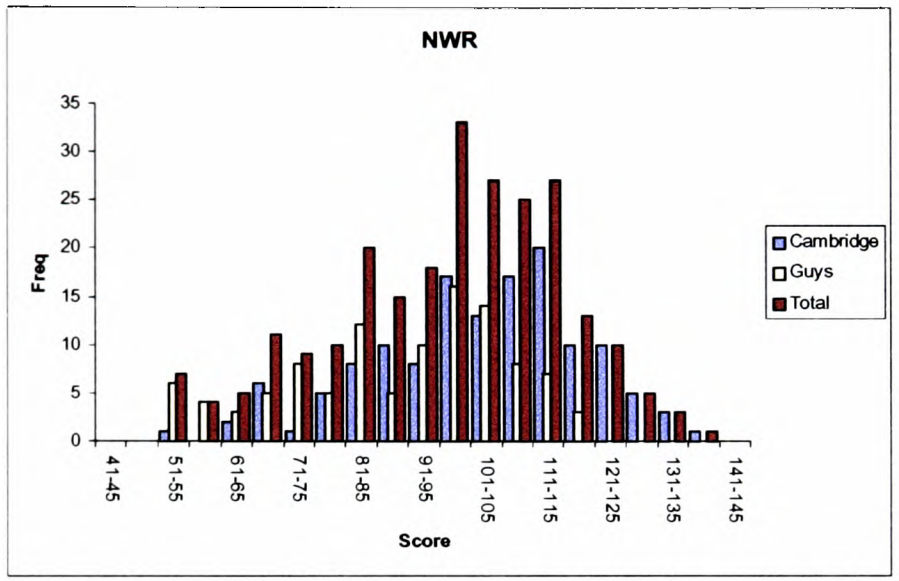
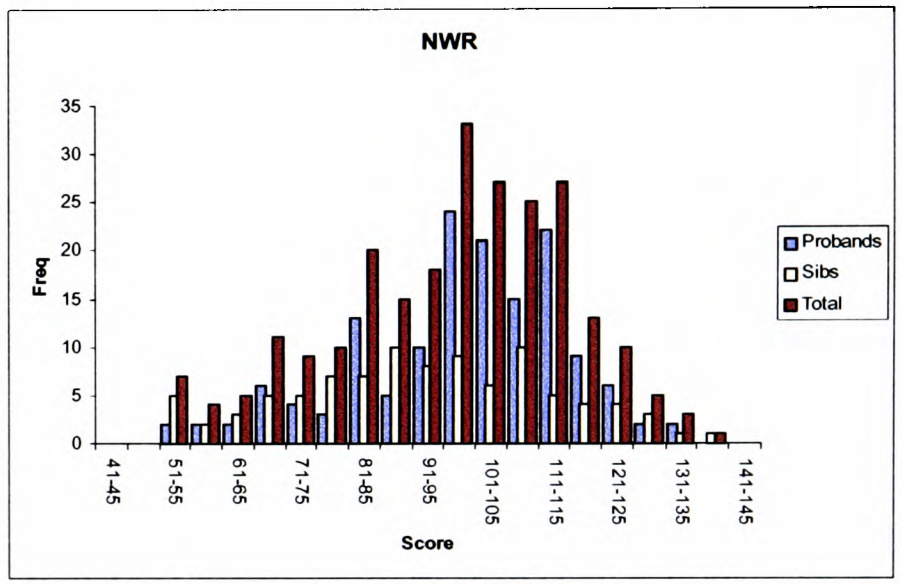
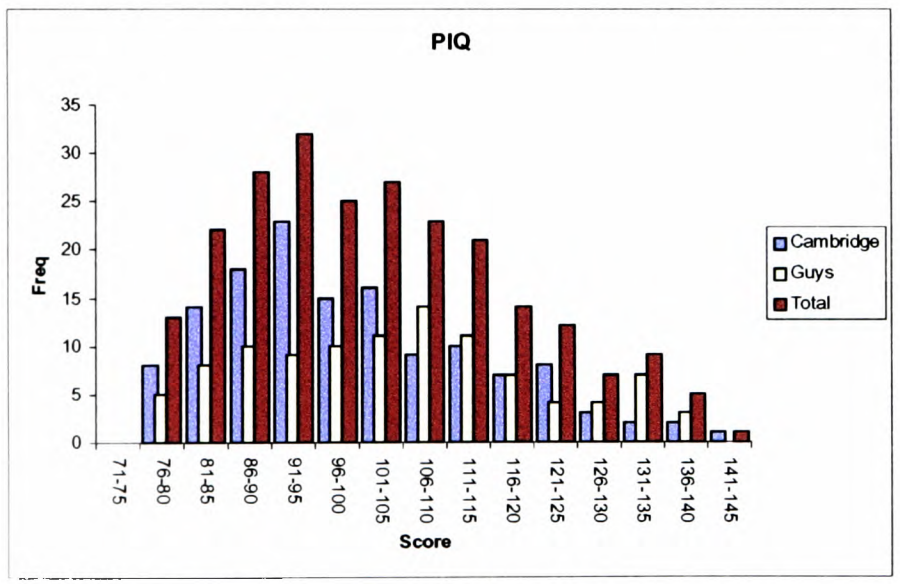


Figure 2.2 - CELF-R Expressive Language Scores (ELS) and Receptive Language Scores (RLS) in Guys, Cambridge, Proband and Sibling Cohorts







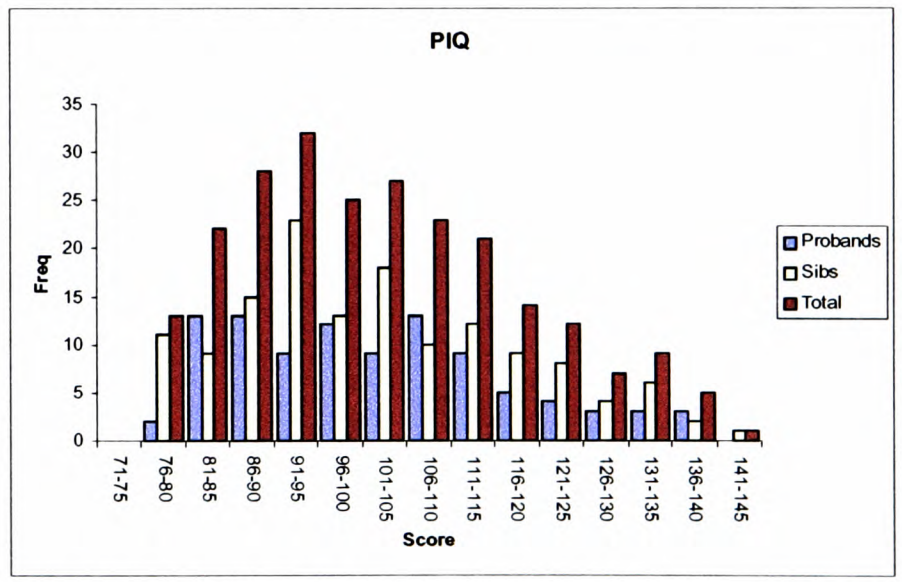


Figure 2.3 - Non-Word Repetition (NWR) and Performance IQ (PIQ) in Guys, Cambridge, Proband and Sibling Cohorts : *The NWR histograms shown here represent the scores after the Guys tests (published version) had been transformed onto that of the Cambridge test (pre-publication version) (see section 2.2ii 'Non-word repetition').*

It is apparent from the above data (Figures 2.2 and 2.3 and Table 2.6), that whilst the average level of performance IQ within the genome screen sample is consistent with that of the general population (i.e. mean of 100), the means of all language-based measurements fall below the expected mean of 100. This is a good indication that the sample selected for the genome screen represents a collection of children whose developmental problems are largely language specific.

Comparisons between proband and co-sib groupings (Table 2.7) indicated that the probands generally attained lower language ratings than their complementary co-sibs. However, whilst the co-sib language scores showed some regression towards the population mean (i.e. 100) they all remained below that expected (Table 2.7). This is attributed to the high number of siblings (~34%) who also displayed signs of language impairments.

2.2v - Data Transformation

The data in table 2.6 demonstrates that, as expected, the Guys group performed at a significantly lower language level than the CLASP group (see Figures 2.2 and 2.3 and Table 2.6). Thus, although both drawn from the general population of SLI children, the two groups differ significantly ($p<0.05$ for RLS, ELS and NWR) in the magnitude of severity of their disorders. This is attributable to the fact that, although the diagnostic criteria applied to both samples were identical, Guy's Hospital represent a clinical, severely affected sample, whereas Cambridge represent a more mainstream epidemiologically selected sample. In order to combine the two cohorts for variance components analysis, which creates a model around a single mean, all phenotypes were standardised to a z score ($z=(x-\mu)/\sigma$, where μ and σ are taken from each separate group). Conversion of the language scores in this manner produces a distribution with a single mean of 0 whilst preserving the variances of the original samples, and thus allows a single analysis of the two groupings within the variance components model.

The standardised scores are hereafter referred to as 'RLStrans', 'ELStrans' and 'NWRtrans' and were used for the combined analysis of both groups for the genome screen.

Table 2.8 presents the correlations between, and familialities for, each of the phenotypes used in the genome screen analysis. The CELF expressive and receptive language scores

showed a high level of correlation with each other and a moderate level of correlation with the non-word repetition scores. Some correlation was found between the CELF scores and non-verbal IQ, whereas the NWR trait showed only a low level of correlation with PIQ (Table 2.8). In general, only a low level of correlation was observed between the language scores of probands and their siblings (Table 2.9). The familality of each language phenotype was estimated within a variance components system in the SOLAR package (Almasy & Blanjero 1998). This program derives a wide estimate of heritability by dividing the variance of the genotype by the variance of the phenotype (V_G/V_P). No account is made for common environmental effects. All three language measures showed significant levels of familiality within the genome screen sample (Table 2.8).

	Total			
	ELS trans	RLS trans	NWR trans	PIQ trans
ELStrans	1.000			
RLStrans	0.746	1.000		
NWRtrans	0.538	0.439	1.000	
PIQtrans	0.396	0.456	0.131	1.000
Familiality	0.725	0.704	0.375	0.582
p	$<1x10^{-7}$	$4x10^{-7}$	$6.6x10^{-3}$	$8.2x10^{-6}$

	Cambridge			
	ELS	RLS	NWR	PIQ
ELS	1.000			
RLS	0.759	1.000		
NWR	0.452	0.436	1.000	
PIQ	0.506	0.566	0.228	1.000

	Guy's Hospital			
	ELS	RLS	NWR	PIQ
ELS	1.000			
RLS	0.706	1.000		
NWR	0.597	0.335	1.000	
PIQ	0.352	0.418	0.171	1.000

Table 2.8 - Correlations and Familialities of Phenotypes Within the Total Genome Screen Set and the Constituent Guy's Hospital and Cambridge Groupings

	Trait			
	ELS trans	RLS trans	NWR trans	PIQ trans
Total	0.396	0.418	0.227	0.323
	ELS	RLS	NWR	PIQ
Cambridge	0.596	0.551	0.289	0.533
Guys Hospital	0.177	0.278	0.154	0.075

Table 2.9 - Proband-Sibling Correlations of Phenotypes Within the Total Genome Screen Set and the Constituent Guy's Hospital and Cambridge Groupings : *For families with more than two children the sibling who had complete set of data and was closest in age to the proband was selected to calculate the proband-co-sib correlations*

2.3 DNA COLLECTION AND EXTRACTION

Whole blood (Guys group) or buccal smear samples (Cambridge group and two non-compliant individuals from the Guys group) were collected from probands and all available siblings and parents, regardless of language ability. All collected samples were accompanied by a deposit form that provided details of the individual and their family (see appendix E).

2.3i - DNA Extraction from Whole Blood

10mls of whole blood was collected in sodium EthyleneDiamineTetraAcetetae (EDTA) tubes (to prevent coagulation) and stored at room temperature (if DNA was to be extracted within 24 hours) or -20°c (if DNA was to be extracted at a later date).

Genomic DNA was extracted using commercial Nucleon kits in accordance with the manufacturer's protocol. The extraction process involves the purification of white blood cells from solution and the extraction of DNA from these cells. The red blood cells are broken down by reagent A, allowing the intact white blood cells to be extracted from solution by centrifugation (Hettich Rotana 46R). The white blood cells are then lysed by a second reagent (reagent B) and the cellular proteins are digested (sodium perchlorate) and removed from solution by a chloroform extraction. The DNA is then precipitated by ethanol extraction. Ethanol acts to remove the residual organic solvents from solution and also de-salts the DNA.

1) Transfer the blood to a 50ml Falcon tube
2) Add 35-40ml reagent A
3) Centrifuge at 2000rpm for 4 minutes *(Reagent A lyses the red blood cells allowing them to remain in solution whilst white blood cells (which contain the DNA) form a pellet during centrifugation)*
4) Discard supernatant *(i.e. red blood cells)*
5) To the pellet add 10ml reagent A and resuspend pellet by gentle shaking *(This second wash in reagent A just removes any remaining red cells from the pellet)*
6) Centrifuge at 2000rpnm for 4 minutes
7) Discard supernatant *(i.e. red blood cells)*
8) To the pellet add 2ml reagent B *(Reagent B lyses the white blood cells and thus releases the DNA into solution)*

9) Resuspend pellet by vortexing

10) To cell suspension add 500µl sodium perchlorate *(Sodium perchlorate digests the proteins from solution)*

11) Mix by inverting tube at least 7 times by hand

12) Add 2ml cold chloroform *(Chloroform extracts the cell lysates from solution)*

13) Mix by inverting tube at least 7 times by hand

14) Add 300µl Nucleon resin *(The nucleon resin forms a visible layer between the chloroform and the DNA solution)*

15) Centrifuge 2000rpm for 3 minutes

16) Transfer the upper layer to a fresh 15ml Falcon tube *(i.e. the aqueous DNA layer)*

17) Centrifuge 2000rpm for 3 minutes

18) Transfer the upper layer to a fresh 15ml Falcon tube

19) Add 2 volumes (~8-10ml) cold 100% ethanol *(The ethanol precipitates the DNA which forms a visible white pellet)*

20) Invert tube several times until DNA is precipitated

21) Fish DNA out from ethanol using a sterile innoculating loop and wash in a 1.7ml microcentrifuge tube containing 1ml 70% ethanol *(The 70% ethanol steps act as a final wash to purify the DNA in solution)*

22) Store -20°c for at least one hour

23) Centrifuge DNA in 70% ethanol at 13000rpm for 20 minutes

24) Pipette off supernatant and resuspend DNA pellet in fresh 1ml 70% ethanol

25) Centrifuge at 13000rpm for 15 minutes

26) Pipette off supernatant and allow DNA pellet to dry at room temperature

27) Resuspend DNA pellet in 400µl 1xTrisEDTA (TE) and leave mixing at room temperature overnight *(TE buffer acts to stabilise the DNA in solution. DNA can be stored in TE (pH8.0) for several years at -20°c)*

Once extracted, 2µl of the DNA solution was diluted in 98µl of distilled water and the Optical Density (OD) of the resultant solution was measured on a UV spectrophotometer (Beckman) at 260nmn (OD_{260}). The concentration of the DNA was estimated by multiplying OD_{260} by 50^2

(dilution factor x conversion factor). 400µl of 'stock DNA' was made up at 80ng/µl and stored at -80°c. The concentration of this stock DNA was confirmed by running 1µl on 2% agarose against 50ng, 100ng and 250ng of a *λHind*III control marker (see section 2.5i 'Agarose gels'). A further 'working stock' was then prepared at 8ng/µl and stored at -20°c.

The yield of DNA from whole blood can vary greatly both between individuals and collections. However, in general 10ml of whole blood can be expected to generate between 200-300µg of DNA.

2.3ii - DNA Extraction from Buccal Swabs

The buccal swab kits were prepared at the Wellcome Trust Centre and mailed to each group prior to sample collection. Each kit consisted of 10 sterile cotton wool buds and a separate 50ml Falcon tube containing 5ml of transport buffer. All individuals were instructed to rub each cotton wool bud firmly around the inside of the mouth (between the cheek and gum) for 20 seconds. This was repeated for all 10 sticks. When used each bud was placed swab end down into the tube containing transport buffer. The buccal swabs were returned to the centre by special post and kept at 4°c until prepared.

DNA extraction from buccal swabs is similar to that of bloods. The buccal sample is de-proteinised by proteinase K; a highly active protease which is purified from the mould *Tritirachium album*. This enzyme has an optimal activity at 65°c and is augmented by the presence of Tris, EDTA and Sodium Dodecyl Sulphate (SDS), all of which are present in the transport buffer. The proteins are digested in solution, and the buccal cells lysed by Guanadine Hydrochloride, a strong denaturing agent that is particularly active in the presence of ammonium actetate. The cellular components are then removed from solution by a chloroform extraction and the DNA purified by ethanol precipitation.

1) Add 100µl of 10mg/ml proteinase K to each buccal swab tube *(Proteinase K de-proteinises the sample)*
2) Spin down briefly
3) Incubate at 65°c for 2-3 hours, or 37°c overnight in a shaking incubator

4) Add 1ml Guanadine Hydrochloride (6M) and 0.5ml Ammonium Acetate (7.5M) *(Guanadine HCl lyses the buccal cells)*

5) Spin down briefly

6) Incubate at 60°c for 1 hour in a shaking incubator

7) Put the cotton wool buds inside two 5ml syringes (5 buds each)

8) Place the syringes into 15ml Falcon tubes

9) Spin at 1000rpm for 5 minutes *(removes solution from cotton wool buds)*

10) Transfer all the liquid (i.e. from the cotton wool buds and from the original tube) into a 15ml Falcon containing 2ml cold chloroform (*Chloroform extracts the cell lysates from solution)*

11) Shake for 10 minutes on a rocker

12) Leave to stand for 1 minute

13) Add 300µl Nucleon silica suspension *(The nucleon resin forms a visible layer between the chloroform and the DNA solution)*

14) Spin at 2500rpm for 5 minutes

15) Transfer the top layer into a new 15ml Falcon *(i.e. the aqueous DNA layer)*

16) Spin 2500rpm for 5 minutes

17) Transfer the top layer into a new 15ml Falcon *(i.e. the aqueous DNA layer)*

18) Add 2 volumes (~10ml) cold 100% ethanol *(The ethanol precipitates the DNA but the pellet is not visible because the DNA yield is low)*

19) Leave at -20°c overnight

20) Spin at 3000rpm for 30 minutes *(pellets the DNA)*

21) Wash pellet twice in 70% ethanol at 3000rpm, 15 minutes *(The 70% ethanol steps act as a final wash to purify the DNA in solution)*

22) Air dry

23) Resuspend in 300µl 0.1xTE

The DNA extracted from buccal swabs is not as pure as that extracted from blood and the optical density of the solution cannot be used to accurately predict the DNA concentration. Thus the concentration of all swab-derived DNAs was estimated by running 1µl of the DNA

solution on 2% agarose against varying concentrations (50, 100 and 250ng) of a λ*Hind*III control marker (see section 2.5i 'Agarose gels'). Working stocks of 8ng/µl were prepared and stored at -20°c. All other stocks were kept at -80°c.

In general, 10 mouthswabs can be expected to yield between 15-25µg of DNA.

2.3iii - Preamplification Extension Protocol (PEP)

In order to acquire enough DNA from the buccal smears to complete a genome screen it was necessary to pre-amplify the DNA derived from all cheek-swab samples. This was achieved using a Preamplification Extension Protocol (PEP) (Zhang et al 1992). The PEP technique involves the random amplification of genomic DNA (using a pool of random 15-mer primers) and can result in a 50- to 100-fold increase in template DNA for subsequent microsatellite amplification. Prior to the genome screen, the amplification of PEP DNA was verified within a series of 27 controls across 20 primers. All controls showed comparable amplification for both genomic and PEP DNAs and no evidence of preferential pre-amplification of specific alleles were seen.

All PEP amplifications were performed in a 100µl reaction in 0.2ml microcentrifuge tubes. The reaction mix was covered in 30µl of mineral oil and the PEP performed on PTC-225 thermocyclers (MJ Research).

Reaction mix

Reagent	Amount
DNA (8ng/µl)	5µl
PEP Primer (random 15mer) (200µM)	20µl
10X buffer (KCl)	10µl
dNTPs (8mM)	2.5µl
$MgCl_2$ (25mM)	4µl
Taq Gold	0.5µl
MilliQ H_2O	To 100µl

PCR program

Block temp	Duration	
95°c	18 minutes	
95°c	1 minute	X49 cycles
37°c	2 minutes	
Ramp to 55°c	@0.1°c/sec	
55°c	4 minutes	
72°c	5 minutes	
15°c	**HOLD**	
END		

Following the PEP all DNAs were extracted from the reaction mix by ethanol precipitation (1x 100% ethanol wash, followed by 2x 70% ethanol washes - see section 2.3i 'DNA extraction from whole blood'). The DNA was then resuspended in 300µl of distilled water.

Because of the random nature of the PEP reaction, it is not possible to visualise PEPed DNA as a discrete band on an agarose gel. In order to quantify the success of the PEP reaction it was therefore necessary to carry out a titration. Serial dilutions (1 in 2, 1 in 5 and 1 in 10) were made for each PEP DNA and used to amplify five different microsatellite markers. All products were run on 2% agarose and compared to the products obtained from three genomic DNA preparations at 8ng/µl. Each PEP reaction was diluted to approximate a 8ng/µl 'working stock' and stored at -20°c. Note that the DNA derived from a PEP reaction is generally less stable than that attained from a straightforward DNA extraction. Thus PEP DNA is particularly prone to degradation if kept at 4°c for any length of time and freeze/thawing cycles should be kept to a minimum.

2.4 THE POLYMERASE CHAIN REACTION

2.4i - The Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) is an enzymatic procedure used for the amplification of specific target regions of DNA, first described by Mullis in 1986.

PCR is performed as a cyclic reaction that involves three discrete steps of denaturation, hybridisation and extension (see figure 2.4 'The Polymerase Chain Reaction'). The double stranded template DNA is denatured to form single stranded DNA and thus allowing the hybridisation of 2 short (~20 nucleotides) primers designed to flank the region of interest. Each primer anneals on opposite DNA strands and is extended in a 5' to 3' reaction by a heat resistant enzyme known as *Taq* polymerase. *Taq* polymerase is a DNA polymerase isolated from the bacteria *Thermophilus acquaticus* found in thermal springs. This polymerase has an optimal working temperature of 72°c and is therefore thermostable throughout the entire PCR reaction. All PCR reactions in this project used Ampli*Taq* Gold DNA polymerase. This is a genetically modified form of the *Taq* polymerase that is isolated from *Escherichia coli* (*E. coli*) and is stable in solution until activated by a high temperature incubation step.

If 100% efficient PCR can result in an exponential amplification of target regions. Thus in a reaction of 35 cycles 2^{35} DNA molecules may be created from every target sequence in the original solution.

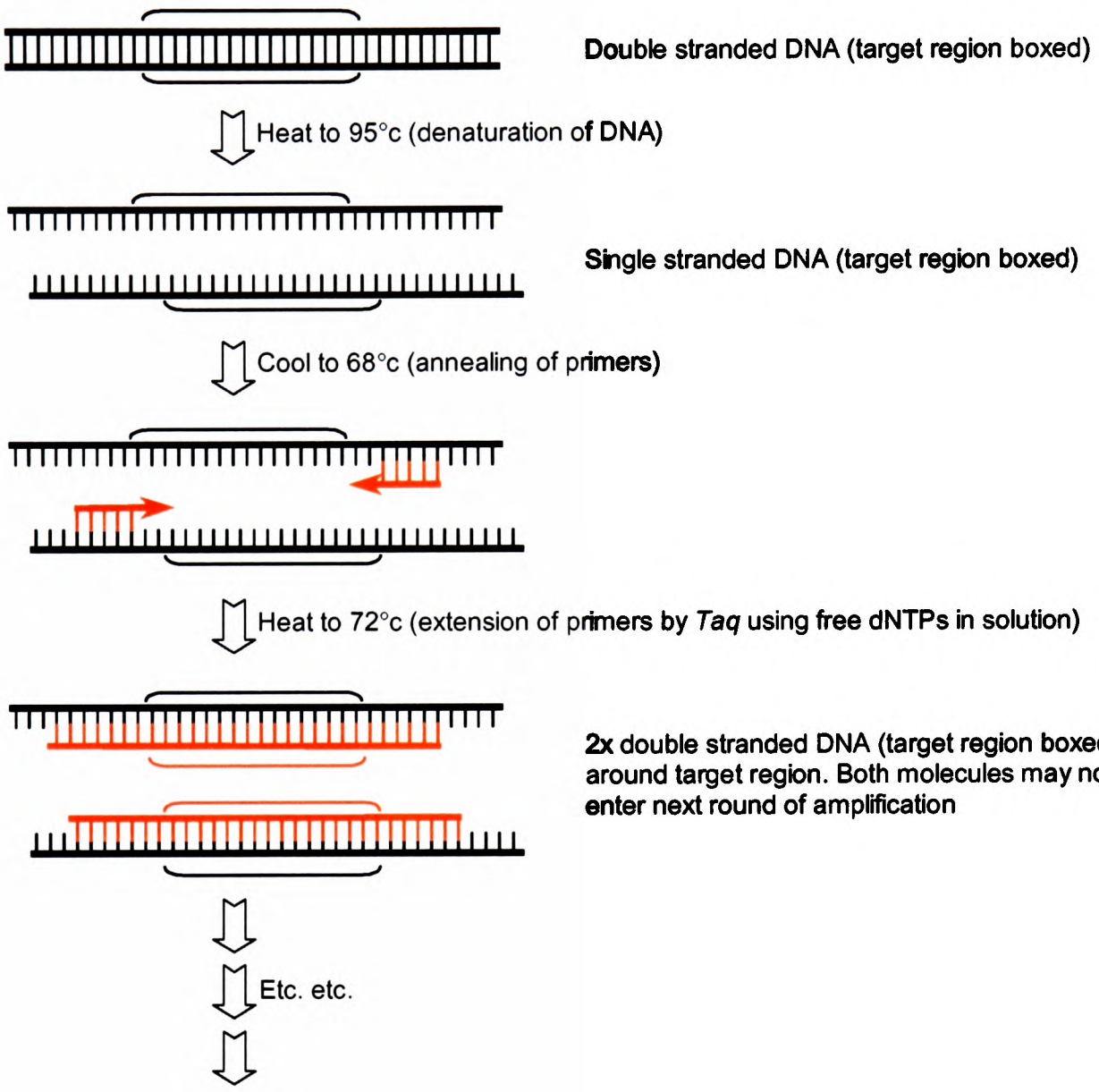


Figure 2.4 - The Polymerase Chain Reaction (PCR)

2.4ii - PCR Conditions and Programs

In general, PCR amplifications were performed in a 15µl reaction using one of two alternate PCR programs (as detailed below). The touchdown protocol was used for primers which failed under the C38 program and employs an increased annealing temperature to aid primer annealing in the early cycles, followed by a lower temperature to ensure efficient amplification in the later cycles.

PCR reaction mix

Reagent	Amount	Comments
DNA (8ng/µl)	5µl	Template DNA
F Primer (10µM)	0.3µl	3'-5' Template specific primers
R primer (10µM)	0.3µl	5'-3' Template specific primers
10X buffer (KCl)	1.5µl	PCR buffer
dNTPs (8mM)	0.4µl	Free dNTPs for primer extension
$MgCl_2$ (25mM)	0.6µl, 1.2µl or 1.8µl	The $MgCl_2$ concentration is variable (see below)
Taq gold	0.08µl	Polymerase for primer extension
MilliQ H_2O	To 15µl	

C38 PCR program

Block temp	Duration		Comments
95°c	18 minutes		Activation period for *Taq* Gold enzyme
95°c	1 second	X38 cycles	Denaturation of DNA
Ta°c	20 seconds		Annealing between primers and DNA
72°c	5 seconds		Extension of primers
72°c	1 minute		Final Extension
15°c	**HOLD**		
END			

Touchdown PCR program

Block temp	Duration	
95°c	18 minutes	
95°c	1 second	X9 cycles
Ta°c	20 seconds	
Ramp to Ta-5°c	@ -0.5°c/cycle	
72°c	5 seconds	
95°c	1 second	X27 cycles
Ta°c -5°c	20 seconds	
72°c	5 seconds	
72°c	1 minute	
15°c	**HOLD**	
END		

Each PCR cycle takes approximately 3 minutes. The exact length and temperature of each stage is specific to the primers, as is the final concentration of $MgCl_2$ in the reaction mix. The Mg^{2+} ions act to aid primer annealing and are essential to dNTP incorporation. Increasing the $MgCl_2$ concentration can act to increase the product yield for some fragments. However, this can also result in a loss of specificity and therefore to the amplification of undesired products. The $MgCl_2$ concentration is therefore optimised according to the primer pair used for each reaction.

The initial incubation at 95°c for 10 minutes acts to activate the *Taq* Gold enzyme. This step activates approximately 40% of the polymerase in the reaction, which is sufficient to ensure

efficient amplification during the early PCR cycles. The additional denaturation stages of the later cycles activate additional enzyme with each cycle of the PCR reaction.

2.4iii - PCR of the Genome Screen Markers

Working stocks of 8ng/µl DNA were stored in 96 well deep plates to allow semi-automated plating using a Hydra96 automatic microdispenser (Robbins Scientific). 5µl of DNA was plated out into 96 well plates and covered in 30µl of mineral oil. The PCR reaction mix (see section 2.4ii 'PCR conditions and programs') was added on top of the oil and spun through the layer by centrifugation. Unless stated otherwise, all PCR reactions were carried out on PTC-225 thermocyclers without heated lids.

All individuals were genotyped for 400 highly polymorphic dinucleotide repeat microsatellite markers, taken from the ABI PRISM LMS2-MD10 panels (version 2). This is a commercially available set of markers that provide a 10cM coverage of the entire genome. The primers are arranged into 28 panels, each of which covers one or two chromosomes. Each primer pair is fluorescently labelled with either a 6-FAM, HEX or NED phosphoramidite (Applied Biosystems (ABI)) allowing the post-PCR pooling of panels of PCR products (see section 2.5ii 'Polyacrylamide gels'). All of the ABI PRISM markers have been designed to produce high quality PCR products with a uniform tail length. Tailing is caused by the addition of non-templated residues (most commonly A) by *Taq* polymerase onto the 3' end of PCR products. Often this polyA addition is incomplete and the length of the tail can vary with magnesium chloride concentration, the length of extension stage, and the sequence of the reverse primer, leading to inconsistencies in product length. All ABI Prism reverse markers have a seven nucleotide modification that promotes consistent adenylation by *Taq*, and therefore acts to produce PCR products of a uniform length (Brownstein et al 1996).

For a list of markers excluded from the LMS2-MD10 please see appendix F.

2.5 AGAROSE AND POLYACRYLAMIDE GELS

2.5i - Agarose Gels

Agarose gels are used to separate DNA molecules according to their molecular weight. The agarose forms a matrix of polymeric molecules into which DNA can be loaded (when

combined with a loading buffer to increase density). The gel is subjected to a electrical field through which negatively charged DNA is pulled towards the anode. The distance of DNA movement is given by the equation

$$\frac{\text{(C-log(size in Kbp))}}{\text{M}}$$

Where C and M are dependent on variables such as temperature, pH, gel concentration etc and thus are given a constant for any given gel.

Hence for any single gel, DNA migration is a function only of the molecular weight of the DNA, with larger fragments remaining closer to the wells and smaller fragments migrating towards the anode.

The addition of ethidium bromide to the agarose solution allows the visualisation of DNA bands under UV light. Ethidium bromide intercalates with the DNA, between adjacent base pairs, and when in this state, fluoresces under UV light.

Agarose gels are used to separate DNA molecules of between a few hundred and a few thousand base pairs in size. Smaller fragments are best isolated on gels that contain a low percentage of agarose. Fragments that differ in size by less than fifty base pairs can be distinguished on polyacrylamide gels.

Because factors such as gel concentration and pH can affect the motility of DNA fragments, a size standard must be loaded onto all gels. Size standards tend to be created from phage genomes that have been digested with restriction enzymes to produce fragments of known sizes (e.g. λ*Hind*III is derived from a *Hind*III digestion of the λ phage genome to produce fragments of 23130, 9416, 6557, 4361, 2322, 2027, 564 and 125bp – see figure 2.5). Most fragments in this project were run against a 1Kbp ladder that produces a wide range of fragments between 100bp and 1000bp (Figure 2.5).

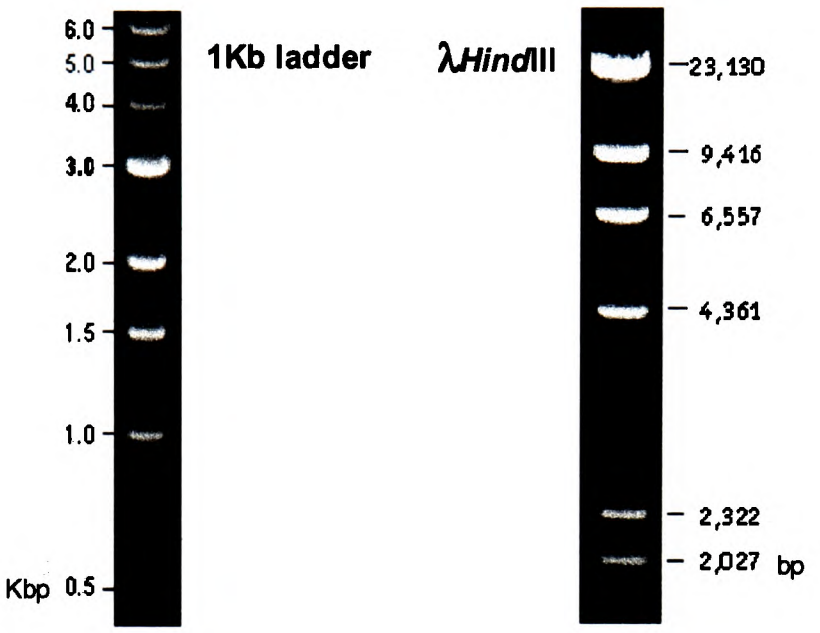


Figure 2.5 - 1Kb Ladder and λ*Hind*III Standards

The integrity of all PCR reactions was verified by running 5µl of the product on a 2% agarose gel against 0.5µg of 1kbp ladder. Gels were produced by dissolving 8g of agarose into 400ml of 1xTris-Borate EDTA (TBE), and adding 15µl of ethidium bromide. Each gel was run at 100mV for 30 minutes and the products visualised using a UV transilluminator (Alpha Innotech Corporation). The size and specificity of each band was checked against that expected and only those products which were clean and strong were used for subsequent genotyping analysis.

2.5ii - Polyacrylamide Gels

Following PCR all products from a given panel were pooled and separated on a polyacrylamide gel by an ABI 373 or an ABI 377 platform. Both of these sequencers use a polyacrylamide gel system that allows the separation of multiple DNA fragments of similar sizes. Pooled DNA fragments from the same individual are loaded into a single lane of the polyacrylamide gel and separated by an electrical field, in a similar manner to that described for agarose gels (see section 2.5i 'Agarose gels'). As the products run through the gel, a laser is used to excite the fluorescent tags (attached to the primers - see section 2.4iii 'PCR of the genome screen markers') and a Charge-Coupled Device (CCD) camera detects the fluorescence given off by each of these tags. The NED, HEX, FAM and ROX dyes all emit

light at a different wavelength and thus can be readily distinguished by the CCD camera (see figure 2.6). The excitation patterns are transformed by a computer to a gel image (see figure 2.7), where ROX appears as red, NED as yellow, FAM as blue and HEX as green.

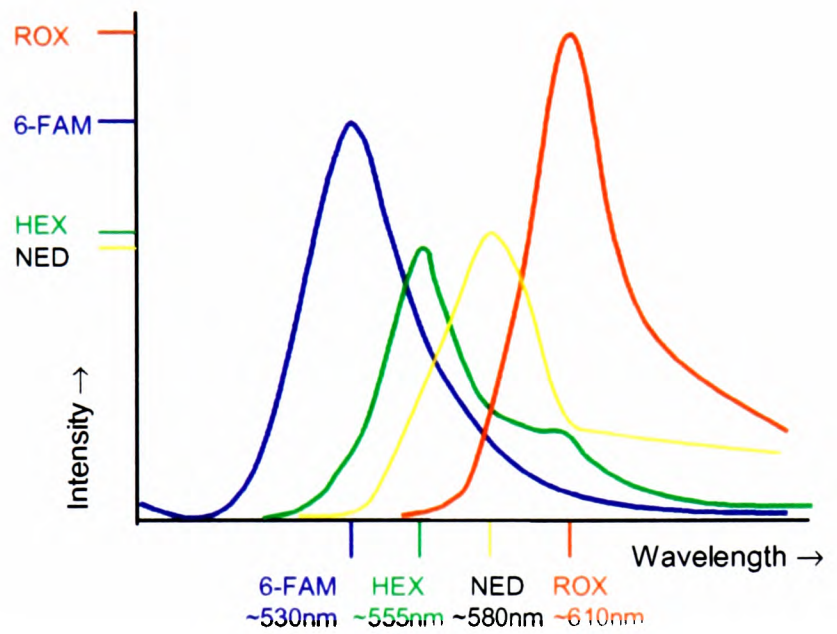


Figure 2.6 - Fluorescence Wavelength and Excitation Efficiency of Dyes : *The differences in wavelength of each fluorochrome mean that the CCD camera can easily detect the differences between each dye. The differences in the excitation efficiency (intensity) of each dye mean that it is necessary to pool the PCR products in different ratios according to the dye attached to the primers used (see below)*

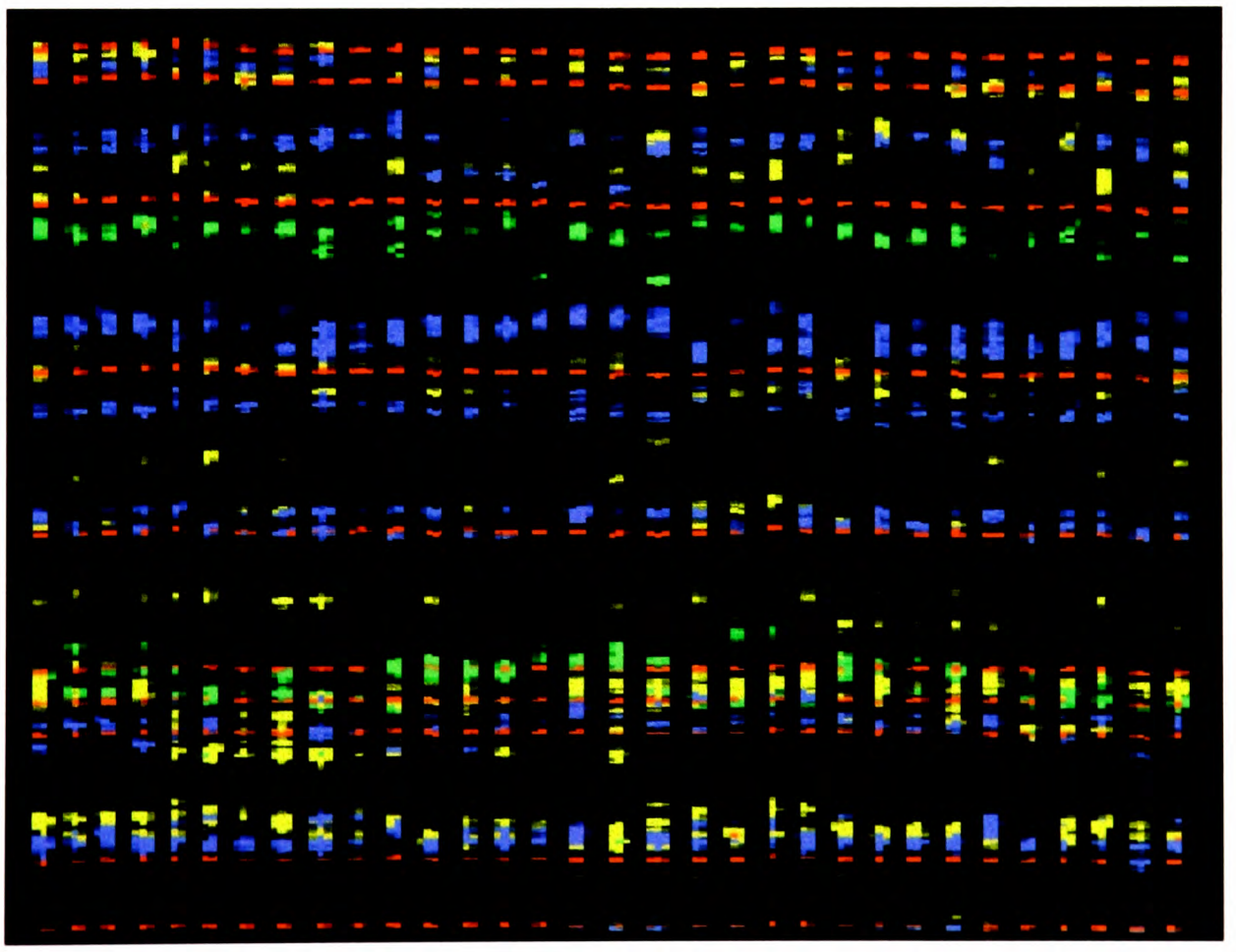


Figure 2.7 - Genotyping Gel : *Each lane on this gel represents 13 PCR products for a single individual. There are 2 HEX markers (green), 6 NED markers (yellow) and 5 FAM markers (blue). The red markers represent the ROX size standard at 35, 50, 100, 139, 150, 200, 250, 300, 340 and 350 base pairs. Note that two given markers may overlap in size range, but can still be distinguished because of the different colour tags.*

Separate FAM, HEX and NED PCR pools were made for each panel using the Hydra96 automatic microdispenser. The ratio of each product in these pools was dependent upon the strength of the PCR product, as estimated from agarose gels.

Each pool was mixed to form a 'master pool' consisting of a 4FAM:8NED:10HEX ratio. This ratio acts to preserve a constant level of fluorescence in the final pool and is necessitated by the differences in the levels of excitation efficiency between fluorescent dyes (see figure 2.6).

15-20µl of master pool was purified by ethanol precipitation (see section 2.3i 'DNA extraction from whole blood'), and the pellets resuspended in 2µl of formamide loading dye with 0.5µl of the size standard dye, ROX. Each sample was denatured at 95°c for 3 minutes and loaded onto a 1% polyacrylamide gel and run on an ABI 373 or ABI 377 platform at 2400v for 4 hours.

2.6 GENOTYPING

2.6i - Data Extraction

Data was extracted from gels using the Genescan software (version 3.1) and passed into the Genotyper program (version 2.0) (ABI). This is a specialised program that allows the visualisation of the genotype data and the automatic sizing of allele data (Reed et al 1994) (Figure 2.8).

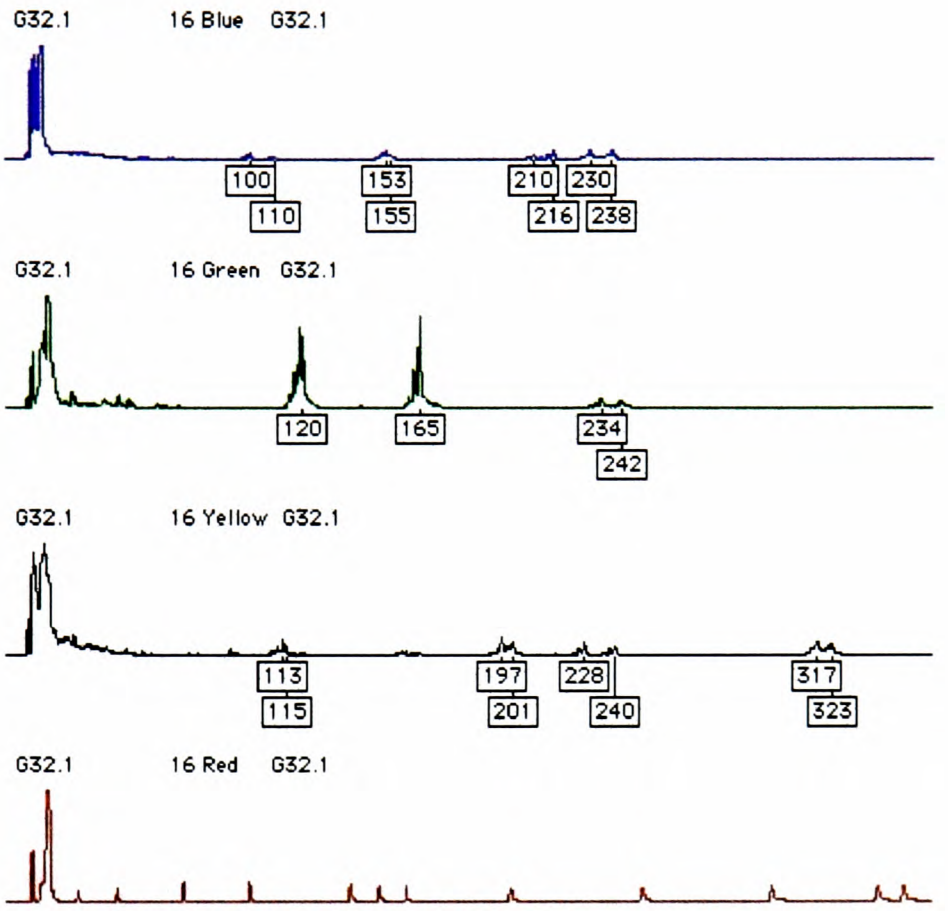


Figure 2.8a - Allele Calling Within Genotyper : *The above figure shows the entire genotyper plot for a panel of 11 markers for a single person. There are 4 FAM (blue), 3 HEX (green) and 4 NED (yellow) markers displayed. The red plot shows the ROX size standard and has peaks at 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340 and 350 base pairs. For each colour, genotyper automatically attaches size labels to each peak above a given intensity and within an expected size range. These are shown beneath the peaks in boxes*

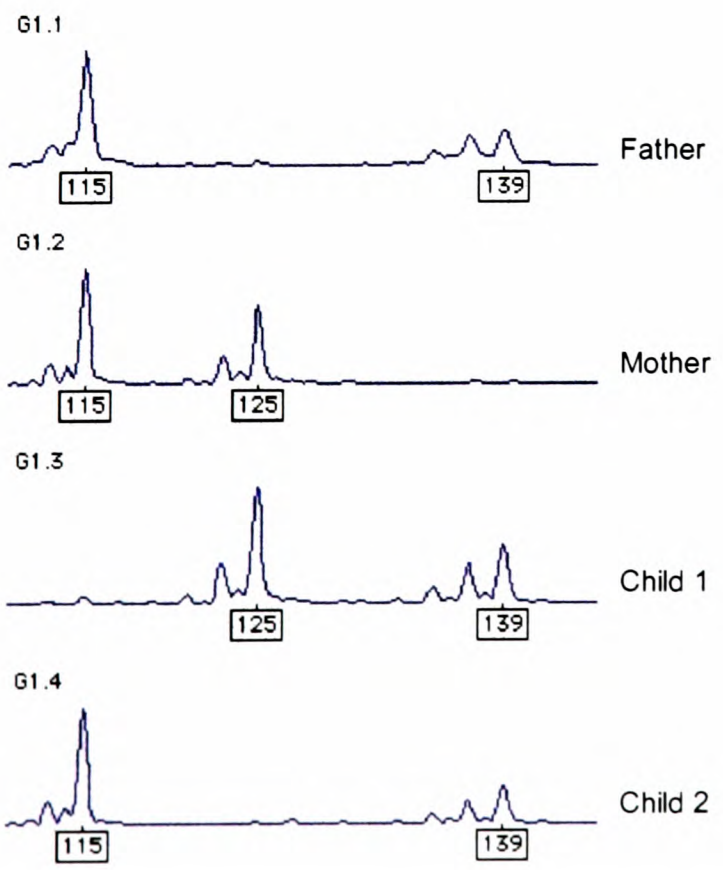


Figure 2.8b - Allele Calling Within Genotyper : *The figure to the left shows the plots for a single marker, in a family of four. Both the mother and the father are heterozygous for this marker (115/139 and 115/125 respectively). One child has inherited a 125 allele from his mother and a 139 allele from his father. The second child has inherited a 115 allele from his mother and a 139 allele from his father. The peaks seen prior to the 'true' allele are typical of a genotyper output and are known as 'stutter bands'. Stutter bands are caused by polymerase slippage during the amplification of the repetitive microsatellite sequence.*

2.6ii - Data Checking

Prior to statistical analyses several data verification steps were performed:

i. The allele calling for every genotype was manually checked for errors within the genotyper package.

ii. Raw allele size data for every family were checked for inconsistencies and deviations from Mendelian inheritance within the Genetic Analysis System (GAS) software (version 2.0) (A Young).

iii. Marker haplotypes were generated within GENEHUNTER2.0 (GH2.0) (Kruglyak et al 1996), and all chromosomes showing an excessive number of recombination events were re-examined at the genotype level.

iv. Corrected data was then run through SIBMED (SIBpair Mutation and Error Detection) (Douglas et al 2000) to identify possible genotyping errors or mutations. SIBMED uses a hidden Markov model to calculate posterior error probabilities for each sib-pair-marker combination, given all the available marker data, an assumed genotype error rate, and a known genetic map. We ran 100,000 Monte Carlo simulations within SIBMED, and assumed a genotype error rate of 1% and a false positive rate of 0.001. All genotypes highlighted by SIBMED were excluded from subsequent analyses.

2.6iii - Data Formats and Information Content Maps

Marker allele frequencies were estimated within RECODE (version 1.4) (D Weeks) and MEGA2 (version 2.2) (Mukhopadhyay et al 1999) was used for the creation of linkage files in a GENEHUNTER2.0 (Kruglyak et al 1996) format. The Discovery Manager™ system (Genomica Corporation) was used for the storage and export of all genotypic data.

Sex-averaged marker maps were taken from CHLC (Co-operative Human Linkage Center) and supplemented with data from Généthon (Dib et al 1996).

Information content maps were produced for each chromosome within MAPMAKER/SIBS (version 2.0) (Krugylak and Lander 1995) and used to determine the markers used in a second round of genotyping involving 101 microsatellites taken from the Généthon map (Dib et al 1996) and the ABI PRISM LMS2-HD5 panels. A list of these additional markers can be found in appendix G. This additional wave of markers allowed the elimination of gaps both in

marker density and information. The typing of these panels was performed in an identical manner to that described for the MD10 panels (see sections 2.4 to 2.6). Final marker density was estimated at less than 8cM for all chromosomes.

2.7 LINKAGE ANALYSIS

2.7i - Haseman Elston and Variance Components Analysis

Haseman-Elston (HE) (Haseman and Elston 1972) and variance components (VC) (Amos 1994; Pratt et al 2000) methods were used within GENEHUNTER2.0 (GH2.0) (Kruglyak et al 1996) to calculate both singlepoint and multipoint LOD scores for all autosomes. The three traits, ELStrans, RLStrans and NWRtrans (see section 2.2 'The genome screen phenotypes'), were analysed for the entire genome as quantitative measures of language ability. Additional multipoint HE and VC analyses were subsequently performed using the Wechsler measure of performance IQ (PIQ) (see also section 2.2 'The genome screen phenotypes') for all areas which showed suggestive linkage to a language trait.

GH2.0 implements a traditional Haseman-Elston regression of squared phenotype differences (D^2) upon estimated Identity By Descent (IBD) sharing (v_i) for each sib-pair at a given genetic locus. At a QTL the variance of D^2. (σ^2_i) is expected to be negatively correlated with the proportion of markers shared IBD (Haseman and Elston 1972).

GH2.0 offers two different weighting schemes to allow for the non-independence of multiple sib-pairs. The first treats all sib-pairs equally and offers no weighting correction for multiple sibs. This leads to an overestimation of sharing probabilities and hence has been suggested to lead to an inflation of LOD scores (Hodge 1984). The second, weights all multiple sib-pair information by the factor (2/no. of sibs). This effectively collapses multiple sib-pairings to produce the same amount of information as a single sib-pair. This leads to an underestimation of sharing and hence may lead to an inflation of type II errors.

For all HE analyses, the first weighting scheme was used and no correction was made for multiple sib-pairings. Simulations, described below (section 2.7iii 'Simulations') indicate that, in this case, the lack of weighting does not lead to an inflation of linkage significance and is therefore valid within our samples.

The VC method derives two maximum likelihood models, both of which dissect the trait variability between siblings into major gene (σ^2a), polygenic (σ^2g) and environmental (σ^2e) variance components. Under the null hypothesis it is assumed that there is no major gene effect (i.e. $\sigma^2a = 0$) and in the alternative model the major gene effect is unrestricted (i.e. $\sigma^2a \neq 0$). Comparison of the likelihood of these two models results in a likelihood ratio estimate and the theoretical significance of linkage effect can be assessed using a standard χ^2 test (Amos 1994). Empirical estimates of the significance of all VC results were derived using simulations as described below. VC analysis was performed using a single mean and with no dominance variance. No adjustment was made for multiple phenotypes. The VC analysis treats each family as a whole and hence does not require any correction for multiple sib-ships.

Regions of 'linkage' were identified as those which exceeded thresholds for 'suggestive linkage' (Lander and Krugylak 1995) under all four types of analysis performed.

2.7ii - X Chromosome Analysis

In the absence of a multipoint sex linked variance components method, linkage to the X chromosome was assessed using HE only. Linkage analyses were performed within MAPMAKER/SIBS (version 2.0) (Krugylak and Lander 1995), under a HE algorithm comparable to that used by GH2.0, described above.

2.7iii - Simulations

Deviations from assumptions made by both of the linkage methods described above (VC and HE) can lead to unpredictable variations in the relationship between nominal p values and LOD scores resulting in both type I and type II errors. The VC method supposes the multivariate normality of data, and the unweighted HE method assumes statistical independence between all sib-pairings in families with multiple sib-ships. We adjusted for any divergence from these assumptions by performing simulations for each phenotype. This allowed an estimation of the empirical pointwise significance of LOD scores.

Pedigree structure and phenotype data were maintained for each family within the genome screen and SIMULATE (J Terwilliger) was used to generate random genotypes for a single marker with four equi-frequent alleles (75% heterozygosity) within this framework. 100,000

replications were run and linkage was assessed for each using both the VC and unweighted HE approaches.

2.7iv - Fine Mapping

Linkage analyses identified two areas of 'suggestive' linkage (chromosome 16q and chromosome 19q). One of these regions (chromosome 19q) was chosen to fine map and an additional 26 markers were chosen to form a 2-3cM grid across the linkage region. These were selected, primarily from the Généthon map (Dib et al 1996) and the ABI PRISM LMS2-HD5 panels and were genotyped and analysed as described above. The fine mapping markers were separated into 4 panels, details of can be found in appendix H.

2.8 DENATURING HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY (DHPLC)

2.8i - Using DHPLC to Detect Polymorphisms

Polymorphisms within the coding sequences of candidate genes were initially detected by Denaturing High-Performance Liquid Chromatography (DHPLC) using the WAVE™ DNA Fragment Analysis System (Transgenomic).

DHPLC involves the fractionation of double stranded DNA on a reverse-phase column and exploits the fact that two completely complementary bound stretches of DNA will have a different melting point from a similar piece of DNA that carries a single base pair mismatch (Oefner & Underhill 1995).

A target region of DNA is amplified from a single individual and the PCR products are melted and allowed to reform under controlled cooling conditions. If the individual is homozygous along the target stretch of DNA, this process results in a population of identical, double stranded, DNA molecules. If, however, the individual carries a polymorphism within the target DNA, a mixture of homo- and hetero-duplex DNA will be formed (see figure 2.10). The homoduplex DNA consists of two completely complementary strands of DNA, whilst the heteroduplex DNA consists of two complementary strands which carry a single base mismatch at the site of the polymorphism.

This duplex DNA is injected into a column within the WAVE machine which contains C_{18} alkylated, polystyrene-divinylbenzene polymeric beads and is filled with two buffers; Buffer A (0.1M TriEthylAmmonim Acetate (TEAA)) and Buffer B (0.1M TEAA and 25% acetonitrile). The TEAA, contained in both buffers, has a high positive charge and binds tightly with both the polymeric beads and with the negatively charged DNA. The acetonitrile in buffer B passes through the column, over the bound DNA and forms the mobile phase of the column. Acetonitrile is an organic solvent that acts to disrupt the base pairing of DNA. Because of it's mismatched base pair, heteroduplex DNA is much more susceptible to this disruption than the tightly bound homoduplex DNA. The concentration of acetonitrile in the column is increased as the DHPLC is carried out.

Once the DNA is bound, the temperature of the column is slowly increased over a controlled, pre-determined range. At a given optimal temperature (which can be calculated from the GC content of the DNA sequence) and acetonitrile concentration, the heteroduplex DNA will begin to melt but the homoduplex DNA will still be strongly bound together. At this point the acetonitrile washes the heteroduplex DNA from the column whilst the homoduplex DNA remains bound to the polymeric beads. A UV light detects the elution of DNA from the column and the presence of a single, or double elutant phase indicates whether the DNA in the column was completely homoduplex (i.e. from a homozygous individual), or a mixture of homo- and hetero-duplex DNA (i.e. from a heterozygous individual) (Figure 2.9).

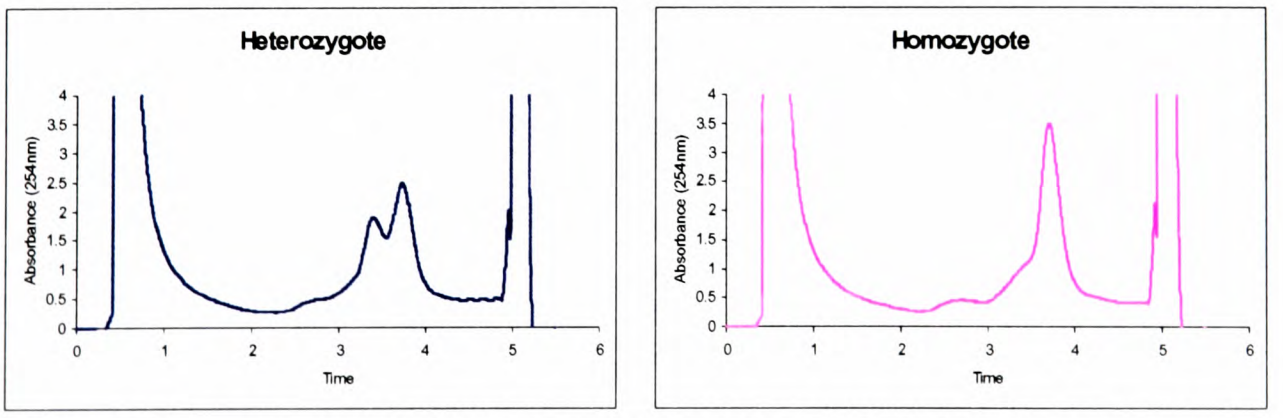


Figure 2.9 - DHPLC Polymorphism Identification : *Output from the Transgenomic Wave machine. The presence of two peaks in the first plot indicates that two populations were washed from the column (i.e. the heteroduplex DNA followed by the homoduplex DNA) and therefore that the individual was a heterozygote. In the second plot, there is one single peak indicating that the DNA was in the form of homoduplexes and therefore that the individual is a homozygote*

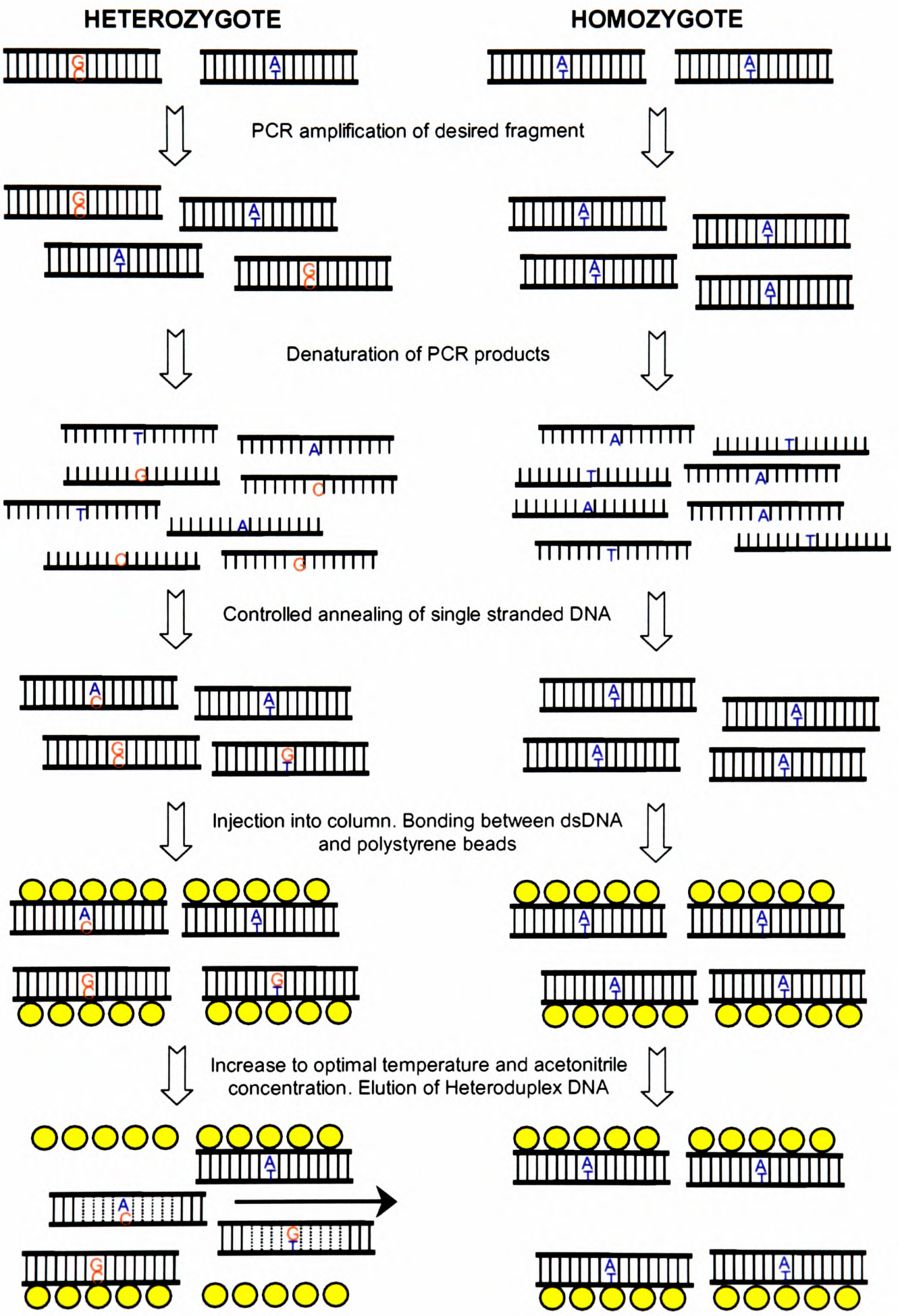


Figure 2.10 - Denaturing High Performance Liquid Chromatography (DHPLC)

2.8ii - DHPLC Analysis of Candidate Genes

Primers pairs were designed to amplify the exons (and surrounding intron-exon boundaries) of each gene of interest (see chapter 4 'Candidate gene analyses'). Each fragment was amplified in a 50µl reaction, on PTC-225 thermocyclers, with a heated lid and no oil, using a touchdown PCR protocol as outlined below.

DHPLC Reaction mix

Reagent	Amount	Comments
DNA (8ng/µl)	5µl	
F-Primer (10pmol/µl)	1µl	
R-Primer (10pmol/µl)	1µl	
10X buffer (KCl)	5µl	
dNTPs (8mM)	1.25µl	
$MgCl_2$ (25mM)	3, 4, 5 or 6µl	
Pfu/*Taq* polymerase mix	0.2µl	*Pfu* polymerase has a proof reading activity
MilliQ H_2O	to 50µl	

DHPLC PCR program

Block temp	Duration	
95°c	10 minutes	
95°c	30 seconds	X14 cycles
Ta + 7.5°c	30 seconds	
Ramp to Ta	@ -0.5°c/cycle	
72°c	30 seconds	
95°c	30 seconds	X20 cycles
Ta	30 seconds	
72°c	30 seconds	
72°c	7 minutes	
15°c	**HOLD**	
END		

The PCR products were checked by running 5µl of PCR product on 2% agarose and all acceptable PCR products were subjected to a hybridisation step, on PTC-225 thermocyclers, as described below.

DHPLC Hybridisation

Block temp	Duration		Comments
95°c	4 minutes		Denatures all DNA fragments
95°c	1 minute	X42 cycles	Cools the mixture of DNA slowly and in a controlled manner - allows the formation of hetero- and homo-duplexes
Ramp to 67.2°c	@ -1.6°c/cycle		
15°c	**HOLD**		
END			

Following the hybridisation step, all fragments were kept at -20°c until ready for DHPLC analysis. DHPLC was performed in a column in the WAVE™ DNA Fragment Analysis System allowing the identification of those fragments that contained a polymorphism. Those fragments that showed a variant elution pattern were directly sequenced (see section 2.9 'Sequencing').

2.9 SEQUENCING

2.9i - Dye Terminator Sequencing

All sequencing was carried out on an ABI 377 sequencer using a cycle sequencing, dye terminator system. This is an automated sequencing system based on the Sanger dideoxy-chain termination method (Sanger et al 1977) and involves the amplification of DNA via three stages (see Figure 2.11). The template DNA is amplified by a normal PCR reaction and this fragment is then amplified further in a sequencing reaction. The sequencing reaction involves the amplification of the target fragment in the presence of a normal concentration of dNTPs (dGTP, dATP, dCTP and dTTP) and a low concentration of dideoxynucleoside triphosphates (ddGTP, ddCTP, ddTTP and ddATP), which lack a 3' hydroxide group. Each of the ddNTPs is labelled with a base-specific fluorophore. Thus the primer extension reaction occurs as normal until a ddNTP is incorporated, at which point the DNA synthesis will cease and the extended fragment is labelled with a fluorophore specific to its ddNTP 3' base. The subsequent separation of the sequencing products on a polyacrylamide gel allows the elucidation of the DNA sequence by a scanning CCD camera (see section 2.5ii 'Polyacrylamide gels') (Prober et al 1987).

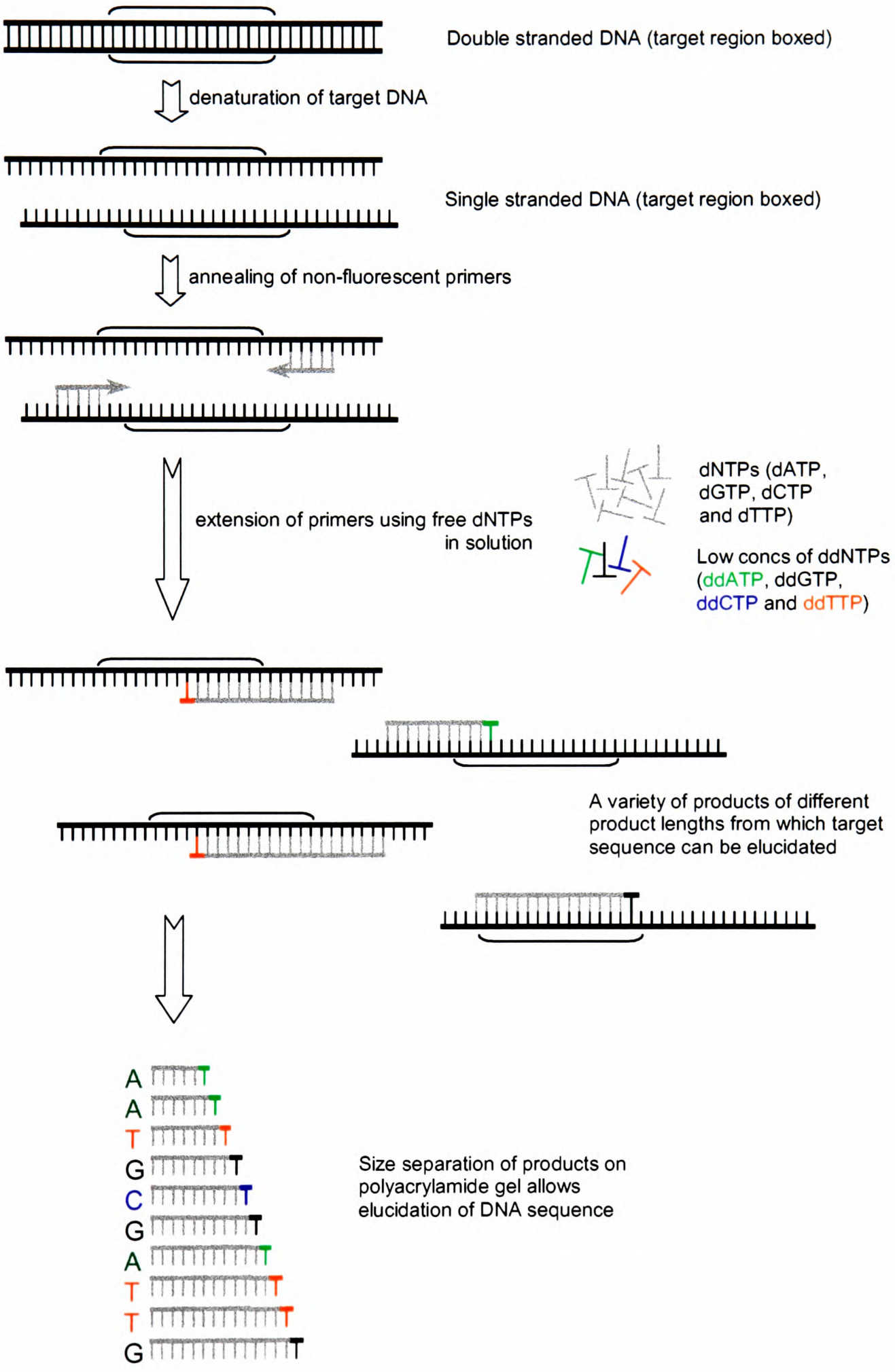


Figure 2.11 - Dye Terminator Sequencing

2.9ii - PCR of Template DNA and Purification of PCR Products

The template DNA was amplified in a 30µl PCR reaction as described in section 2.4ii (PCR conditions and programs'). The PCR product was purified on a 96 well Millipore purification plate. These plates contain a nitro-cellulose filter membrane that binds DNA of between 100 and 1000 base pairs. The PCR reagents and genomic DNA are drawn through this filter by a vacuum manifold (Millipore), leaving the PCR products bound to the membrane surface.

1) Load the PCR products onto the plate

2) Place the plate on top of a vacuum manifold

3) Apply vacuum at 10 inches Hg until the wells have emptied (*The filters appear shiny when dry)*

4) Resuspend the fragments in 50µl MilliQ dH_2O to each well

5) Place the plate on top of a vacuum manifold

6) Apply vacuum at 10 inches Hg until the wells have emptied *(This additional wash acts to purify the PCR products further)*

7) Resuspend the fragments in 15µl MilliQ dH_2O *(The samples are resuspended in a reduced amount of water to concentrate the products in solution)*

8) Mix samples on a plate mixer for 5 minutes

9) Transfer to a microcentrifuge tube

10) Check 5µl of recovered fragment on 2% agarose against 0.5µg of 1kbp ladder

2.9iii - Sequencing reactions

Two 10µl sequencing reactions were prepared for each DNA sample; one with the forward primer and one with reverse primer. Sequencing reactions were performed on PTC-225 thermocyclers, with no oil and a heated lid.

Sequencing reaction mix

Reagent	Amount	Comments
Big Dye Terminator	2µl	Commercially available sequencing mix (see below)
2.5x dilution buffer	2µl	
Primer (F or R)(10µM)	0.4µl	
PCR product	30-40ng (usually 3µl)	
MilliQ dH_2O	To 10µl	

Sequencing program

Block temp	Duration	
96°c	10 secs	X34 cycles
50°c	5 secs	
60°c	4 mins	
15°c	HOLD	
END		

The BigDye Terminator mix contains an Ampli*Taq* DNA polymerase, fluorescently labelled ddNTPs, dNTPs, $MgCl_2$, and the reaction buffer necessary for the sequencing reaction.

2.9iv - Ethanol Precipitation of Sequencing Products

Following the sequencing reaction the PCR reagents are removed from solution by an ethanol precipitation in the presence of sodium acetate. This monovalent cation acts to reduce the co-precipitation of free dNTPs in solution.

Ethanol Precipitation of sequencing products

Reagent	Amount
Sequencing reaction	10µl
NaAc pH5.2 3M	2µl
100% Ethanol	30µl

1) Leave on ice for 15-30 mins
2) Spin 30 mins 13000rpm 4°c
3) Remove ethanol
4) Add 100µl 70% ETOH
5) Spin 15 mins 13000rpm 18°c
6) Remove ethanol
7) Air dry pellet at room temperature
8) Resuspend pellet in 2µl of formamide loading buffer

The sequencing products were separated on a polyacrylamide gel on an ABI 377 (as described for genotyping - see section 2.5ii 'Polyacrylamide gels').

Sequence data were extracted from gels using the Sequencing Analysis program (ABI) and analysed using the Sequence Navigator software (ABI). Sequence Navigator provides a

visual representation of the sequence data and allows the alignment and correction of base calling by the Sequencing Analysis program (see figure 2.12).

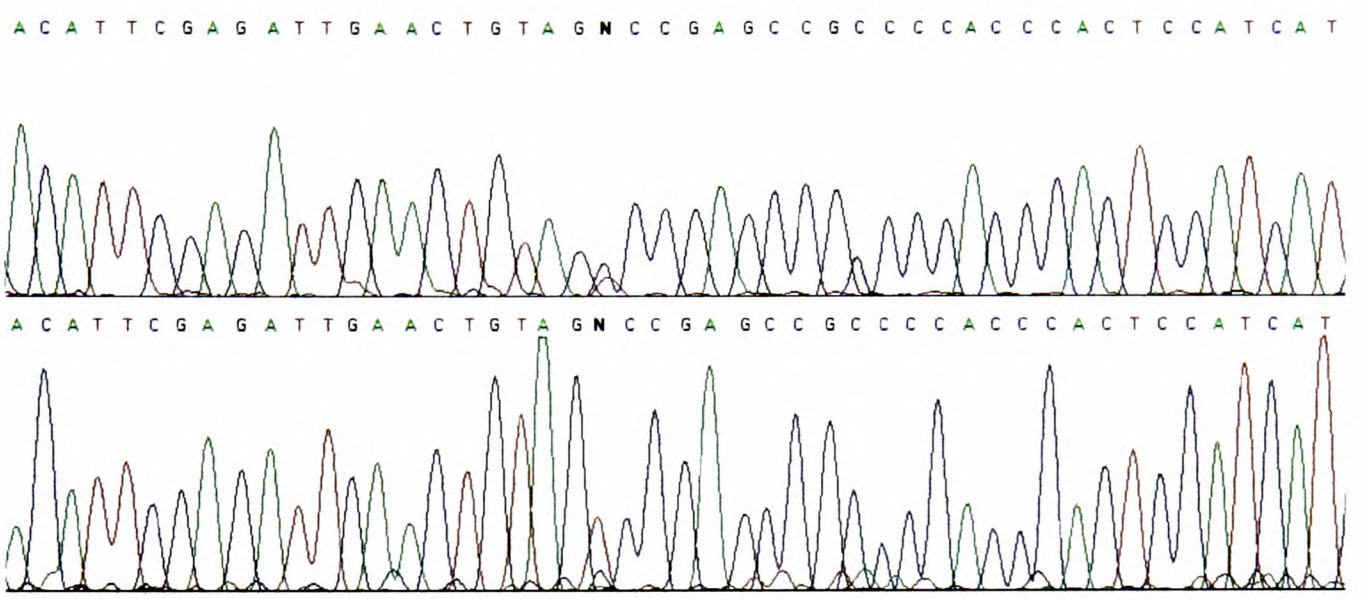


Figure 2.12 - Sequencing Plots : *Forward (top) and reverse (bottom) sequences of a stretch of DNA for a single individual, as viewed in sequence navigator. Guanine is represented by blue peaks, Adenine by green, Cytosine in blue and Thymine by red peaks. Note that the sequence involves a C to T polymorphism (marked by the N)*

2.10 SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

2.10i - SNPs

In general, tests of association (see section 4.2 'Association') are better suited to the analysis of single nucleotide polymorphisms (SNPs), rather than microsatellite markers. SNPs involve only a single base change and therefore more commonly found throughout the genome (approximately once every 1Kbp) than their microsatellite counterparts (Stoneking 2001). They are biallelic and therefore relatively easy to type and show only low rates of mutations. All these features make them very amenable to association analysis (The International SNP Map Working Group 2001).

Ventures by both the Human Genome Project (HGP) and the SNP Consortium have lead to an exponential increase in the mapping and characterisation of SNPs over the last decade. At the last freeze, the SNP database (dbSNP – SNP Consortium) contained details of 4.2 million SNP markers (build ID 105).

2.10ii – SNP Typing

All SNPs used were chosen from the SNP database (dbSNP) and were selected to be typeable by restriction enzyme analysis.

Restriction enzymes are endonucleases that recognise a specific DNA nucleotide sequence (usually palindromic in nature) and cleave the DNA at, or near to, these recognition sites by the digestion of internal phosphodiester bonds. These enzymes form part of the host-specific response in bacteria and act to protect the bacteria from foreign DNA and have many applications in molecular genetics, one of which is to type sequences which vary at a single base. If this base forms part of the recognition sequence then one variant will be cut by the enzyme whilst its alternative will remain intact. Thus DNA sequences can be easily typed by the digestion of target fragments and the separation of products on 1% agarose.

A variety of SNPs were chosen to cover each region of interest and primers were designed to allow the amplification of each appropriate region by PCR. Each SNP was then validated by sequencing within 24 unrelated controls. From a total of 26 SNPs, all selected from dbSNP, 11 were found to be polymorphic by this method (see table 2.10).

SNP ID	Region	Polymorphic?	Major Allele Freq. in Controls
NUMB-R			
2561530	*NUMB-R*	No	—
2604887	*NUMB-R*	Yes	0.62
2561534	*NUMB-R*	No	—
2604883	*NUMB-R*	No	—
2604881	*NUMB-R*	No	—
2561537	*NUMB-R*	Yes	0.50
2604878	*NUMB-R*	No	—
2561540	*NUMB-R*	No	—
2561546	*NUMB-R*	No	—
2604892	*NUMB-R*	Yes	0.69
2561551	*NUMB-R*	Yes	0.67
FOXP2			
1852469	*FOXP2*	No	—
1527158	*FOXP2*	No	—
1450832	*FOXP2*	No	—
923875	*FOXP2*	Yes	0.56

Table 2.10 - SNP Validation : *For details of the gene or region of interest please see relevant sections.*
***For FOXP2** please see section 4.4 'Forkhead bOX domain P2 (FOXP2)'*
***For NUMB-R** please see section 4.3 'Numblike (NUMB-R)'*

SNP ID	Region	Polymorphic?	Major Allele Freq. in Controls
RP11-105N14			
1564610	RP11-105N14	No	—
2371790	RP11-105N14	Yes	0.75
1025820	RP11-105N14	No	—
1026886	RP11-105N14	Yes	0.93
1441165	RP11-105N14	Yes	0.59
1441169	RP11-105N14	Yes	0.63
1372132	RP11-105N14	No	—
1488984	RP11-105N14	Yes	0.70
1219987	RP11-105N14	No	—
723621	RP11-105N14	No	—
977251	RP11-105N14	Yes	0.94

Table 2.10 - SNP Validation (cont.) : *For details of the gene or region of interest please see relevant sections.*
For RP11-105N14 please see section 5.6 'FISH results'

For each validated SNP, the flanking region was amplified in a 15µl PCR reaction as usual (see section 2.4ii 'PCR conditions and programs') and the resultant products digested in a 10µl reaction in a 0.2ml microcentrifuge tube.

Digestion reaction mix

Reagent	Amount	Comments
PCR product	3-5µl	As estimated from agarose gel
Restriction enzyme	5units	
10x Restriction buffer	1µl	The restriction buffer is determined by the enzyme used
100x Bovine Serum Albumin (BSA)	0.1µl	BSA is sometimes required to supplement certain enzymes
MilliQ dH_2O	To 10µl	

Details of digestion reactions can be found in appendices Q and T.

Digestion products were separated and typed on 1% agarose gels.

2.10iii – Quantitative Transmission Disequilibrium Test (QTDT)

A test of association was performed by the Quantitative Transmission Disequilibrium Test (QTDT) program (Abecasis et al 2000) which employs a variance components model that partitions association into between- and within- family components. The association program allows the analysis of quantitative trait scores and thus circumvents the need for a qualitative affection status. In addition, the package includes a permutation framework, which allows the derivation of empirical *P*-values for the sample being evaluated. This corrects for small

sample sizes or deviations of quantitative traits from multivariate normality (as are often seen in selected samples such as ours). 1000 simulations were performed for all of the markers typed across chromosomes 2, 16 and 19 for each of three genome screen language traits (i.e. ELStrans, RLStrans and NWRtrans - see section 2.2 'The genome screen phenotypes').

2.11 FLUORESCENCE IN-SITU HYBRIDISATION (FISH)

2.11i - Epstein-Barr Virus Transformation of Cultured Lymphocytes

The DNA for FISH was collected as 10ml whole blood in heparin. Each sample was sent to the European Collection of Cell Cultures (ECACC) at Porton Down, where immortalised lymphoblastoid cell lines were created by Epstein-Barr Virus (EBV) transformation. A sample of each cell line was returned to us for culture. Once immortalised, the lymphocyte cell line can be kept growing indefinitely to provide a constant source of DNA from the patient sampled.

2.11ii - Cell Culture

Two separate cell cultures were created for each sample; one for chromosome harvest, and the second for cryopreservation to maintain a stock. Cells were maintained in vented tissue culture flasks at 37°c and 6% CO_2. Cultures were grown in RPMI 1640 media that contains the necessary amino acids, glucose, salts, vitamins and other nutrients required for cell growth. All medium was supplemented with 10% Foetal Calf Serum (FCS), 2mM L-Glutamine (an unstable amino acid which needs to be added freshly to the medium), 500U/ml penicillin and 5µg/ml streptomycin (to reduce bacterial contamination).

RPMI-1640 contains a pH indicator which changes colour to indicate the acidity of the medium. When fresh, the medium is pH7.2-7.4 and red in colour, over time, as the medium becomes more acidic, it changes to a yellow colour. This colour change provides an indication both of the cell growth rate and the nutrient requirements of the culture.

2.11iii - Cryopreservation of Cells

1) Maintain the culture for three to four days until a cell density of approximately 1×10^6 cells/ml is achieved

2) Ensure that the cells have been freshly fed the night before cryopreservation is performed

3) Pour off the excess media and transfer cells to a 15ml Falcon tube.

4) Spin at 1000rpm for 5-10 minutes *(pellet the cells)*

5) Pour off the excess media and gently flick the bottom of the tubes to resuspend the cells.

6) Using a 2ml pipette add 1ml of freeze mix and quickly mix the cells up and down

7) Transfer the mixture to labelled cryovials and leave at room temperature until ready to freeze

Cells were frozen in a controlled manner to -140°c in a cryopreserver (PLANER kryosave), before transferring to a liquid nitrogen store at -194°c. These cell lines were kept as stocks and can be revived for use at a later date if necessary.

2.11iv - Harvesting Lymphoblastoid Cells

Once a growing lymphoblast cell line has been established, chromosomes can be harvested and spread on slides for use in in-situ hybridisation. Chromosomes are collected from metaphase cells, where they take the form of highly condensed chromatin. The harvest process therefore involves the synchronisation of the cell cycle within the growing culture to maximise the number of metaphase chromosomes present. This synchronisation is achieved by the addition of a cell cycle block such as 5-bromodeoxyuridine (BuDr), which blocks mitosis at the synthetic (S) phase of interphase, or colchicine, which prevents spindle formation and therefore causes arrest at metaphase. Once all cell cycles are aligned, thymidine is added to the culture to restore normal growth. The cells are subjected to a hypotonic shock treatment (with potassium chloride), which causes the chromosomes to swell, and then fixed in an acidic solution, which removes excess proteins and water and disrupts the cell membrane. The released chromosomes are spread onto a slide upon which they can be easily visualised with a microscope.

1) Set up 10ml of rapidly dividing cells in fresh RPMI 1640 media

2) Add 0.4ml Budr to the cells and mix *(Budr causes the cells to arrest in S-phase)*

3) Leave cells to grow at 37°c for 17 hours

4) Pipette the culture into a 15ml Falcon tube and centrifuge for 5 mins at 1200rpm

5) Pipette off most of the supernatant (leave ~100µl)

6) Resuspend the cell pellet in 10ml fresh RPMI 1640 media

7) Centrifuge for 5 mins at 1200rpm

8) Pipette off most of the supernatant (leave ~100µl) *(these washes ensure that the Budr is removed from the culture)*

9) Resuspend the cell pellet in 10ml fresh RPMI 1640 media and return to a vented culture flask

10) Add an excess (0.1ml) of thymidine (10^{-5}M) and mix *(thymidine releases the cell cycle block exerted by Budr)*

11) Leave cells to grow at 37°c for 6-7 hours *(the cell cycle progresses throughout this growth period. After 6-7 hours most cells should be in metaphase)*

12) Add 0.2ml colchicine (100µg/ml) and mix *(colchicine causes the cells to arrest at metaphase)*

13) Leave cells at 37°c for 5-10 minutes

14) Pipette the cells into a 15ml Falcon tube and centrifuge for 5 mins at 1200rpm

15) Pipette off most of the supernatant (leave ~100µl)

16) Flick the tube to resuspend the pellet

17) Add 5ml of 0.56% KCl *(KCl is a hypotonic which causes the chromosomes to swell)*

18) Incubate at 37°c for 10 minutes

19) Centrifuge at 1000rpm for 5 mins

20) Pipette off most of the supernatant (leave ~100µl)

21) Resuspend the pellet

22) Add 10 drops of fixative in a dropwise fashion, mixing after every drop *(fixative is an acidic solution which removes excess water and disrupts the cell membranes)*

23) Make the volume up to 10ml with fixative

24) Resuspend the pellet

25) Spin at 1000rpm for 5 mins *(this wash and centrifugation removes the proteins and excess water from solution)*

26) Pipette off most of the supernatant (leave ~100µl)

27) Resuspend the pellet in 10ml of fixative

28) Leave -20°c overnight

2.11v - Slide Creation

The fixed cell suspensions were then used to make cell spreads for use in FISH. The cells were washed a further two times in fixative (as described above), and the pellet resuspended in enough fixative to form a milky white solution. 2µl of cell suspension was then dropped onto a wet slide suspended above a 37°c water bath (to increase humidity). 20 slides were prepared for each individual. The slides were allowed to dry at room temperature for 2 hours and stored in a sealed slide box with silicone beads at -20°c until required.

2.11vi - Slide Preparation

Before probes can be used on the chromosomes, they must be denatured and fixed to allow probe hybridisation. This denaturation is achieved by incubation in formamide, and rehydration through an alcohol series.

1) Incubate the slide in 50ml prewarmed 70% formamide/0.6x Sodium Chloride/Sodium Citrate (SSC) for 2.5 minutes (N.B. TIMING IS CRUCIAL – start the clock as the first slide goes in). Do not incubate more than three slides in one jar.

2) Take the slides out of the formamide and incubate in cold 2xSSC for 5 minutes

3) Pass the slides through an alcohol series *(75%x1, 90%x1, 100%x2)* for a minimum of 30 seconds in each alcohol.

4) Air dry the slides

Slides were prepared in this way immediately prior to probing.

2.11vii - Bacterial Artificial Chromosomes as FISH Probes

Bacterial Artificial Chromosome (BAC) clones were chosen to flank the expected position of the breakpoint and ordered from the Children's Hospital Oakland Research Institute (CHORI). All clones were chosen from the RPCI-11 library (Osoegwa et al 2001), which consists of a series of pBACe3.6 vectors (Frengen et al 1999), into which the entire genome of a random anonymous human male subject has been cloned.

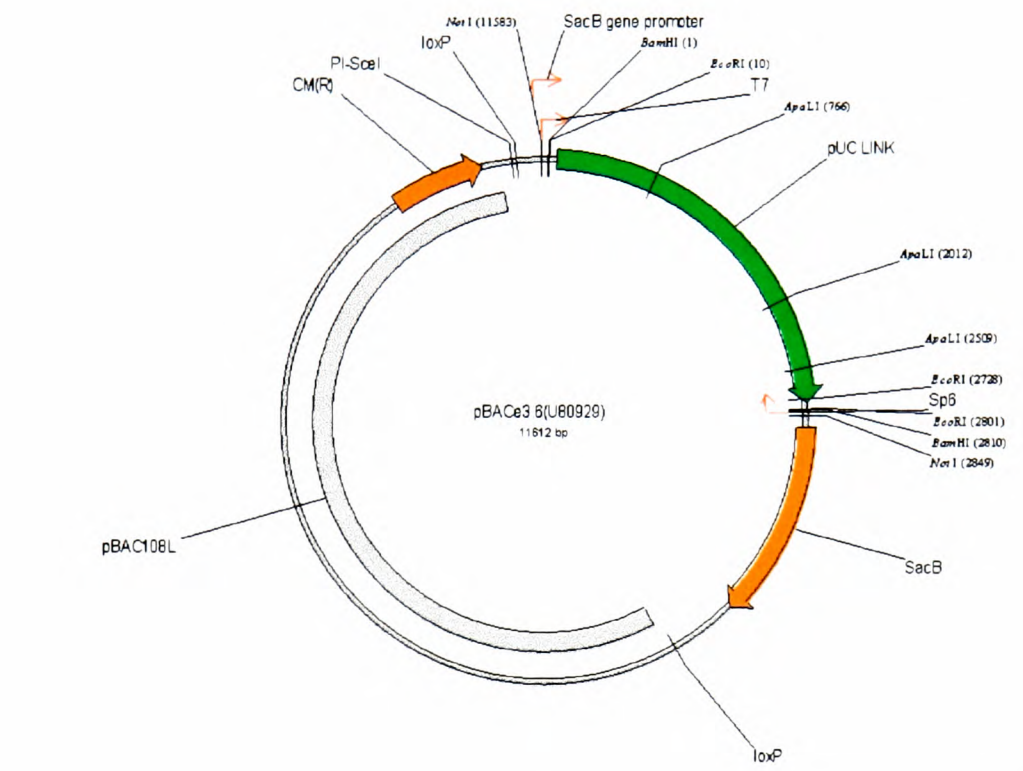


Figure 2.13 - The pBACe3.6 Vector : *The pBACe3.6 vector is an 11.6Kb BAC that carries multiple restriction sites for cloning and a chloramphenicol resistance gene (CM(R)) for selection. This plasmid is present in a high copy number within the E. coli host and so produces an efficient yield of DNA for probe manufacture*

2.11viii - Preparation of FISH Probes

All clones were propagated within a culture of *E. coli* and then purified from culture using a miniprep protocol. Clones are grown in Luria Broth (LB) overnight and the bacterial cells are recovered from culture by centrifugation. A mixture of detergent (SDS) and alkali (sodium hydroxide (NaOH)) are added to the bacterial cells, causing them to lyse and directly disrupting the base pairing of the host genome. The plasmids escape this denaturation, as they are small and circular. Ammonium acetate is added to precipitate the SDS, NaOH and the resultant lysates. The precipitate is spun out of solution and the plasmid DNA recovered from solution by an isopropanol precipitation. Isopropanol is often used instead of ethanol in the precipitation of small amounts of DNA, as a reduced volume of isopropanol is required for precipitation. It should be noted however, that isopropanol is less volatile than ethanol and is therefore harder to remove from solution. Following DNA purification, a small amount of Rnasel is added to the solution to remove all RNA from preparation.

1) Streak out each BAC culture on a plate of LB agar, containing 25µg/ml chloramphenicol and incubate at 37°c overnight
2) Pick four clones, for each BAC and grow each in 20ml LB and 400µl (25µg/ml) chloramphenicol at 37°c for 16 hours
3) Take off 850µl culture and add 150µl glycerol to keep as a stock. Freeze at -70°c
4) Spin the remaining culture at 2000rpm for 20 minutes *(pellets bacterial cells)*
5) Discard the supernatant and allow the pellets to air dry for a few minutes
6) Add 300µl GTE *(lyses the cells)*
7) Resuspend the pellet completely
8) Transfer to a 2ml eppendorf tube
9) Add 600µl of 1% SDS + 0.2M NaOH (freshly made) *(further lyses the cells and destroys the bacterial genome)*
10) Mix by inverting the tube several times until the mixture becomes viscous
11) Put the tube on ice
12) Add 500µl 7.5M Ammonium Acetate *(neutralises the NaOH)* and mix immediately by inverting several times. A white precipitate should form
13) Leave on ice for 10 minutes inverting occasionally
14) Spin 13000rpm, 30 minutes *(pellets the cell lysates from solution)*
15) Pour the supernatant into a new 2ml eppendorf
16) Spin 1300rpm, 30 minutes
17) Pour the supernatant into a new 2ml eppendorf
18) Add 700µl cold isopropanol and invert the tube several times *(precipitates the DNA)*
19) Spin 1300rpm, 30 minutes
20) Wash the pellet with 70% ethanol
21) Spin 1300rpm, 30 minutes
22) Resuspend the pellet in 100µl 1xTE + 1µl RNase (do NOT dry, do NOT speedvac) *(degrades any residual RNA in solution)*
23) Leave on rocker for 1-2 hours *(to resuspend the DNA)*

To ensure that no rearrangements had occurred within the clones, 4µl of recovered DNA was digested with 0.5µl (10U) of *EcoR*I overnight at 37°c. The digestion products were loaded onto a 1% long agarose gel and run at 40mV overnight. The banding patterns of all clones were compared against each other. This step also acts to confirm the efficiency of the miniprep, as solutions that contain residual cell lysates will be resistant to enzyme cleavage.

A further 2µl of DNA was run on 2% agarose against 50ng, 100ng and 250ng λ*Hind*III to quantify amount of DNA recovered. All clones were stored at 4°c until required.

2.11ix - Nick Translation

A single sample of DNA was chosen for each clone, and subjected to a nick translation assay. This process uses two *E. coli* enzymes (DnaseI and DNA polymeraseI (PolI)) to label the recovered DNA with a fluorescent dye (biotin or DIgoxiGenin (DIG)) so that it may be used as a probe for FISH analysis. DnaseI is used to create single stranded nicks within the double stranded probe DNA and then PolI is used to fill in these gaps. PolI has an exonucleolytic activity that digests the DNA downstream of the nick, and then synthesises a new strand to replace it. The presence of fluorescent dUTPs in the nick translation reaction mix ensures that the new strand created by PolI is fluorescently labelled. The probe is then purified by an ethanol precipitation with an excess of carrier DNA. This carrier DNA aids the efficient extraction of the small amounts of probe DNA from solution (Feinberg & Vogelstein 1983).

The nick translation reactions were performed in a 0.5ml microcentrifuge tube on PTC-225 thermocyclers, with a heated lid as detailed below. If the concentration of the recovered DNA was estimated to be below 60ng/µl then a 50µl nick translation reaction was prepared with double the amount of reagents.

Nick Translation Reaction Mix

Reagent	Amount	Comments
Miniprep DNA	500ng	DNA concentration is estimated from the miniprep gel as discussed above (see section 2.11viii)
0.2mM dUTP	1.25µl	dUTP is labelled with a fluorescent marker (biotin or digoxigenin)
0.1mM dTTP	2.5µl	The concentration of dTTP is kept low so as to encourage the incorporation of labelled dUTP
dNTP mix (dGTP, dATP and dCTP)	5.0µl	The remaining dNTPs are not labelled
Nick translation buffer	2.5µl	Creates optimal conditions for DnaseI and PolI enzymes
Nick translation enzyme	5.0µl	DnaseI and PolI
dH_2O	To 25µl	

Nick Translation Reaction

Block temp	Duration	Comments
16°c	10 hours	Nick translation reaction
65°c	10 minutes	Denaturation of nick translation enzyme
4°c	**HOLD**	
END		

The efficiency of all nick translations were checked by running 10µl of nick product on 2% agarose, against 10µl ϕ*Hae*III marker. This ensured that all products were between 200 and 600bp in size.

Labelled probes were purified from solution by an ethanol precipitation in the presence of an excess of carrier DNA as detailed below.

Ethanol Precipitation Mix

Reagent	Amount	Comments
Nick reaction DNA	20µl	
Precipitation mix	5µl	Containing carrier DNA
3M NaAc	2.5µl	Reduces the co-precipitation of free dNTPs in solution.
Cold 100% ethanol	60.0µl	

The precipitation mix was incubated at –20°c for 2 hours and the DNA pelleted by centrifugation at 1300rpm for 30 minutes at 4°c. The excess ethanol was removed and two further 70% ethanol washes are performed at 13000rpm at 4°c for 5 minutes each. The pellet was allowed to air dry at room temperature and then resuspended in 20µl 1xTE. All probes were stored at –20°c.

2.11x - Probe Preparation

Purified probes were denatured and added onto ready prepared slides (see section 2.11vi 'Slide preparation'). Prior to denaturation, all probes were mixed with Cot^{-1} DNA (i.e. highly repetitive DNA), which hybridises to the repeat elements in the single stranded probe DNA, thus masking any common, non-specific stretches of DNA.

Cot^{-1} reaction mix

Reagent	Amount
Purified probe DNA @ 20ng/µl	4µl
Cot^{-1}	1µl

The Cot^{-1} reaction mix was allowed to dry completely on a heated block at 60-65°c for 10-15 minutes. The dried DNA was resuspended in 5µl of hybridisation buffer and 1µl of this mix was placed in a 0.2ml microcentrifuge tube and heated to 80°c for 5 minutes, followed by a 15-30 minute incubation at 37°c. The denatured probe was then pipetted onto the prepared slide (see section 2.11vi 'Slide preparation'), covered with a coverslip and sealed with a layer of glue. The slide was then incubated in an airtight moist chamber at 37°c overnight to allow the probe to hybridise to the denatured chromosomes.

2.11xi - Probe Detection

The signal emitted by the biotin and dioxygenin probes alone is often too weak to detect. The probe signal must therefore be amplified by adding labelled antibodies that bind to the probe thus increasing the signal intensity. Often, several layers of antibodies are added on top of each other. The exact combination of antibodies used depends on the dUTP used for the original labelling (see table 2.11). Probes that were originally labelled with biotin can be boosted by the addition of the avidin antibody. Each avidin antibody is capable of binding three biotin molecules, thus an initial layer of avadin is followed by a second layer of biotin and a final layer of an avadin/FluoresceinIsoThioCyanate (FITC) complex. Probes that were originally labelled with dioxigenin can be enhanced by the addition of an anti-dioxigenin antibody. Multiple anti-dioxigenin antibodies are able to bind to a single dioxigenin molecule, thus an initial layer of anti-dioxigenin is followed by a single layer of Alexa-488 labelled antibody. Following antibody addition, the chromosomes are counterstained in a propidium iodide or 4',6-DiAmidino-2-PhenylIndole (DAPI) stain.

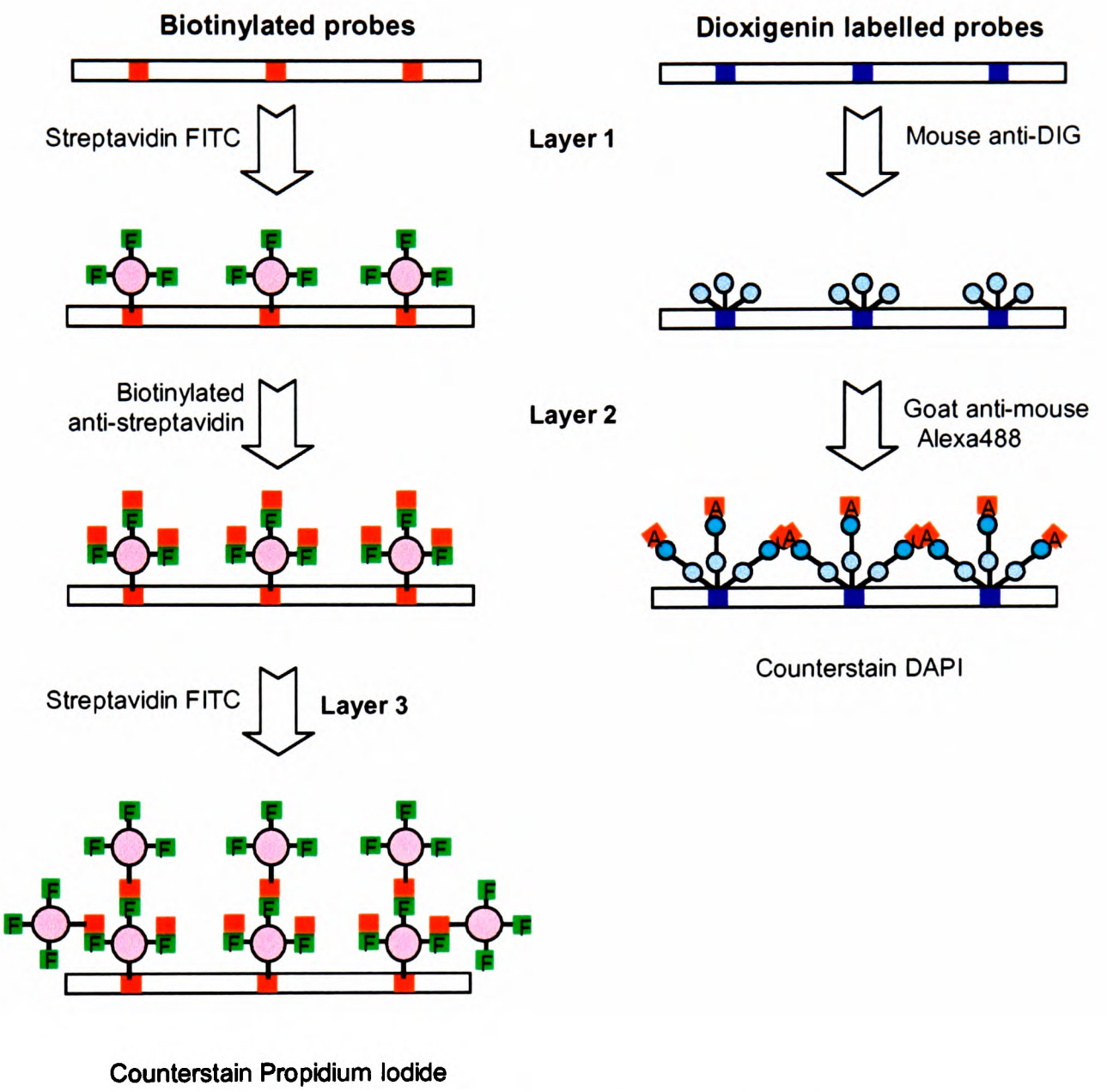


Figure 2.13 - Antibody Layers for Biotin and Dioxigenin Probes : *F = FITC*
A = Alexa 488

1) Remove the slides from their moist chamber (see section 2.11x 'Probe preparation') and remove the glue with forceps

2) Float the cover slips off the slides by soaking slide in 2xSSC room temperature for a few minutes

3) Incubate the slides in 50% formamide/1xSSC prewarmed to 43°c for 10 minutes

4) Transfer slides into 2xSSC and incubate at 43°c for 5 minutes

5) Transfer slides into 1xPhosphate Buffered Saline (PBS) and leave at room temperature for 5 minutes

6) Put a few drops of PBS on coverplate holder and place slide onto holder, DNA side down

7) Put coverplate holder and slide into box until you hear a click

8) Fill the reservoir with PBS/Tween and allow to drain. If the flow rate is very fast check the slide is placed correctly in the holder

9) Add 150µl blocking solution into reservoir and leave for 10 minutes

10) Make up the antibody dilutions for use later (see section 2.11xii 'Antibody layers')

11) Fill the reservoir with PBS/tween and allow to drain

12) Mix and spin layer 1 antibody solution at 13000rpm for 1 minute

13) Add 150µl layer 1 antibody solution to reservoir and incubate rack for 30 minutes at 37°c

14) Fill the reservoir with PBS/tween and allow to drain x2

15) Repeat steps 12 to 14 for each antibody layer used

16) Remove the slides from the coverplate holder and stand on a rack in the dark to drain for 10-20 minutes

17) Counterstain the slides with propidium iodide or DAPI

18) Cover with a large square coverslip and leave at 4°c for at least 1 hour

2.11xii - Antibody Layers

The antibody layers used for each probe type are detailed below.

For Biotin labelled Probes

Antibody layer 1 (amounts for 1 slide)

Antibody	Amount
Streptavidin FITC	0.66µl
Milk + PBS	200µl

Antibody layer 2 (amounts for 1 slide)

Biotinylated anti-streptavidin (goat)	2µl
Milk + PBS	200µl

Antibody layer 3 (amounts for 1 slide)

Streptavidin FITC	0.66µl
Milk + PBS	200µl

For DIG labelled Probes

Antibody	Amount
Mouse anti-DIG	0.8µl
Milk + PBS	200µl

Goat anti-mouse Alexa 488	3.33µl
Milk + PBS	200µl

No third layer

Table 2.11 - Antibody Layers for FISH : *FITC labelled probes appear green on the slide and Alexa488 labelled probes appear red*

It is possible to obtain dual staining on the same slide (see table 2.12). Two probes are prepared one with biotin and one with dioxigenin, as described above, and hybridised to the same slide. This technique allows the positioning of two probes on a chromosome in relation to each other.

For dual detection the counterstain was performed with both To-Pro-3 and DAPI.

For Biotin and DIG Dual Analysis Probes	
Antibody layer 1 (amounts for 1 slide)	
Antibody	**Amount**
Streptavidin TXRD	0.66µl
Mouse anti-DIG	0.8µl
Milk + PBS	200µl
Antibody layer 2 (amounts for 1 slide)	
Biotinylated anti-streptavidin (goat)	2µl
Gt anti mouse Alexa 488	3.33µl
Milk + PBS	200µl
Antibody layer 3 (amounts for 1 slide)	
Streptavidin TXRD	0.66µl
Milk + PBS	200µl

Table 2.12 - Antibody Layers for Dual Layer FISH : *FITC labelled probes appear green on the slide and Alexa488 labelled probes appear red*

CHAPTER 3

The SLI Genome Screen

3.1 GENE MAPPING IN COMPLEX DISORDERS

Although there is ample evidence to suggest that some aspects of language impairment may be genetically determined (see section 1.4 'Is SLI genetic?'), it is generally accepted that most cases of SLI are complex in nature. Complex, or polygenic, disorders involve normal variations within many genes that interact with each other, and the environment, to cause disease. In complex disorders each susceptibility variant confers an increased risk of disease development but no allele alone is sufficient or necessary to produce the disease phenotype. Complex disorders may involve a single major gene, which is influenced by modifier loci, or several genes, all of which have a minimal effect upon overall disorder susceptibility.

3.1i - Functional and Positional Cloning

There are two classical approaches to gene mapping; functional cloning and positional cloning.

Functional cloning involves the identification of candidate genes on the basis of pre-existing biochemical knowledge. For a disease in which the underlying aetiology is clear, functional cloning provides a powerful method of gene identification (e.g. Lesch Nyhan syndrome - Jolly et al 1983). However, for diseases where the biological basis is ambiguous, the number of possible candidate genes is often too high to vindicate a functional cloning approach. In such cases, a positional cloning technique can prove more successful. Positional cloning allows the mapping of disease genes to specific sub-chromosomal regions, even when little, or nothing, is known about the biochemical basis of the disorder (e.g. Duchenne Muscular Dystrophy (DMD) – Koenig et al 1987). This process basically involves the reconstruction of inheritance patterns for the entire genome within families affected by the disorder, thus allowing the detection of regions that are passed onto affected children more often than would be expected by chance alone. This method of gene mapping is often referred to as linkage analysis.

Both the functional and positional cloning techniques are well established in the characterisation of monogenic disorders, and are particularly powerful when the parameters of inheritance (e.g. the mode of inheritance, penetrance, phenocopy rate and disease allele

frequency etc) can be fully defined within the analysis model. In complex diseases, however, phenomena such as incomplete penetrance (i.e. when individuals with the mutation do not develop the disorder), phenocopies (i.e. when individuals with the disorder do not possess the mutation), and variable expressivity (i.e. where a given genotype may result in a range of phenotypes) are widespread. These factors are characteristically hard to measure and can complicate the specification of an accurate disease model. In addition, because complex disorders can involve several genes, the effect size of each gene variant is reduced over that in a single-gene disorder and the individual genes are therefore harder to detect. Moreover, the exact combination of genes underlying the disorder may vary from person to person (i.e. heterogeneity) and thus it is not usually possible to determine a clear disease haplotype. Thus although it is theoretically possible to apply a positional-cloning-based approach to the detection of complex genes, these methods necessitate large sample sizes, highly informative genetic markers and powerful model-free analyses. The reduced power associated with the linkage analysis of complex disorders often results in the identification of less clearly defined chromosomal regions, which can be large in size. Thus a second stage of fine mapping linkage or association analysis is often required prior to candidate gene analysis.

3.1ii - Recombination

Although linkage analyses theoretically involve the scanning of entire genome, they do not require the sequencing of every base or the study of every gene. Instead, they exploit the fact that DNA is inherited as mosaic blocks from parents to offspring. During the production of gametes, chromosomes become entangled with each other to form chiasmata. At these points of cross-over, the DNA can recombine, swapping genetic material between chromatid pairs. Recombination is a highly random process and therefore, even two cells that have inherited the same array of parental chromosomes are likely to differ. The presence of a chiasma in a given stretch of DNA directly inhibits the recombination process in neighbouring regions (i.e. interference), thus resulting in the inheritance of DNA as blocks of a predictable size. On average there are 1.5 recombination events on every chromosome per generation. As a consequence of the recombination and interference processes, the closer together that two genes are on a chromosome, the more likely that they will be inherited within the same

block unit of DNA. Thus by sampling regions of DNA along each chromosome, it is possible to infer the origin of the majority of the genome without the need to sequence every base.

3.1iii - Recombination Fractions

The probability that two loci will be inherited within the same block of DNA is measured by the recombination fraction, θ. In model organisms, these recombination fractions can be simply calculated by counting the number of recombinant and non-recombinant offspring. Recombination fractions can vary between 0 (for two loci that lie so close together on a chromosome that recombination never occurs between them) and 0.5 (for two loci that are so far apart a recombination event always occurs between them). Two loci that have a recombination fraction of zero are said to be tightly linked and always exist within the same linkage group. Because recombination rates depend heavily upon the distance between two loci, they can be used to provide a measure of genetic distance. This is achieved by the use of a mapping function, of which there are two principal formats. The Haldane function ($\theta=0.5(1-e^{-2x})$, where x is the genetic distance) assumes that recombination is random, whereas the Kosambi function ($\theta=0.5\tanh(2x)$) allows for interference. For small values of θ (<0.1 i.e. linked loci) the Haldane and Kosambi mapping functions provide a similar estimate of genetic distance, which is directly equivalent to the recombination fraction expressed as a percentage. For larger recombination values, interference becomes more important and the estimates of genetic distance provided by the Kosambi and Haldane functions can vary. Genetic distances are measured in Morgans (M) and centiMorgans (cM). Two loci which show a recombination frequency of 1% (i.e. we would expect a recombination to occur once in 100 meioses) are said to be 1centiMorgan (1cM) apart. Generally, in humans, a genetic distance of 1cM equates to a physical distance of 1Mb of DNA sequence, although it should be noted that this relationship is not steady and the frequency of recombination events can vary between individuals and chromosomal regions. For example, the recombination rate is higher in males than females and is usually increased in telomeric regions.

3.1iv - Linkage Analysis

The linkage analysis of two genetic markers simply involves the study of how often those two markers mutually co-segregate due to their physical proximity on the chromosome (i.e. the

evaluation of the recombination frequency). In a similar manner, it is possible to test for linkage between a directly measured phenotypic trait and a genetic locus by investigating the relationship between the marker locus and the variance of that trait.

Essentially, all forms of linkage analysis rely upon the estimation of a likelihood ratio which compares the probability of obtaining a given set of data under a null hypothesis, where θ is 0.5 (i.e. no linkage between marker and trait), to that under the alternative hypothesis, where θ is less than 0.5 (i.e. linkage between marker and trait). The ratio of the likelihood of each of these models gives the odds for, or against, linkage (z), and the logarithm of these odds is the LOD score (Logarithm of ODds). The maximum LOD score (MLS) for a given locus is obtained by testing different values of θ within H_1 and deriving a LOD using the most likely value of θ (Nyholt 2000).

The LOD score provides a direct measurement of linkage and thus, by testing a series of loci spread evenly across the genome (see section 3.1v 'Microsatellite markers and Identity by Descent (IBD)'), it is possible to identify those positions which are likely to be linked to genes affecting the trait under study. This process is known as a genome screen. As a general rule, any genomic region with a LOD of above 3.0 is considered to be linked to the phenotype of interest. For a mongenic disorder, this refers to a ratio of 1000:1 in favour of linkage. Note, however, that this is not equivalent to a probability of 0.001, as the LOD score refers to the probability of a ratio between two likelihoods, rather than a single absolute probability value (Lander & Kruglyak 1995). LOD scores can be multiplied by $2(\log_e 10)$, to derive a χ^2 statistic, which can then be converted to a p-value (Lander & Kruglyak 1995).

3.1v - Microsatellite Markers and Identity By Descent (IBD)

A genome screen typically involves the typing of a series of non-functional genetic markers, known as microsatellites. Microsatellites are selectively neutral, non-coding regions of DNA that consist of short (2-3bp), tandem repeat stretches of sequence (e.g. CACACACA). The repetitive nature of the microsatellite sequence can lead to 'slippage' of the polymerase enzyme during DNA replication. This slippage results in the alteration of the number of repeat units in the microsatellite and therefore maintains a high number of alleles and an increased variability between individuals. For example, one individual may have five repeat units

(CACACACACA) where another may have nine (CACACACACACACACACACA). Microsatellites are evenly spread throughout the genome (~1 every 10^6bp) and have a low mutation rate making them ideal for linkage analysis.

The typing of microsatellites basically involves recording the length of the microsatellite repeat (determined by the number of repeat units present) for all individuals across a number of families and for hundreds of markers, spread evenly across the genome. This is achieved by PCR, using unique primers that flank each microsatellite sequence, and polyacrylamide gels, to separate the allelic fragments (see methods).

The resultant microsatellite size data provides a measurement of genetic identity between sib-pairs. If two siblings carry the same allele for a given polymorphic genetic marker they are said to be Identical By State (IBS). If these alleles were inherited from the same ancestral chromosome they are said to be Identical By Descent (IBD). In any family, for a given locus, the expected ratio of offspring sharing 0, 1 or 2 alleles IBD is 1:2:1 (see figure 3.1).

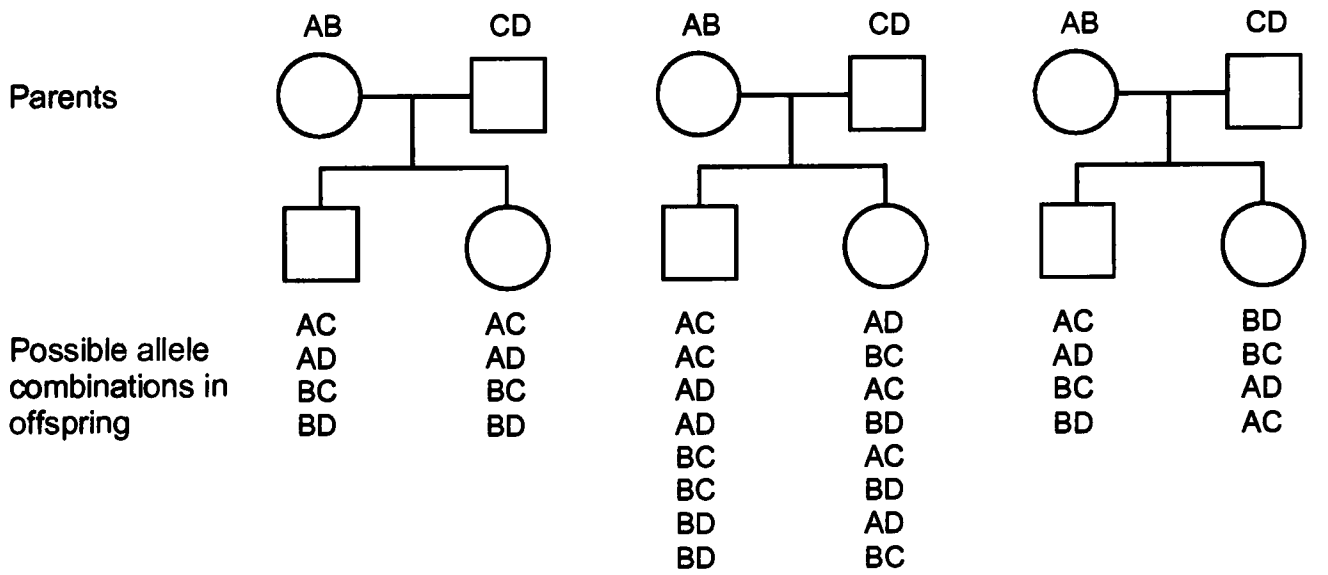


Figure 3.1 - Identity By Descent (IBD) : *At any given locus the probabilities of two siblings sharing 0, 1 or 2 alleles Identical By Descent (IBD) are 0.25, 0.5 and 0.25 respectively. Note that if one parent is homozygous or if two parents share an allele, then it may not be possible to determine the full extent of IBD sharing. Note that it is possible for individuals to be IBS without being IBD, but not vice versa*

The power of linkage analyses can be increased exponentially when the identity by descent, rather than IBS, can be explicitly determined. Thus, microsatellites are usually typed within

parents and children allowing the tracking of alleles between parental and offspring chromosomes. The extent to which a marker can be used to unambiguously indicate the genetic identity between individuals is known as the Polymorphic Information Content (PIC) (Botstein et al 1980). PIC provides a measure of the informativeness of the given marker, and is closely related to the marker heterozygosity (i.e. the frequency of heterozygotes in a population for a given marker). In a genome screen, the information content can be maximised by performing a multipoint analysis. This technique uses the data of flanking, linked loci and a function of genetic distance to infer both missing IBD values and the extent of IBD sharing in regions between markers. Multipoint analysis is therefore heavily dependent upon the correct estimation of intermarker distances (Elston 1998).

3.1vi - Parametric and Non-Parametric Linkage

For a monogenic disorder, linkage analysis can be performed within a pre-determined parametric model in which all the variables of inheritance are given (e.g. the mode of inheritance, penetrance, phenocopy rate and disease allele frequency etc). When the model is correctly defined this provides an efficient method of gene detection. However, mis-specification of the parameters can lead to unpredictable variations in the linkage results obtained (Clerget-Darpoux et al 1986). Thus for complex traits, where the model of inheritance is rarely known, a non-parametric or model-free mode of analysis is more appropriate.

There are various methods of non-parametric analyses suitable for the linkage analysis of complex traits. For qualitative, or binary, traits these tests identify regions where affected sib-pairs show an increased rate of sharing, above that predicted by Mendelian segregation (see figure 3.1). For quantitative, or continuous, traits, they look for regions where there is a constant relationship between the phenotypic similarity and genetic identity of sibling-pairs. This is achieved by testing for the relationship $\sigma^2_0 > \sigma^2_1 > \sigma^2_2$, where σ^2_i is the trait variance of the sibs who share i alleles IBD (Kruglyak & Lander 1995).

3.1vii - Haseman Elston Analysis

The Haseman Elston method (Haseman & Elston 1972) tests for this relationship by performing a linear regression of the sib-pair trait differences upon the proportion of alleles

shared IBD. The regression equation is given by $D^2=\beta\pi+\alpha$ where D^2 is the squared phenotype difference of a given sib-pair, and π is the proportion of alleles shared IBD. The slope of this regression line (β) yields a direct measure of the relationship between the phenotype and genotype and is expected to be negative at a Quantitative Trait Locus (QTL) (i.e. as the sib difference increases, their genetic identity decreases). In the HE approach, a maximum likelihood value of β is derived and the comparison between the probability of this value of β against that of the null hypothesis ($\beta=0$) provides a likelihood ratio from which a LOD score can be derived. The Haseman Elston analysis is a robust technique, which makes few assumptions about the input data. It does however have problems with the weighting of non-independent data derived from multiple sib-ships (see section 2.7i 'Haseman Elston and Variance Components analysis') and is susceptible to error when there is a large amount of missing data.

3.1viii - Variance Components Analysis

A second approach to QTL mapping is Variance Components (VC) analysis (Lange et al 1976, Hooper & Matthews 1982). This approach models the variance of the trait under study within a maximum likelihood framework for each marker locus. The trait variance is divided into genetic and environmental components (V_G and V_e), within which the genetic factor may be further dissected to reflect major gene (V_a) and polygenic (V_g) effects. At a QTL we would expect most of the trait variance to be explained by the major gene variable.

As for a standard linkage analysis, two models are derived, one under the assumption of no linkage (i.e. $V_a = 0$) and a second, maximum likelihood model, within which V_a is unconstrained. Within large samples and under an assumption of trait normality, the likelihood ratio of these two models, can be equated to provide a LOD score (Pratt et al 2000).

The VC approach models each pedigree as a complete unit and therefore does not encounter problems with the weighting of multiple sib-pairs (Pratt et al 2000). However, the assumption of multivariate trait normality places it's own limitations upon the VC method. The analysis of variance components has been shown to be specifically sensitive to skewed or leptokurtic

data, particularly in pedigrees where the family members are highly correlated at the phenotypic level (e.g. in affected sib-pair samples) (Allison et al 1999).

The success of both HE and VC analyses is limited by the sample size used for the genome screen, the number of genes involved in the disorder and the way in which those genes interact. For a sample size of 100 sib-pairs, with a sibling risk ratio (λs) of 2.1 and a linkage threshold of 1.0 it is estimated that a genome scan approach has over 90% power to detect a complex disease susceptibility locus (Weeks & Lathrop 1995). Note, however, that this estimate assumes that the marker and the disease locus are completely linked.

Comprehensive genome screens involve the testing of multiple loci, and are therefore subject to different LOD thresholds from those set for parametric analyses (see section 3.1iv 'Linkage analysis'). For genome scans, a LOD score of above 2.2 ($p=7x10^{-4}$) is considered to be 'suggestive' of linkage (i.e. a result we would expect to occur once by random in an entire genome screen) and a LOD of 3.6 ($p=2x10^{-5}$) is a 'significant linkage' (i.e. a result which can be expected to occur randomly 0.05 times in a genome scan). (Lander & Kruglyak 1995). It should be noted, however, that even significant linkages are expected to be true in only 95% of cases, and thus the replication of genome screen linkages represents an important step in the gene mapping process.

3.1ix - Simulations

Linkage results are directly influenced by the heterogeneity of both the phenotype and genotype data within the sample selected, and are therefore unique to each study cohort. VC results are particularly sensitive to the phenotype trait distributions, whilst HE results can be sensitive to family structures. Thus the interpretation of all linkage results is best achieved through the derivation of LOD-significance distributions from simulation of the data in hand (Lander & Schork 1994). These simulations involve the generation and linkage analysis of random marker genotypes, within a fixed phenotypic and family structure framework, and thus allow the estimation of empirical p-values that can be applied to the LOD scores obtained.

The LOD-significance distributions derived from simulations can be regarded to approximate a multipoint situation (where ~70-80% of IBD information is extracted) and are therefore applicable for estimating the pointwise (or nominal) significance of linkage peaks (Fisher et al

2002a). However, it is important to note that simulations yield only nominal p-values (i.e. the probability of obtaining a given LOD score at a specific locus by chance) and are not adjusted for the scanning of the whole genome.

3.2 GENOME SCREENS

3.2i – Overview of Genome Screens for Complex Disorders

In a recent review, Altmüller et al (2001) created a database of 101 whole-genome scans, which between them covered a total of 31 complex human diseases. Whilst the majority of these genome screens had been completed for tangible, but complex, medical conditions such as asthma, inflammatory bowel disease and diabetes, a surprising number had also been carried out for complex psychiatric illnesses such as autism, schizophrenia and bipolar affective disorder. Most of these neurological screens relied upon a qualitative coding of affection status, with only a minority employing disease-associated quantitative traits. Most investigations used samples of affected sib-pairs, selected from outbred, general populations and all relied upon the typing of microsatellite repeat markers. In total, only 4% of all the genome screens performed yielded a highly significant result, whilst 24% produced a significant linkage, 47% showed suggestive linkage and 24% found only nominal evidence of linkage.

The most striking conclusion of the Altmüller review was that, across the disorders for which more than one screen had been completed, replication of linkage was a rare occurrence. Furthermore, when a 'replication' was reported, the overlap between the two genetic regions was highly variable. Moreover, for those disorders where multiple genome scans have been completed (e.g. asthma and type 2 diabetes), it is not uncommon for some level of linkage to be reported for almost all of the autosomes (see table 3.1).

Disease	No. of genome screens	Chromosomes for which positive findings have been reported	Chromosomes for which linkages have been replicated
Asthma	7	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 19, 21	**5q**, 6p, 12q
Type 1 diabetes	8	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 16, 22, X	**6p**, 11p, 16q
Inflammatory Bowel Disease (IBD)	8	1, 2, 3, 4, 5, 6, 7, 10, 12, 14, 16, 17, 18, 19, 22, X	6p, **12q, 16**
Autism	5	1, 2, 5, 6, 7, 8, 13, 16, 18, 19, X	2q, **7q**, 16p

Table 3.1 – The Overlaps Between Genome Screens of Complex Disorders : *The chromosomes shown in bold represent the most highly replicated chromosomal regions*

Over the 101 studies reviewed, the HLA region on chromosome 6p was the most common area for a linkage result to occur. This region has been linked to several complex disorders including type-1 diabetes, multiple sclerosis, rheumatoid arthritis, psoriasis, inflammatory bowel disease and asthma.

The Altmüller study concluded that, beyond increased sample sizes and reduced heterogeneity, no one factor of study design had a consistent or significant impact upon the outcome of a genome screen.

3.2ii – Genome Screens for Disorders Related to SLI

Autism, dyslexia and Attention Deficit Hyperactivity Disorder (ADHD) are three childhood illnesses that are often associated with SLI. These four disorders have many shared clinical features, show a high level of co-morbidity and often cluster within the same families. However, the mechanisms underlying the phenotypic and epidemiological overlaps between these conditions remain unknown.

Although sibling risk ratios vary widely between the four disorders (60-100 in autism, 5-10 in dyslexia, 5-7 in ADHD and 7-10 in SLI), all show a high level of familiality and are thought to be subject to strong genetic influences. However, it is proposed that in the majority of cases, these impairments will involve complex genetics, with several distinct loci interacting to produce a genetic liability for the onset of each disease (Bishop et al 1995, Pickles et al 1995,

Smalley 1997, Fisher et al 1999). Interestingly, while all four conditions affect many more boys than girls, none show any strong evidence of sex linkage.

Over the last decade, systematic genome-wide scans have been performed for autism, dyslexia, ADHD and SLI. It is hoped that studies such as these will not only allow the identification of the genes underlying each individual disorder, but will also serve to clarify the extent to which the phenotypic overlaps can be explained by shared genetic factors between these disorders.

3.2iii – Autism

Autism (MIM209850) is a neurodevelopmental disorder characterised by deficits in reciprocal social interaction and communication, accompanied by repetitive and stereotyped behaviours and interests (World Health Organisation 1993, American Psychiatric Association 1994).

Although the presence of autistic features traditionally precludes a diagnosis of SLI (see section 1.2v 'Autism and social competence'), the features of the two disorders show marked overlaps. Expressive language deficits form a major component of the autistic diagnostic criteria (Lord et al 1994), and many children with SLI display mild autistic-like features (e.g. joint attention deficits – see section 1.1ii 'The SLI phenotype') (Rapin & Allen 1983). Furthermore, there exists a distinct group of children who present with definite language problems but yet fail to meet strict diagnostic criteria for either disorder. Several researchers present an argument for the formation of a 'semantic-pragmatic' classification to capture these 'borderline autistic/language-impaired' cases (Rapin & Allen 1983, Bishop & Rosenbloom 1987).

Over the last decade, several genome screens have been completed for autism (see table 3.2). The results of these, alongside other molecular work, has highlighted several chromosome regions which may be involved in the onset of autism, and has allowed the beginning of candidate gene analysis in at least one of these regions.

Sample subgroups	No. of ASP / fams	Regions with LODs > 1.0	MLS
International Molecular Genetic Study of Autism Consortium (IMGSAC) (1998)			
UK families	66 families	4p, 7q, 16p, 19p, 22p	3.55 (7q)
All families	99 families	4p, 7q, 10p, 16p, 22p	2.53 (7q)
Philippe et al (1999)			
All families	51 ASP	6q, 15p, **19p**,	2.23 (6q)
Risch et al (1999)			
Wave 1	97 ASP	1p, 1q, 7p, 11p, 13q, 15q, 17p, 18q, 20p	1.87 (1p)
Wave 1 + wave 2	147 ASP	1p, 7p, 17p, 18q	2.15 (1p)
Collaborative Linkage Study of Autism (CLSA) (1999)			
All families	75 families	4, **7q**, 8p, 11, 13p	3.0 (13p)
Buxbaum et al (2001)			
Stage 1 (full genome screen (not X))	35 families	1q, 2q, 5p, 9q	2.25(2q)
Liu et al (2001)			
Broad affection status (autism or PDD)	118 ASP	2p, 5p, 8q, 11p, 15q, **16p**, **19p**, 19q 20p, Xq	2.55 (5p)
Narrow affection status (autism)	75 ASP	3p, 5q, **16p**, **19p**, 19q, Xq	2.53 (19p)
IMGSAC (2001a)			
Type 1/1 sib-pairs (autism only)	43 ASP	**2q**, 4, 15p	2.21 (15p)
Type 1/2 sib-pairs (autism or PDD)	84 ASP	**2q**, **7q**, 9p, 14, **16p**, 17p, 22	3.66 (2q)
Type 1/1 + Type 1/2	127 ASP	**2q**, **7q**, 9p, 10p, 15p, **16p**, 17p,	4.80 (2q)
All families (autism, PDD and borderline autism/PDD)	152 ASP	**2q**, **7q**, 8, 9p, 10p, **16p**, 1p7	3.74 (2q)
Shao et al (2002)			
Stage 1 (full genome screen)	46 ASP	**2q**, 3p, **7q**, **15q**, 18p, **19q**, **Xq**	1.81(X)
Stage 2 (follow up on chromosomes 2q, 3p, 7q, 13p, 15q, 18p, 19q, Xq)	96 ASP	**2q**, 3p, **7q**, **15q**, **19q**, **Xq**	2.49(X)

Table 3.2 - Genome Screens of Autism : *All regions reported to have LODs of above 1.0 are shown. Note that the methods by which these LOD scores were calculated varied between studies. Chromosomes given in bold represent those which replicate the findings of a previous genome screen. The maximum LOD score for the entire genome is given and the chromosome on which it was found is shown in brackets*

***ASP** = Affected Sib-Pairs*

Across the eight autism genome screens, the most consistently replicated region is 7q31-33 (*AUTS1*) (see tables 3.2 and 3.3). This locus provided the strongest evidence for linkage in the original genome screen (IMGSAC 1998) and has since been replicated with varying degrees of success by subsequent scans. Furthermore, in a recent meta-analysis (Badner & Gershon 2002) including four of the autism genome screens (IMGSAC 1998, CLSA 1999, Philippe et al 1999 and Risch et al 1999), the 7q region was also highlighted as the most likely position for an autism-linked locus (P=0.00014).

Study	7q MLS	Position	Comments
IMGSAC (1998)	3.55	146cM	
Philippe et al (1999)	0.83	135cM	
Risch et al (1999)	0.93	137cM	
Ashley-Koch et al (1999)	1.77	129cM	Study of chromosome 7q only
CLSA (1999)	2.2	104cM	
Auranen et al (2000)	0.57	170cM	Study of 10 selected chromosomal regions
Buxbaum et al (2001)	0.75	147.2cM	
Bradford et al (2001)	1.4	103.7cM	Study of chromosomes 7q and 13q only
Liu et al (2001)	2.13	165cM	After fine mapping chromosome 7q
IMGSAC (2001a)	3.20	120cM	
Shao et al (2002)	1.66	128cM	Study of 8 selected chromosomal regions

Table 3.3 - Evidence for a Chromosome 7q Locus in Autism

The evidence for the existence of an autism gene in the *AUTS1* region is strengthened by cytogenetic studies of individuals with chromosome 7 abnormalities. Ashley-Koch et al (1999) described a family with a pericentric inversion of the long arm of chromosome 7 (Inv(7)(q22-q31.2)) transmitted from an unaffected mother to all three of her children. Interestingly, two of the three siblings in this family were affected by autism, and the third presented with a severe expressive language disorder. Vincent et al (2000) characterised a translocation transmitted from an unaffected mother to an autistic child (t(7;13)(q31.2;q21)) and mapped the breakpoint within a highly conserved, brain expressed gene of unknown function (*RAY1*) between markers D7S2460 and D7S633. Warburton et al (2000) described two unrelated cases, one with autism and a second with a severe expressive language impairment, both of whom

showed de-novo abnormalities involving breakpoints on chromosome 7q31 (inv(7)(p12.2;q31.3) and t(2;7)(p23;q31.3) respectively).

However, as a result of the variation in the exact location of *AUTS1* across studies (see table 3.3), the chromosome 7q critical region remains very large (~50cM) and contains a large number of potential candidate genes. Thus, all attempts to identify a critical gene within this region have, as yet, been unsuccessful (see table 3.4).

Gene	Description	Comments
Cheung et al (2001)		
***CORTBP2* (cortactin-binding protein-2)**	Novel cortactin binding protein. May be involved in cell signalling in the cortical cytoskeleton	
Results	1 missense mutation found in 90 autistic probands. Frequency not reported	Missense mutation found at a similar frequency in normal controls
	No significant association found	
Persico et al (2001)		
***PAI-1* (plasminogen activator inhibitor-1)**	Inhibits plasminogen activator genes. Plasmin is a serine protease involved in many physiological processes including neuronal migration and synaptic remodelling	
Results	No linkage or association found in 167 trios with autistic probands	
Wassink et al (2001)		
***WNT2* (wingless-type MMTV integration site family member 2)**	The WNT family of genes influence the development and patterning of numerous organs and systems, including the central nervous system.	
Results	2 missense mutations found in 135 autistic probands. However, these mutations were transmitted from unaffected parents	
	Weak association found with 3'UTR SNP (P=0.013)	
Newbury et al (2002)		
***FOXP2* (Forkhead box gene P2)**	Transcription factor mutated in the KE family (see chapter 4 'candidate gene analyses')	
Results	No evidence for association found	
	No coding changes found in 48 autistic probands	

Table 3.4 - Candidate Gene Studies for Autism at the *AUTS1* locus

Gene	Description	Comments
Bonora et al (2002)		
***PEG1/MEST* (mesoderm specific transcript homologue)**	Imprinted gene expressed in embryonic development in mesodermal tissues and developing brain. Peg1/Mest knockout female mice show abnormal maternal behaviour	
Results	No coding changes detected in 48 autistic individuals	
	No significant association found	
	No abnormal methylation patterns detected (imprinting normal)	
***COPG2* (coatomer protein complex subunit gamma 2)**	Imprinted gene involved in vesicle trafficking of proteins from the golgi complex to the endoplasmic reticulum	
Results	No coding changes detected in 48 autistic individuals	
	No significant association found	
	No abnormal methylation patterns detected (imprinting normal)	
***CPA1* (carboxypeptidase A1)** ***CPA5* (carboxypeptidase A5)**	Zinc containing exopeptidases. Catalyse the release of carboxy-terminal amino acids from proteins	
Results	**CPA1 -** 1 missense mutations found in 135 autistic probands.	Missense mutations found at a similar frequency in normal controls
	CPA5 - 3 missense mutations found in 135 autistic probands.	
	No significant association found	

Table 3.4 - Candidate Gene Studies for Autism at the *AUTS1* locus (cont.)

Two other regions that are often studied with respect to autism are chromosomes 15q and 17q.

Rearrangements of chromosome 15q11 are the most frequently reported chromosomal abnormalities in autism, particularly in cases where the autistic phenotype is accompanied by severe mental retardation or seizures (Flejter et al 1996, Rineer et al 1998, Wolpert et al 2000). This region is of particular interest as it is subject to imprinting and contains the genes involved in Prader Willi and Angelman syndromes. Chromosome 15q candidate genes include the γ-aminobutyric acid ($GABA_A$) receptor gene cluster (Cook et al 1998, Maestrini et al 1999) and the ubiquitin protein ligase E3A gene (*UBE3A*) (Veenstra-VanderWeele et al 1999).

Many functional studies of autism have focussed upon candidate genes selected from the serotonin pathway. An increased level of serotonin is often detected in the urine of autistic patients (Piven et al 1991), and some individuals show an amelioration of behavioural symptoms when treated with medication that targets serotonin activity (Gordon et al 1993, McDougle et al 1996). The serotonin receptor gene (*5-HTT*) on chromosome 17q is probably the best studied serotonin gene with regards to autism (Cook et al 1997, Klauck et al 1997, Zhong et al 1999). Other genes include the serotonin receptors *5-HT7* on chromosome 10 (e.g. Lassig et al 1999) and *HTR2A* on chromosome 13q (e.g. Herault et al 1996, Veenstra-VanderWeele et al 2002).

It should be noted, however, that the majority of genome screen studies have failed to find linkage to chromosomes 15 and 17, and that investigations of the candidate genes mentioned above have yielded variable degrees of association. It is generally accepted that the chromosome 15 and 17 loci may be more relevant to a limited or specific sub-set of autistic patients.

3.2iv - Dyslexia

Dyslexia is defined as an unexplained difficulty in the acquisition of written language skills despite adequate intelligence and opportunity. Relatives of probands affected by dyslexia are subject to an increased risk of language impairment (Gallagher et al 2000), and studies of children selected for language impairments often report a high incidence of literacy problems

(Tallal et al 1989). Furthermore, both dyslexics and language-impaired children tend to experience non-linguistic deficits which are common to both disorders (e.g. clumsiness and poor timing) (Bishop & Edmundson 1987, Powell & Bishop 1992, Habib 2000). The strong links between dyslexia and SLI have often led to the speculation that language impairments and reading disabilities may represent different manifestations of similar neurological deficits (Snowling et al 2000). However, it is apparent that language problems are neither sufficient nor necessary to result in a dyslexic phenotype (or vice versa), and any aetiological interactions between SLI and dyslexia are predicted to be complex in nature.

For the mainpart, genetic studies of dyslexia have relied upon a candidate gene approach basing gene selection upon various theoretical conjectures. These studies have concentrated on chromosomes 1, 2p, 6p and 15q.

The first linkage study of dyslexia (Smith et al 1983) revealed a region on chromosome15q (*DYX1*), which has since been both replicated (Grigorenko et al 1997, Schulte-Körne et al 1998), and disputed (Bisgaard et al 1987, Cardon et al 1994).

Links between dyslexia and autoimmune disorders (Galaburda et al 1985, Pennington et al 1987, Hugdahl et al 1990) have concurrently fuelled genetic studies in regions important to the immune system, principally the HLA region on chromosome 6p. Whether, the hypothesised links between autoimmunity and dyslexia are true or not, the HLA region (*DYX2*) has proven to be the best replicated locus in dyslexia thus far (Cardon et al 1994, Cardon et al 1995, Grigorenko et al 1997, Fisher et al 1999, Gayán et al 1999, Kaplan et al 2002), with only two studies to date failing to find linkage to this region (Field & Kaplan 1998, Sawyer et al 1998). A second genomic locus that has immunological links and has been cited with respect to dyslexia is the area around the Rhesus factor (Rh) locus on chromosome 1p. Rabin et al (1993) found linkage to this region in a collection of three-generation pedigrees apparently affected by a monogenic form of dyslexia, and Froster et al (1993) reported a translocation involving this region which co-segregated with a delayed speech and dyslexic phenotype.

In 1999 Fagerheim et al performed the first genome-wide scan for reading disability using 36 individuals selected from a large Norwegian family in which dyslexia appeared to be inherited

as an autosomal dominant disorder. A maximum LOD score of 4.32 was obtained for a region on chromosome 2p15-16 (*DYX3*) which contained a 4cM haplotype that co-segregated with the dyslexia phenotype.

However, to date, there has been only one full genome screen that has used nuclear families affected by the more typical, genetically complex forms of dyslexia (Fisher et al 2002a). This screen involved 208 families selected from two independent sites, one in the US, and one in England. Phenotype data reflecting several aspects of the reading process (phonological awareness (PA), phonological decoding (PD), orthographic decoding (OD) and reading ability (RA)) were collected from each group. However, logistical constraints meant that the exact nature of these quantitative measures differed slightly between the two cohorts and the US and UK groups were therefore separated for the genome screen analysis. Linkage analyses were performed using both singlepoint and multipoint VC and DeFries-Fulker methods.

In general, the level of concordance between the linkage results for the two groups is strikingly low, with only one area (18p11 – *DYX*QTL18) being identified by both samples (see table 3.5).

Interestingly, whilst the UK sample showed a moderate level of linkage to chromosome 6p (*DYX2*) (see table 3.5), the US group showed no significant linkage within this region. Both samples yielded modest linkages to chromosome 2p (see table 3.5) and in the US, this region was found to overlap with the *DYX3* locus previously described by Fagerheim et al (1999).

	UK			US		
Locus	**Empirical p**	**Trait**	**Analysis**	**Empirical p**	**Trait**	**Analysis**
2p15-16	**0.0007**	**OC**	**Singlepoint DF**	**0.001**	**OC**	**Multipoint DF**
				0.001		**Singlepoint VC**
3q13	nr	—	—	0.0003	PA	Multipoint DF
6p21	0.00001	PD	Singlepoint DF	0.002	PD	Singlepoint DF
13q22	nr	—	—	0.001	OC	Multipoint VC
18p11	**0.00001**	**RA**	**Multipoint VC**	**0.0004**	**RA**	**Multipoint DF**
				0.0004		**Singlepoint DF**
18q22	nr	—	—	0.00003	RA	Multipoint DF
21q21-22	nr	—	—	0.00003	OC	Multipoint DF
Xq26	0.001	RA	Multipoint VC	nr	—	—

Table 3.5 - Major Locl Detected by Dyslexia Genome Screen : *nr = Result not recorded (p<0.01). Empirical probabilities were derived for all linkages using simulations. The probability is shown for the highest single result within the given region. The trait given is the one with which the most significant result was obtained. For trait codes please see text (section 3.2iv). The type of analysis shown is that which yielded the most significant result. Those results shown in bold denote a similar region of linkage that was consistent across the UK and US screens.*

In conclusion, although the first genome-wide scan for dyslexia endorsed previous findings on chromosomes 6p and 2p to a certain extent, no such support was found for the chromosome 1 and 15 regions. In addition, the use of this non-biased genome-wide approached allowed the identification of a novel locus on chromosome 18, which appeared to have a stronger influence upon the dyslexia phenotype than any of the loci which have previously been studied.

3.2v – Attention-Deficit Hyperactivity Disorder (ADHD)

Attention Deficit Hyperactivity Disorder (ADHD) (MIM143465) is characterised by developmentally inappropriate behaviour which involves inattention, hyperactivity and impulsivity. Symptoms include restlessness, distractibility, forgetfulness and risk-taking behaviour (American Psychiatric Association 1994). As mentioned in the introduction, there is a strong co-morbidity between SLI and ADHD and many children with language impairments also suffer from additional attentional problems. However, like dyslexia, the overlap between ADHD and SLI is not complete and the development of one disorder is not completely reliant upon the existence of the other.

The majority of genetic studies of ADHD have opted for a functional approach, studying candidate genes selected on the basis of neurochemical deficits associated with ADHD. ADHD is commonly treated with chemical stimulants (e.g. amphetamines) which alter dopamine transport within the brain (Spencer et al. 1996) and mice that are knocked-out for genes in the dopmaine pathway demonstrate extreme altered states of activity (Xu et al 1994, Baik et al 1995, Accili et al 1996, Giros et al 1996). Thus the majority of ADHD candidate genes are involved in the dopamine pathway. Positive associations to ADHD have been reported by a number of researchers at both the dopamine transporter gene (*DAT1*) locus on chromosome 5p and a dopamine receptor gene (*DRD4*) locus on chromosome 11p (see table 3.6). However, even at these loci, the strength of the association is often tenuous, and there are a significant number of groups who have failed to detect the expected relationship.

Position	Positive association	Negative association
Interleukin-1 Receptor Antagonist (*IL-1Ra*)		
2q14	Segman et al 2002	
Dopamine Receptor D3 (*DRD3*)		
3q13		Barr et al 2000(a), Payton et al 2001, Muglia et al 2002
Dopamine Receptor D5 (*DRD5*)		
4p16	Daly et al 1999, Barr et al 2000(b), Tahir et al 2000(a)	Payton et al 2001
Dopamine Transporter (*DAT1*)		
5p15	Cook et al 1995, Gill et al 1997, Waldman et al 1998, Daly et al 1999, Barr et al 2001(a), Curran et al 2001	Palmer et al 1999, Holmes et al 2000, Roman et al 2001
Dopa Decarboxylase (*DDC*)		
7p12		Hawi et al 2001
Dopamine Beta Hydroxylase (*DβH*)		
9q34	Daly et al 1999, Roman et al 2002	Payton et al 2001
Dopamine Receptor D4 (*DRD4*)		
11p15	LaHoste et al 1996, Smalley et al 1998, Swanson et al 1998, Faraone et al 1999, Barr et al 2000(c), Holmes et al 2000, Muglia et al 2000, Sunohara et al 2000, Tahir et al 2000(a), Faraone et al 2001, Roman et al 2001, Holmes et al 2002	Castellanos et al 1998, Eisenberg et al 2000, Hawi et al 2000(a), Kotler et al 2000, Barr et al 2001(b), Payton et al 2001, Todd et al 2001
Tyrosine Hydroxylase (*TH*)		
11p15		Comings et al 1995, Barr et al 2000(c), Payton et al 2001
Neuronal Nicotinic Acetylcholine Receptor α7 Subunit (*CHRNA7*)		
15q14		Kent et al 2001
Norepinephrine Transporter (*NET1*)		
16q12		Barr et al 2002
Serotonin Transporter (*5-HTT*)		
17q11	Manor et al 2001	Retz et al 2002

Table 3.6 - Candidate Gene Studies of ADHD

Position	Positive association	Negative association
Synaptosomal Associated Protein of 25kDa (*SNAP-25*)		
20p11	Barr et al 2000(d), Mill et al 2002	
Catechol-O-Methyltransferase (*COMT*)		
22q11	Eisenberg et al 1999	Barr et al 1999, Hawi et al 2000(b), Tahir et al 2000(b), Payton et al 2001
Monoamine Oxidase A (*MAOA*)		
Xp11	Jiang et al 2001	Payton et al 2001
Monoamine Oxidase B (*MAOB*)		
Xp11		Jiang et al 2001

Table 3.6 - Candidate Gene Studies of ADHD (cont.)

To date, only one genome screen has been performed for ADHD (Fisher et al 2002). This screen involved 126 affected sib-pairs and used both broad and narrow definitions of affection status alongside a quantitative measure based on the number of DSM-IV ADHD criteria that the individual met.

The most striking result from the genome screen was the disparity between the qualitative and quantitative analyses (see table 3.7). Multipoint analyses of the qualitative data yielded an MLS of 1.66 (p=0.0028) on chromosome 10q using a broad definition of affection, whilst the QTL mapping remained below 1.0 for this region. Similarly, the quantitative genome screen highlighted a region on chromosome 12p (multipoint MLS 2.6 p=0.0003), but the qualitative analysis remained below 0.79 in this area. Singlepoint analyses yielded a qualitative MLS of 2.10 for marker D5S418 (broad affection status), and a quantitative MLS of 1.59 for marker D12S1725. Again the LODs for the alternative analyses remained below 1.0 at these markers. Across the entire genome, no one single region was consistently highlighted as being important across both screens (see table 3.7).

	Singlepoint Analyses			Multipoint Analyses		
	Qualitative		Quantitative	Qualitative		Quantitative
	Narrow	Broad	QTL	Narrow	Broad	QTL
5p12	**1.88**	**2.10**	nr	1.05	1.04	nr
10q26	0.98	1.11	nr	1.32	**1.66**	nr
12p13	nr	nr	**1.59**	0.79	nr	**2.60**
12q23	1.11	0.95	nr	**1.54**	1.16	nr
16p13	nr	nr	nr	0.75	**1.51**	nr

Table 3.7 - Genomic Regions Yielding LOD Scores > 1.5 in ADHD Genome Screen : *All available LOD scores are given for any region that gave a LOD score above 1.5 under any of the analyses performed. All LOD scores of above 1.5 are given in bold. nr = LOD score not recorded*

In conclusion, the genome screen indicated that, rather than involving a single major gene, ADHD is likely to be caused by combinations between several loci, which individually have only a minimal impact upon the disorder. However, because this is the first such screen, it is hard to evaluate the effect of the identified QTLs with respect to the ADHD phenotype.

It is worth note that the ADHD genome screen reported the exclusion of 96% of the genome at λ_s of ≥3. Even under the strictest criteria, all of chromosomes 18, 19, 21, 22 and X, and thus the candidate genes *COMT*, *MAOA* and *MAOB* could be excluded (see table 3.8).

	Strictest Exclusion Criteria (Narrow Affection Status, λs=2)		Most Relaxed Exclusion Criteria (Broad Affection Status, λs=3)	
Percentage of Chromosome Excluded	**Chromosomes Included in Exclusion Category**	**Candidate Genes on These Chromosomes**	**Chromosomes Included in Exclusion Category**	**Candidate Genes on These Chromosomes**
0-20%				
21-30%	12			
31-40%	9, 16, 20	*DβH, NET1, SNAP-25*		
41-50%				
51-60%	1, 7, 8, 14, 17	*DDC, 5-HTT*		
61-70%	2, 5, 10, 11	*IL-1Ra, DAT1, DRD4, TH*		
71-80%	3, 4, 15	*DRD3, DRD5, CHRNA7*		
81-90%	6, 13		2, 11, 12, 16, 17	*IL-1Ra, DRD4, TH, NET1, 5-HTT*
91-99%			5, 7, 8, 10	*DAT1, DDC*
100%	18, 19, 21, 22, X	*COMT, MAOA, MAOB*	1, 3, 4, 6, 9, 13, 14, 15, 18, 19, 20, 21, 22, X	*DRD3, DRD5, DβH, CHRNA7, SNAP-25, COMT, MAOA, MAOB*

Table 3.8 - Loci Excluded by the ADHD Screen

3.2vi - Molecular Genetic Studies of SLI

To date, aside from the present study, there have been only two linkage studies performed for language impairment (Fisher et al 1998, Bartlett et al 2002).

Fisher et al (1998) performed a parametric genome screen using a single, large pedigree (known as KE), in which a severe speech and language impairment is inherited as an autosomal dominant trait. A full genome screen within the KE family found linkage to a region of 7q (known as *SPCH1*) which was found to overlap with the *AUTS1* region for autism (Fisher et al 1998) (see section 3.2iii 'Autism'). Subsequent candidate gene analyses later allowed the identification of the gene mutated within this family as *FOXP2* (Lai et al 2001). For a full description of the KE phenotype and the relevant gene please refer to chapter 4 ('Candidate gene analyses'). Aside from 7q, no specific chromosomal regions have been investigated with regards to speech and language impairments.

Bartlett et al (2002) also used a parametric design, with binary indices of affection status, to investigate language impairment within 86 individuals collected from branches of five Canadian families. Although not inter-related or a population isolate, these families were selected from a relatively homogeneous sample of individuals with Celtic ancestry. All families were assessed with a comprehensive battery of tests covering expressive (Test of Language Development (TOAL:2, TOLDP:2 or TOLDI:2)) and receptive (Token test) language skills, IQ (Wechsler Intelligence Scale (WISC, WAIS, WIPPSI or WAIS)), non-word (word attack) and single word (word identification) reading and oral structure and function. All individuals with hearing impairments, motor impairments, autism, schizophrenia or other known neurological disorders were excluded. The scores on these tests were then used to derive three diagnostic categories.

A statistical cut-off was applied to spoken language scores and used to define a group of 'language-impaired' individuals, a non-word reading-IQ discrepancy criterion was employed to identify a group of 'reading impaired' individuals, and a set of more relaxed conditions (e.g. a history of speech therapy or reading intervention) were used to specify a 'clinically impaired' group. Note that these diagnostic cohorts were not mutually exclusive and many individuals were classified as affected under all three schedules. Each of these three diagnostic

classification groups was then analysed for linkage under both a dominant and a recessive model of inheritance. Model parameters were estimated assuming a population prevalence of 7% and genome-wide empirical p-values were determined via the simulation of 1,500 sets of 400 markers. Genome-wide linkage analyses revealed two regions of linkage, one on chromosome 13q21 and a second on chromosome 2p22. The chromosome 13q peak was seen under a recessive model of reading impairment and reached a significant level of linkage with a four-point maximum LOD score of 3.92 (genome-wide $p<0.01$) at marker D13S800. Linkage to chromosome 2 was found using a recessive model of language impairment and attained a four-point MLS of 2.79 (genomewide $p<0.06$).

Interestingly, both the chromosome 13 and the chromosome 2 loci identified by Barlett et al (2002) coincide with regions that have previously been linked to both autism and/or dyslexia. In a genome-wide screen for autism, the Collaborative Linkage Study of Autism (CLSA 1999) obtained their strongest linkage result (MLS=3.0, p=0.008) with the same chromosome 13 marker (D13S800) as Bartlett et al (2002). Furthermore, when a follow-up study stratified the CLSA cohort according to the language status of the proband, they found that the linkage signal on chromosome 13 came predominantly from the language-abnormal (i.e. onset of phrase speech > 36 months) proband subset (Bradford et al 2001). In addition, a recent meta-analysis of four genome screens for autism indicated that chromosome 13q formed the second most significant region of linkage (after 7q31) across the entire genome (p=0.0006) (Badner & Gerschon 2002). In a genome-wide screen for dyslexia, Fisher et al (2001) found that this region of chromosome 13q was also linked (p=0.001) to a measure of forced orthographic choice (i.e. rapid discrimination between a target word and a phonologically identical word (e.g. pane vs. pain)) within a sample of American reading impaired families. As mentioned in section 3.2iv ('Dyslexia'), this study also found modest linkage to the *DYX3* locus on chromosome 2p, which corresponds to the region identified by Bartlett et al (2002). The overlaps between these genome screens provide further endorsement for the complexity of the interactions between SLI, autism and dyslexia (see section 3.2ii 'Genome screens of disorders related to SLI') and may indicate that the phenotypic attributes shared by these neurodevelopmental disorders are reflected by the presence of a common genetic aetiology.

As demonstrated by dyslexia and ADHD (see sections 3.2iv 'Dyslexia' and 3.2v 'Attention-Deficit Hyperactivity Disorder (ADHD)'), non-parametric genome screens can be crucial to the study of complex disorders, as they make no prior assumptions about the basis of the disease nor target specific chromosomal regions for analysis. The work contained in this thesis details a genome-wide screen for generalised and genetically complex speech and language impairments and describes two novel loci which appear to have a significant genetic effect upon the development of SLI. This work represents a major step in the clarification of the genetic mechanisms behind SLI and may lead to a better understanding of the processes involved in language acquisition whilst facilitating better diagnosis and treatment for individuals with language impairments.

3.3 GENOME SCREEN RESULTS

3.3i – Genome Screen Results

The DNA collection, PCR and genotyping procedures have all been previously described in chapter 2 ('methods'). Likewise, the genome screen design and linkage analyses of the data is described both above (section 3.1 'Gene mapping in complex disorders') and in section 2.7 ('Linkage analysis').

In short, 473 individuals (98 families, 219 total sib-pairs) were recruited from two independent centres (Guys Hospital and the Cambridge Language and Speech Project (CLASP)) and typed at 500 microsatellite markers. Genotyping was performed in two stages. An initial set of 400 markers was taken from the ABI PRISM LMS2-MD10 collection. This was followed by a second wave of 101 markers, chosen to fill in gaps in information content across the genome. The final average information content across the genome was estimated at 71%.

Three quantitative measures of language abilities (CELF receptive language score (RLStrans) and expressive language score (ELStrans) and non-word repetition (NWRtrans)) were assessed for linkage across the entire genome, using both singlepoint and multipoint unweighted Haseman Elston (HE) and Variance Components (VC) methods within the GH2.0 package. X chromosome HE analyses were performed within the MAPMAKER/SIBS package. Simulations indicated that the lack of multiple sib-pair weighting within the HE

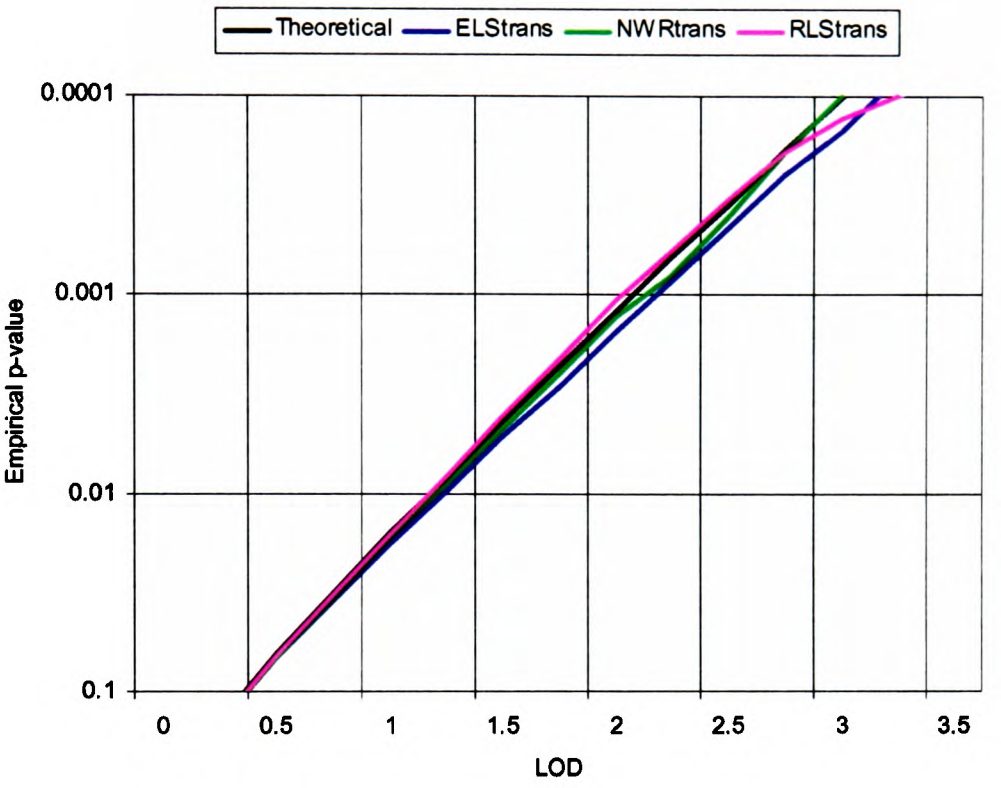
analyses did not result in increased type I errors in our data set (see figure 3.2a).

In areas which showed a suggestive level of linkage, an additional measure of non-verbal IQ (PIQ) was analysed in an identical manner to the other three traits. Although it was felt that the exclusion of all individuals with a low performance IQ (<80) would sufficiently control for the influence of general intelligence upon language ability, it was envisaged that this extra analysis would enable the direct quantification of these effects upon the linkages found.

Empirical probability distributions were derived from the random simulation of genotype data within a preserved pedigree and phenotype framework. Simulations were performed for all three language traits (ELStrans, RLStrans and NWRtrans) under HE and VC analyses. In general, the HE empirical distributions coincided with those of the theoretical distribution for each phenotype tested (see figure 3.2a). However, for VC analysis, whilst ELStrans and RLStrans behaved as predicted by theory, the theoretical p-values for NWRtrans were found to be over-conservative (see figure 3.2b). Thus the empirical, pointwise significance of all NWRtrans LOD scores is increased above the theoretical probability of obtaining LODs of that value. Empirical p-values are reported for all regions yielding LOD scores of above 1.0 (see table 3.3).

The results of the genome screen can be found below. Multipoint HE and VC plots are shown in figures 3.3 and 3.4, and the results of all analyses in regions that yielded LOD scores of above 1.0 are detailed in table 3.9. The plots for all individual chromosomes can be found in appendix J.

A. Haseman Elston (unweighted)



B. Variance Components

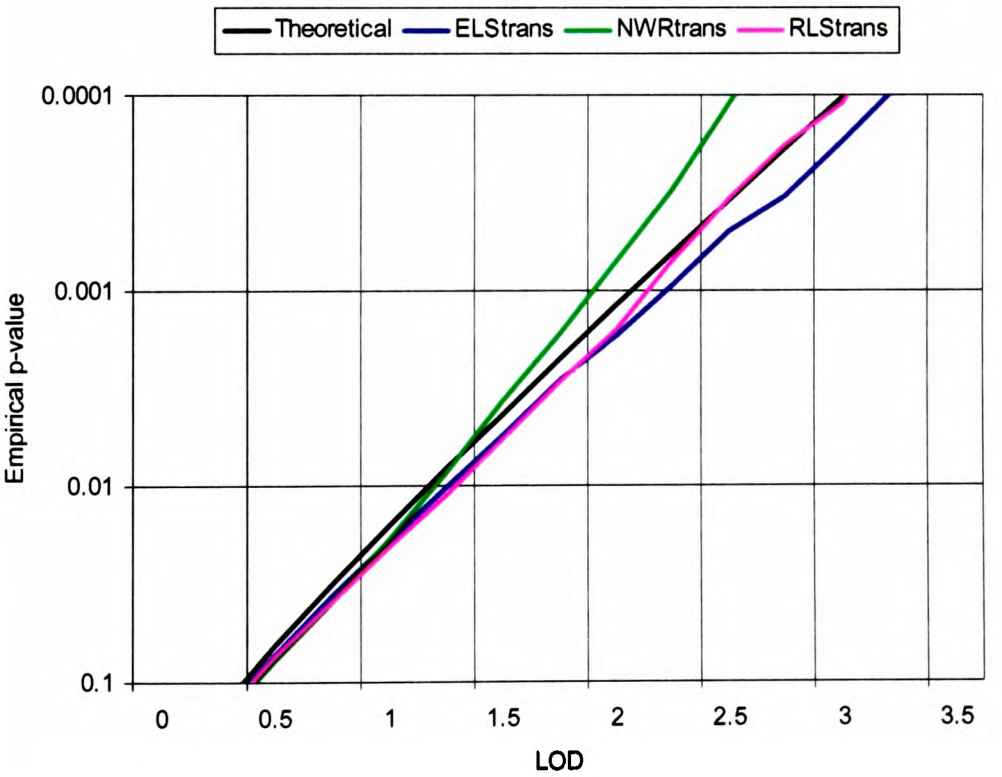


Figure 3.2 - LOD Significance Distributions for Each Measure Used in the Genome Screen : *Empirical p-values may differ from theoretical p-values due to sample-specific deviations from assumptions critical to different analyses (see section 2.7iii). Comparisons between empirical and theoretical probability distributions thus allow the quantification of these deviations that are specific to each phenotype. Black lines show the theoretical probability for any given LOD under the appropriate analyses. Coloured lines represent phenotype-specific empirical probabilities for any given LOD*

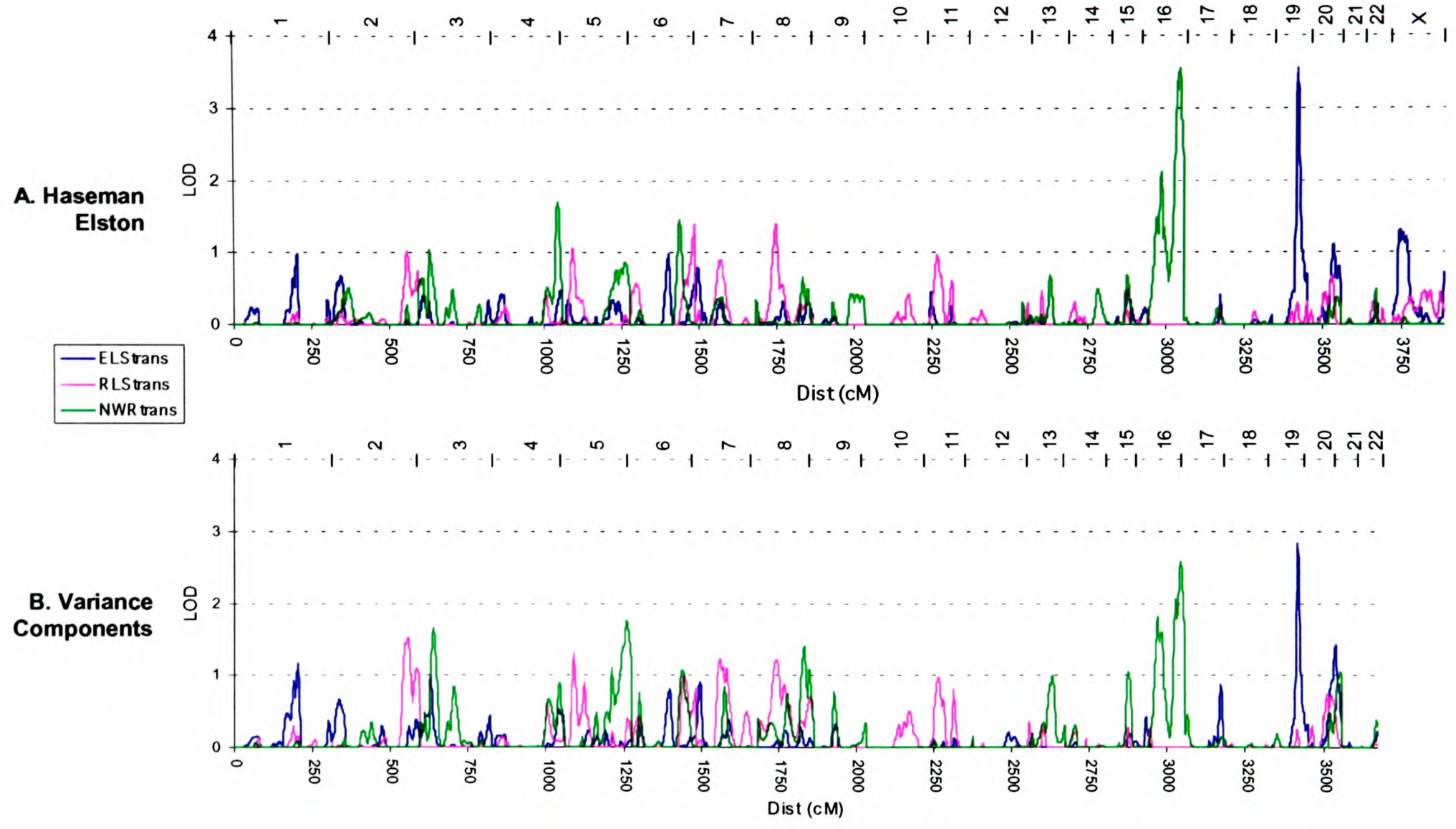


Figure 3.3 - Genome-Wide Multipoint Linkage to SLI (organised by analysis type) : *The x-axis shows cumulative distance (in Haldane centiMorgans) and chromosome numbers are displayed along the top of the graph. Note the magnitude of the chromosome 16 and 19 linkages in relation to the background. No analysis of chromosome X was performed for variance components (see section 2.7ii)*

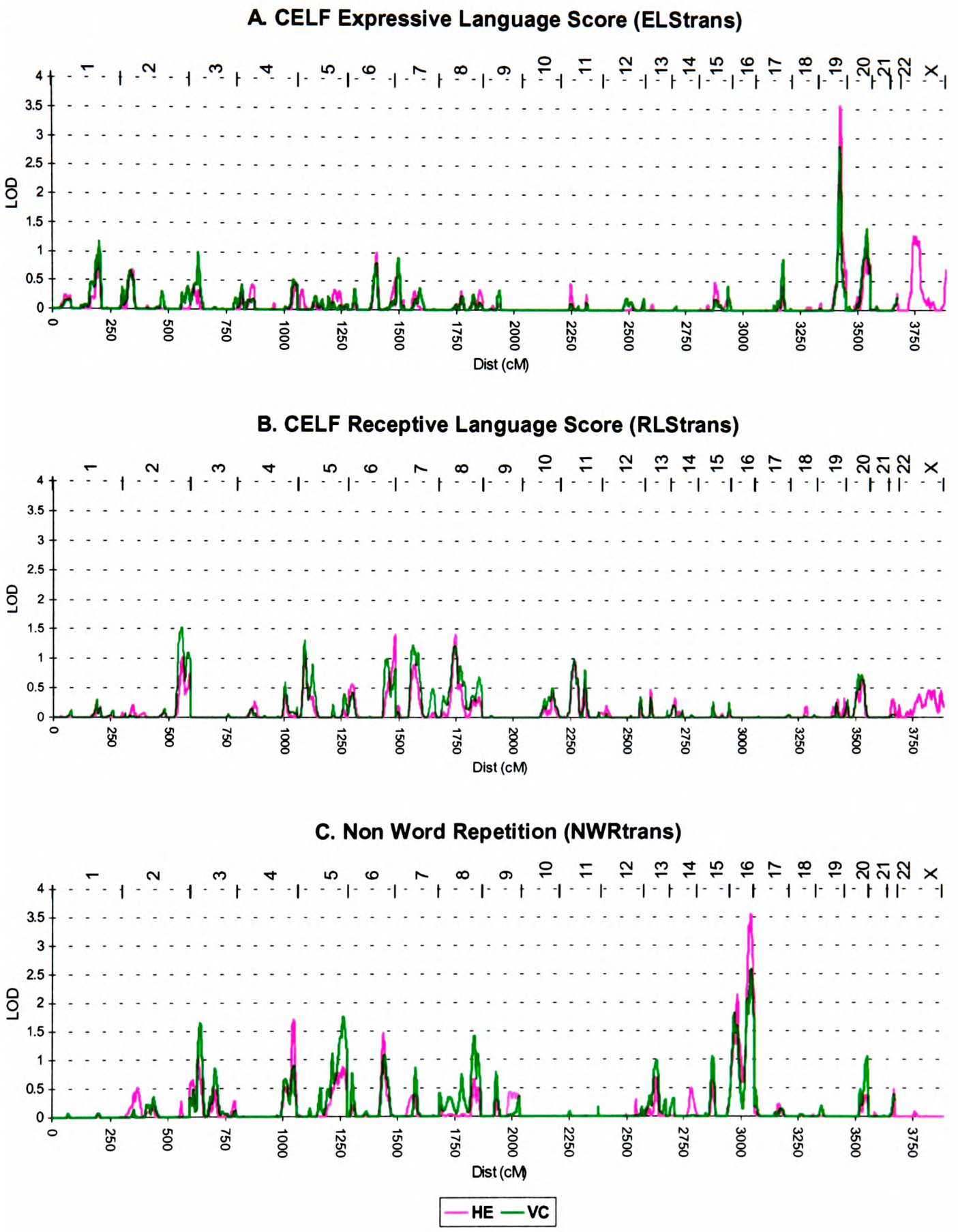


Figure 3.4 - Genome-Wide Multipoint Linkage to SLI (organised by phenotype) : *The x-axis shows cumulative distance (in Haldane centiMorgans), and chromosome numbers are displayed along the top of the graph. Note the magnitude of the ELStrans and NWRtrans linkages in relation to the background*

		MLS				
Chr	**Position**	**HE Multipoint**	**VC Multipoint**	**HE Singlepoint**	**VC Singlepoint**	**Trait**
1	1q24	0.98 (0.01924)	1.16 (0.01246)	1.25 (D1S218)	1.25 (D1S218)	ELS trans
2	2q36	1.03 (0.01468)	1.52 (0.00574)	0.98 (D2S206)	1.01 (D2S338)	RLS trans
3	3p24	1.05 (0.01534)	1.64 (0.00238)	1.10 (D3S2338)	1.42 (D3S1266)	NWR trans
4	4q35	1.70 (0.00296)	0.89 (0.02632)	1.00 (D4S1535)	0.63 (D3S1535)	NWR trans
5	5p15	1.06 (0.01310)	1.32 (0.00950)	1.88 (D5S416)	2.12 (D5S416)	RLS trans
5	5q34	0.86 (0.02506)	1.76 (0.00160)	0.63 (D5S2073)	1.76 (D5S1960)	NWR trans
6	6q25	1.46 (0.00522)	1.07 (0.01554)	1.23 (D6S441)	0.86 (D6S441)	NWR trans
7	7q11	0.90 (0.02146)	1.24 (0.01544)	0.61 (D7S684)	1.06 (D7S684)	RLS trans
8	8p12	1.41 (0.00508)	1.22 (0.01220)	1.00 (D8S260)	1.56 (D8S260)	RLS trans
8	8q24	0.66 (0.04378)	1.40 (0.00534)	0.81 (D8S272)	1.10 (D8S272)	NWR trans
15	15q22	0.69 (0.03990)	1.05 (0.01620)	0.80 (D15S153)	1.19 (D15S153)	NWR trans
16	**16q24**	**3.55 (0.00003)**	**2.57 (0.00008)**	**2.77 (D16S516)**	**2.29 (D16S516)**	**NWR trans**
19	**19q13**	**3.55 (0.00004)**	**2.84 (0.00027)**	**2.49 (D19S908)**	**2.22 (D19S908)**	**ELS trans**
20	20q13	1.11 (0.01414)	1.43 (0.0063)	0.93 (D20S171)	0.67 (D20S171)	ELS trans
X	Xp11	1.30	—	—	—	ELS trans

Table 3.9 - LOD Scores >1.0 : *Multipoint LOD scores are given with empirical p values, calculated by simulations (as described in section 2.7iii), in brackets. Singlepoint LOD scores are given for the marker with the highest score in the area of linkage. These are given as a guide to the level of singlepoint support given to the multipoint result. The trait given is the phenotype for which the highest LOD is seen. Other traits may also show linkage to the same region at lower levels. Figures in bold denote those results considered to be suggestive*

As shown above (see figures 3.3 and 3.4 and table 3.9), genome-wide analyses highlighted two prominent areas of linkage; one on chromosome 16 and one on chromosome 19. Whilst several other regions were found to have LODs of above 1.0 (see table 3.9), only these two regions exceeded the threshold of 'suggestive linkage', as proposed by Kruglyak and Lander (i.e. a LOD of 2.2 or greater – see section 3.1viii 'Variance Components analysis').

Furthermore, they did so under all four types of analysis performed (multipoint HE and VC and singlepoint HE and VC).

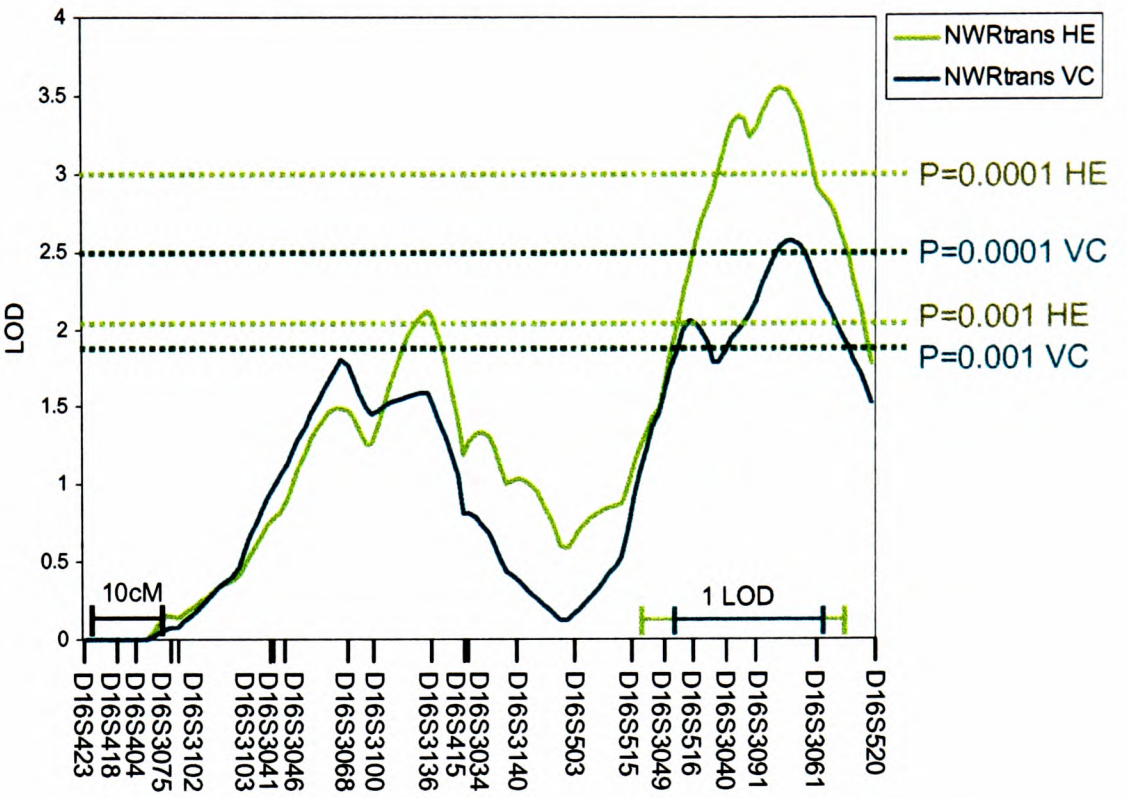
The chromosome 16 locus is linked to the NWRtrans trait and spans approximately 40cM of 16q, from D16S515 to D16S520 (see figure 3.5a). Whilst the HE maximum LOD score (MLS) for this region reached 3.55, the VC analysis yielded a somewhat lower MLS of 2.57. However, empirical probability distributions drawn from simulated data, indicated a general deflation of VC LOD scores for the NWRtrans trait (see figure 3.3b). In 100,000 simulations a VC LOD of above 2.57 was seen only eight times (i.e. pointwise empirical p=0.00008), and thus is consistent with the HE result (empirical p=0.00003). Furthermore, chromosome 16 yielded the most significant singlepoint result of the genome (D16S516 LOD=2.77), with a cluster of three markers directly under the peak of linkage giving singlepoint LOD scores of above 1.5 (see table 3.10).

Across the chromosome 16q linkage area, the LOD scores for ELStrans and RLStrans remained below 0.42 and 0.27 respectively (data not shown).

The chromosome 19 locus is linked to ELStrans and covers approximately 30cM of 19q, from D19S220 to D19S418 (see figure 3.5b). Again, this QTL was evident both under HE and VC multipoint analyses (HE LOD=3.55, VC LOD=2.84) and was further supported by the singlepoint analyses, where two adjacent markers yielded LOD scores exceeding 1.5 (see table 3.11). Simulations indicated that the ELStrans measure behaves as predicted by theory for both HE and VC analyses, and therefore empirical p-values can be taken to be representative of nominal p-values at the chromosome 19 locus (see figure 3.2).

For chromosome 19q, the highest LOD scores found for RLStrans and NWRtrans were 0.33 and 0.20 respectively (data not shown).

A. Chromosome 16



B. Chromosome 19

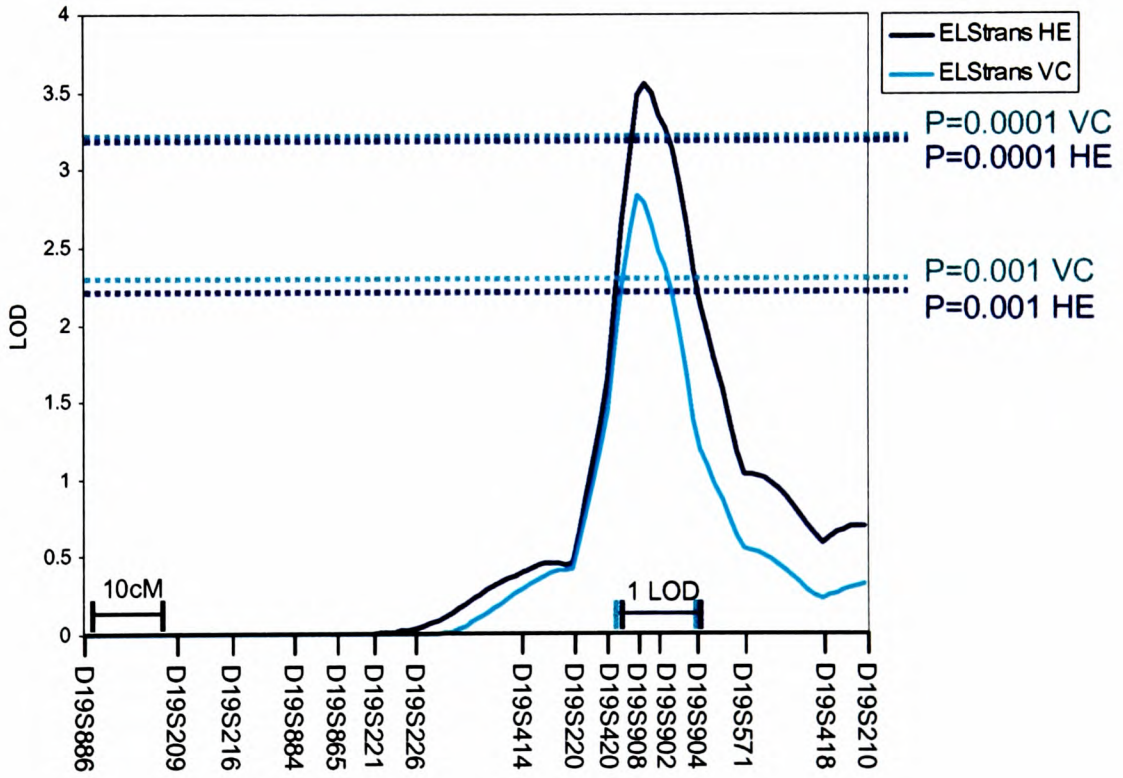


Figure 3.5 - Multipoint Linkages to Chromosomes 16 and 19 : *For both figures, the x-axis shows the positions of all markers typed, and a 10cM (Haldane) bar is given for reference. 1 LOD intervals are given for both the HE and VC peaks. Light green/dark blue dotted lines show the LOD thresholds for empirical p values of 0.001 and 0.0001 under HE analysis. Dark green/light blue lines show the same LOD thresholds under VC analysis*

		HE			VC		
Dist (cM)	Marker	ELS trans	RLS trans	NWR trans	ELS trans	RLS trans	NWR trans
9.8	D16S423	0.24	0.00	0.00	0.40	0.00	0.00
14.4	D16S418	0.00	0.00	0.00	0.04	0.00	0.11
17.4	D16S404	0.14	0.00	0.00	0.02	0.00	0.00
23.2	D16S3075	0.55	0.156	0.42	0.08	0.10	0.01
24.2	D16S3102	0.00	0.85	0.16	0.12	0.71	0.07
34.4	D16S3103	0.47	0.00	0.14	0.00	0.00	0.22
40.6	D16S3041	0.00	0.74	0.40	0.00	0.00	0.00
42.0	D16S3046	0.00	0.00	0.57	0.00	0.00	0.64
52.1	**D16S3068**	0.00	0.00	1.09	0.00	0.03	**1.54**
56.3	D16S3100	0.00	0.00	0.00	0.00	0.00	0.01
65.8	D16S3136	0.00	0.04	0.73	0.00	0.03	0.84
71.4	D16S415	0.00	0.00	0.47	0.00	0.00	0.35
71.5	D16S3034	0.00	0.00	0.88	0.00	0.00	0.54
78.8	D16S3140	0.00	0.00	0.60	0.01	0.00	0.50
88.3	D16S503	0.00	0.076	0.24	0.00	0.12	0.03
97.7	D16S515	0.00	0.00	0.00	0.00	0.00	0.00
104.0	D16S3049	0.00	0.00	0.00	0.00	0.00	0.04
108.2	**D16S516**	0.00	0.00	**2.78**	0.00	0.01	**2.29**
113.2	**D16S3040**	0.00	0.00	**2.24**	0.00	0.00	**1.58**
119.0	**D16S3091**	0.00	0.00	**1.95**	0.00	0.00	1.41
130.0	D16S3061	0.00	0.00	1.49	0.00	0.01	0.99
139.4	D16S520	0.00	0.00	1.20	0.00	0.00	1.21

Table 3.10 - Singlepoint LOD Scores for Chromosome 16 : *Cumulative distances across the chromosomes are given in Haldane centiMorgans. All LOD scores above 1.5 are marked in red bold text*

		HE			VC		
Dist (cM)	Marker	ELS trans	RLS trans	NWR trans	ELS trans	RLS trans	NWR trans
0.0	D19S886	0.00	0.22	0.00	0.00	0.03	0.00
12.1	D19S209	0.15	0.00	0.14	0.10	0.00	0.44
20.7	D19S216	0.00	0.00	0.00	0.00	0.00	0.12
28.7	D19S884	0.00	0.00	0.00	0.00	0.01	0.00
34.5	D19S865	0.00	0.00	0.00	0.00	0.00	0.00
39.4	D19S221	0.11	0.00	0.00	0.01	0.00	0.00
45.5	D19S226	0.24	0.06	0.00	0.03	0.14	0.00
60.3	D19S414	0.09	0.48	0.00	0.01	0.06	0.04
67.9	D19S220	0.16	0.00	0.00	0.19	0.00	0.00
72.7	D19S420	1.01	0.80	0.00	0.93	0.85	0.01
76.7	**D19S908**	**2.49**	0.04	0.08	**2.22**	0.08	0.33
79.7	**D19S902**	**1.74**	0.75	0.00	0.92	0.37	0.00
85.4	D19S904	0.81	0.01	0.00	0.09	0.00	0.00
91.8	D19S571	0.12	0.00	0.00	0.20	0.00	0.00
102.9	D19S418	0.43	0.48	0.00	0.13	0.00	0.00
109.5	D19S210	0.74	0.32	0.00	0.65	0.05	0.00

Table 3.11 - Singlepoint LOD Scores for Chromosome 19 : *Cumulative distances across the chromosomes are given in Haldane centiMorgans. All LOD scores above 1.5 are marked in red bold text*

Analysis of PIQ on chromosomes 16 and 19 demonstrated only a minimal level of linkage in these regions. Thus general intelligence effects probably do not contribute to either of the language-related linkages, and both loci can be expected to reflect language specific influences (see figure 3.6).

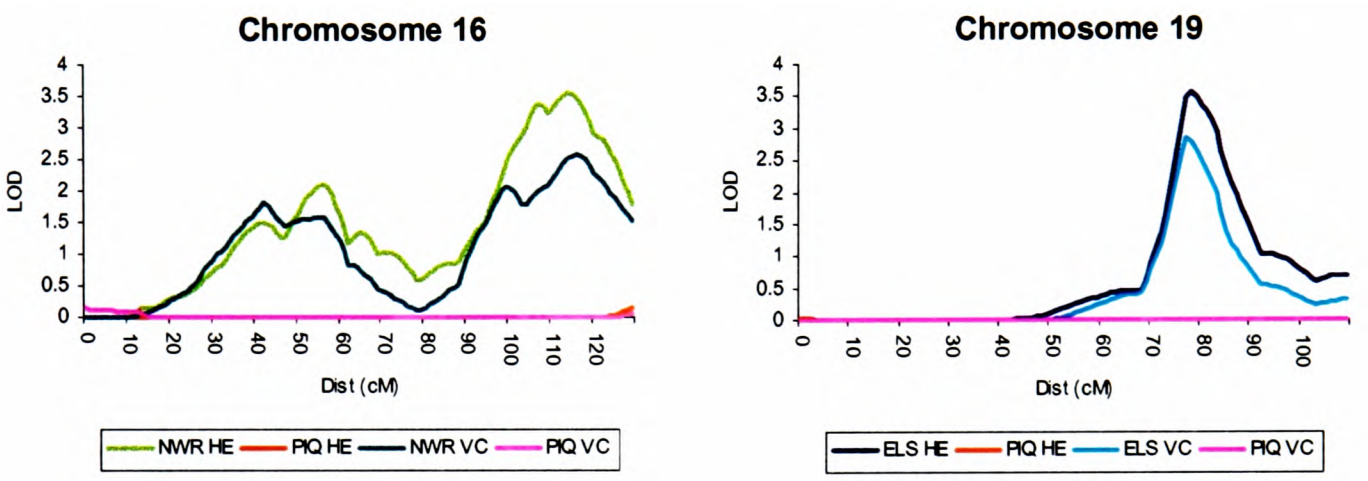


Figure 3.6 - HE and VC Multipoint Linkage to PIQ on Chromosomes 16 and 19 : *Under both HE and VC analyses, the PIQ LOD score did not exceed 0.15 on either chromosome*

In order to quantify the contribution of the Guys and Cambridge groupings to the overall linkages, we divided the genome screen sample into its constituent groupings and re-analysed chromosomes 16 and 19. Figure 3.7 shows that both the clinical and epidemiological cohorts contribute equally to both peaks of linkage.

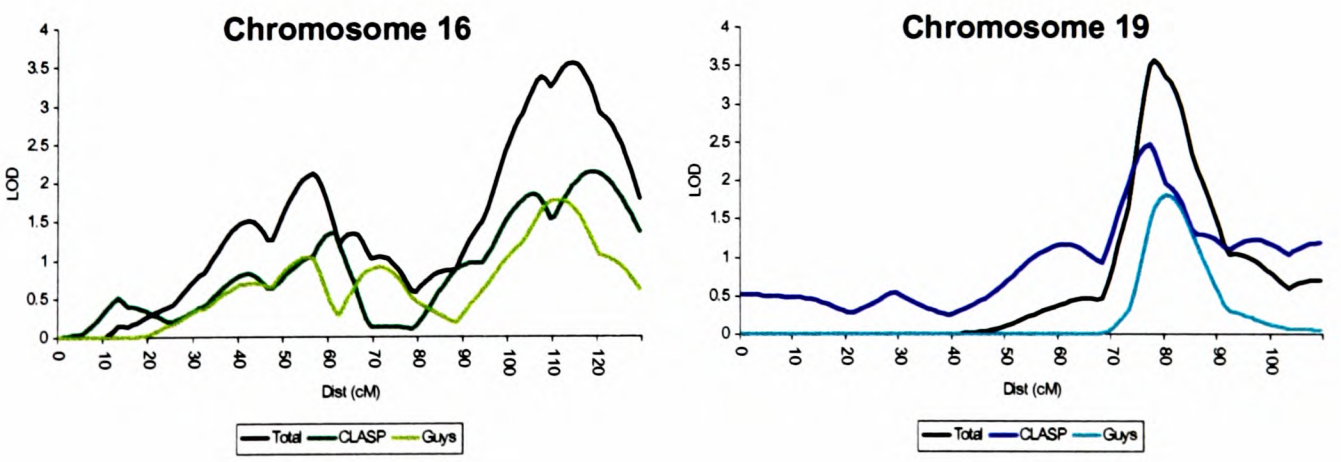


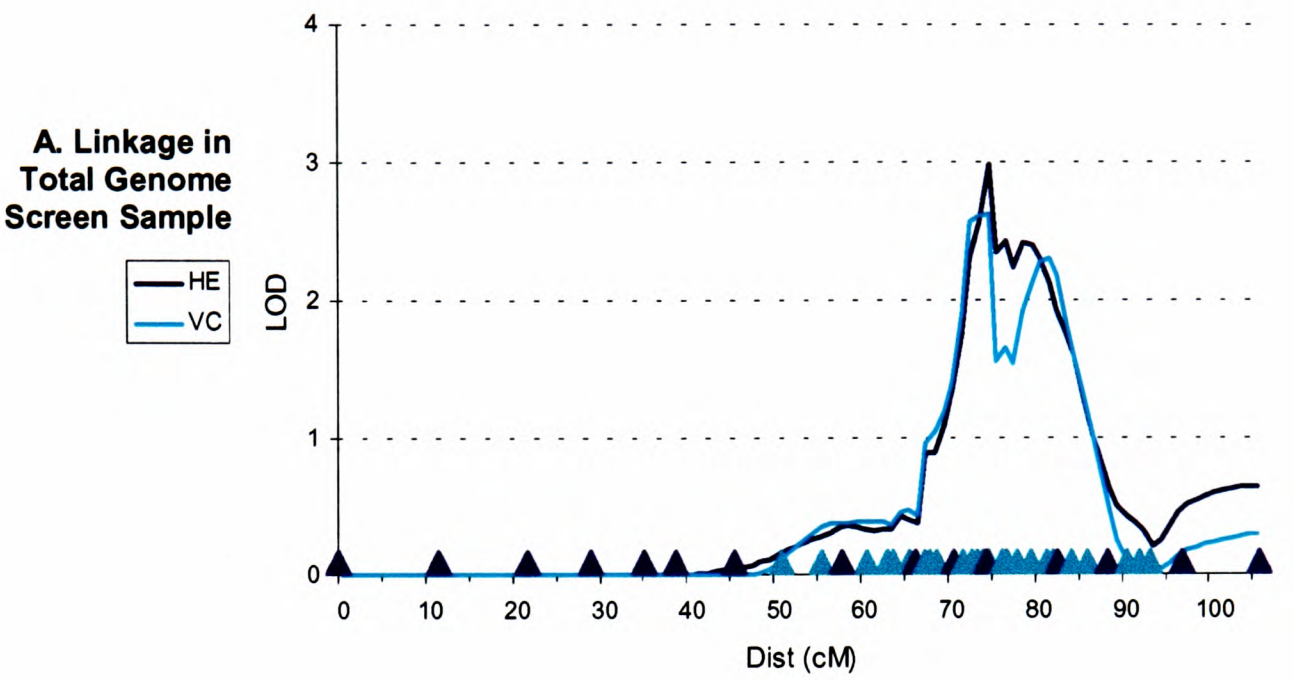
FIGURE 3.7 - Linkage to Chromosomes 16 and 19 in the Guys and CLASP Samples : *Traces are shown for HE analyses only for NWRtrans on chromosome 16 and ELStrans on chromosome 19. The x-axis shows cumulative distance across the chromosome in Haldane centiMorgans*

3.3ii - Fine mapping of Chromosome 19

As discussed in the introduction, the mapping of genes for complex disorders often requires a second round of linkage analysis to allow the reduction of the genomic regions under investigation (see section 3.1i 'Functional and positional cloning'). Thus the chromosome 19q locus was selected to be studied in further detail. 26 markers were chosen from chromosome 19q, and genotyped and analysed in an identical manner to those markers used for the genome screen. Details of the analysis procedure can be found in chapter 2 ('methods'), and the fine mapping panels are outlined in appendix H. Singlepoint and multipoint results for chromosome 19q are shown in table 3.12 and figure 3.8 respectively.

Dist (cM)	Marker	HE	VC
45.5	**D19S226**	0.24	0.03
50.9	D19S566	0.44	0.40
55.6	D19S882	0.22	0.30
57.8	**D19S414**	0.09	0.01
60.6	D19S225	0.02	0.12
63.3	D19S425	0.01	0.00
63.4	D19S208	0.19	0.09
65.5	D19S224	0.13	0.11
66.0	**D19S220**	0.16	0.19
67.0	D19S881	0.36	0.53
67.5	*NUMB-R*	0.60	0.69
68.2	D19S223	0.45	0.80
70.4	**D19S420**	1.01	0.93
72.2	D19S178	2.61	2.39
73.6	*APOC2*	1.89	1.45
73.7	*APOE*	0.06	0.17
74.0	D19S903	1.61	1.15
74.1	**D19S908**	2.49	2.22
74.3	D19S412	1.12	0.46
76.9	**D19S902**	1.74	0.92
77.0	D19S606	0.05	0.07
78.3	D19S596	1.45	0.69
79.6	D19S879	1.30	0.49
81.8	D19S867	1.25	1.38
82.3	D19S246	1.26	1.46
82.4	**D19S904**	0.81	0.09
84.1	*KLK*	0.16	0.01
85.8	D19S553	0.23	0.82
88.4	**D19S571**	0.12	0.20
90.3	D19S888	0.20	0.00
92.1	D19S921	0.82	0.59
93.3	D19S572	0.41	0.06
97.1	**D19S418**	0.43	0.13
105.1	**D19S210**	0.74	0.65

Table 3.12 - Singlepoint LOD Scores for Chromosome 19q Fine Mapping : *Singlepoint HE and VC LOD scores are given for ELStrans only. No results are given for the original markers on chromosome 19p, outside the region of linkage (see table 3.5 for these results).*
Markers shown in bold were used for the original genome screen analysis. All other markers were added for fine mapping analyses. Cumulative distances across the chromosomes are given in Haldane centiMorgans. All LOD scores above 1.5 are marked in red text



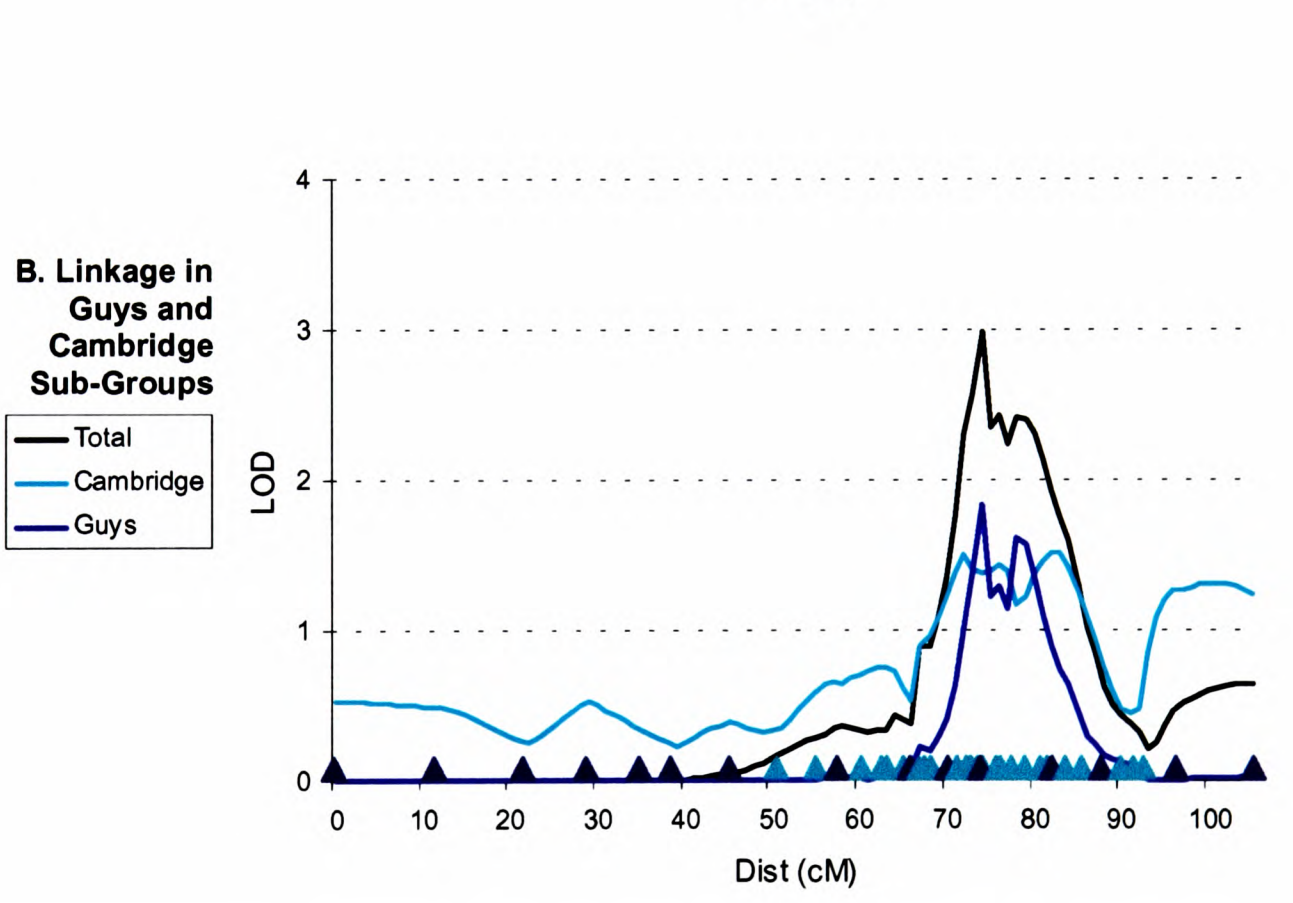


Figure 3.8 - Multipoint LOD Scores for Chromosome 19q Fine Mapping : *Traces are shown for ELStrans only. The x-axis shows cumulative distance across the chromosome in Haldane centiMorgans. The triangles show the positions of markers used for the fine mapping analyses. Dark blue triangles represent original markers, used for genome screen analysis. Light blue triangles represent supplementary markers added for the fine mapping. The chromosome 19 linkage remains after fine mapping under both HE and VC analyses (figure A). Again, linkage is demonstrated within both the clinical and epidemiological sub-groups (figure B).*

3.4 DISCUSSION

This genome screen was the first molecular genetic study of generalised SLI and implicated two loci believed to influence language-related traits. Evidence for these QTLs was drawn from four complementary analyses (multipoint HE and VC and singlepoint HE and VC) and both loci were shown to be relevant within clinical and epidemiological samples.

One important feature of this study is the use of quantitative measures of generalised language abilities. The lack of consensus as to the aetiological basis of SLI often makes the derivation of a consistent qualitative affection status unfeasible. The use of quantitative traits circumvents this issue and, in complex cognitive disorders, has been demonstrated to provide a suitable means of investigating underlying genetic effects (Cardon et al 1994; Fisher et al 1999; Gayán et al 1999). A quantitative trait approach does however, create its own issues, perhaps the most pertinent of which is the selection of phenotypes for the appraisal of disorder severity. In the diagnosis of SLI both ICD-10 and DSM-IV guidelines require a substantial discrepancy between non-verbal IQ and verbal abilities (see section 1.2i 'Non verbal IQ'). However, whilst the enforcement of discrepancy scores acts to aid the elimination of general IQ effects, they are generally felt to result in an over-restrictive phenotype which is susceptible to compound errors. Also relevant to the current study is the finding that discrepancy scores show only a minimal level of heritability and hence may not reflect the underlying genetic influences involved in SLI (Bishop et al 1995). The SLIC genome screen therefore chose to employ two language measures derived from broad phenotype batteries, alongside a single specific measurement of phonological short-term memory. All three of these traits have been demonstrated to be significantly heritable and are good predictors of language abilities (Semel et al 1992, Bishop et al 1995, Bishop et al 1999a).

The importance of phenotype selection in the quantitative study of complex disorders is confirmed by the results of the current genome screen. Intriguingly, across the entire genome, only a minimal level of linkage is seen to measures of receptive language abilities – the strongest RLStrans result was seen on chromosome 2q and peaked at 1.52. These genetic results are entirely consistent with the previously reported lack of probandwise concordance between twins with pure receptive impairments (Bishop et al 1995). In contrast,

measurements of expressive language skills and phonological short-term memory, both of which have previously been demonstrated to be under strong genetic influence (Bishop et al 1995, Bishop et al 1999a), yielded the two most significant linkage results of the genome screen. Furthermore, despite the fact that all three measures were moderately correlated within the genome screen sample (see table 2.8), the level of concordance between phenotypes within each of the regions of linkage was found to be minimal. Linkages on chromosomes 16 and 19 were seen to be specific to NWRtrans and ELStrans respectively, with no corresponding peaks seen for the other measures. Thus, whilst all three measures represent consistent and reliable indicators of language abilities, only ELStrans and NWRtrans were able to demonstrate the underlying genetic components. Furthermore, both ELStrans and NWRtrans appear to reflect distinct genetic factors. It should be noted, however, that studies of dyslexia (Grigorenko et al 1997. Fisher et al 1999) indicate that the dissection of a complex trait in such a simple manner is not always appropriate and that inferences relating specific loci to distinct components of language impairment should be viewed with caution (Fisher et al 1999), especially in small sample sizes.

In genome screens for complex traits it is not uncommon to see a shift between the original peak and replication peaks or to find linkage to alternate phenotypes from that originally reported (Altmüller et al 2001). Thus the independent reproduction of the chromosome 16 and 19 loci within the constituent epidemiological and clinical groups was particularly striking. The observation of linkage in exactly the same region to the same phenotypes across two separate groups with such different origins provides further endorsement for the QTLs reported here.

The chromosome 19 locus is further substantiated by the fine-mapping results that continue to support the presence of a language-related locus in this region. Although the LOD score decreased slightly after fine mapping, it is still suggestive of linkage and the locus is again reflected within the total, epidemiological and clinical samples, under all the types of analysis performed. The linkage area was not reduced in width by the fine mapping panel and the linkage curves were seen to become slightly 'spiky'. This spikiness is not uncommon in fine mapping investigations (IMGSAC 2001b) and is thought to result from inaccuracies in the

order and distances of the high-density marker map. The fine mapping panel did not significantly increase the information content across the 19q region and a denser marker map, perhaps using single-nucleotide polymorphisms (SNPs) may be required to resolve this linkage region significantly.

The use of simulations allowed the derivation of empirical p-values for all peaks obtained. This is a practice that is strongly recommended for whole-genome screen studies (Lander & Kruglyak 1995), and is becoming increasingly common. The use of simulations is ratified by the NWR results on chromosome 16. Although the VC results (LOD=2.57) for non-word repetition appeared somewhat less significant than those obtained for HE (LOD=3.55), simulations demonstrated that the empirical probability of the VC result was entirely consistent with that found for the HE (HE empirical p = 0.00003, VC empirical p = 0.00008). Furthermore both results verged upon the threshold for 'significant linkage' as suggested by Lander and Kruglyak (i.e. p=0.00002) (Lander and Kruglyak 1995).

Our sample contained a male to female ratio of approximately 3:2, which is consistent with the male predominance reported in previous studies (Stevenson & Richman 1976). However we found no strong evidence for a major sex locus in the HE analysis of the X chromosome. Although a LOD score of 1.30 was found on Xp for ELStrans, LODs remained below 0.5 across the entire X chromosome for all other measures (see appendix J).

Interestingly, neither of the loci identified here overlap with those highlighted by the Bartlett et al genome screen (Bartlett et al 2002). Around marker D13S800, our LOD remained below 0.77 and on chromosome 2p the LOD scores never exceeded 0.68 (see appendix J). This apparent lack of consensus is probably attributable to differences in study design. Although both screens were related to language impairments, each took a completely different approach to their subject selection, phenotyping and genetic strategies. Whilst the present study used small nuclear families and quantitative measures of language abilities (expressive language, receptive language and non-word repetition) alongside non-parametric analyses. Bartlett et al used large extended pedigrees, a selection of binary affection statuses (language impairment, reading impairment and clinical impairment) and parametric genetic analyses (under both a recessive and a dominant model of inheritance). Each study design

has its own strengths — the use of quantitative traits circumvents the need for the derivation of a consistent affection status, whilst the use of large pedigrees can act to increase the homogeneity of the sample — and it is possible that both have revealed loci that are of general importance to the SLI phenotype. As with other complex disorders, only independent replication of these studies will enable the evaluation of the importance of each of the identified chromosomal regions.

Neither did we find any evidence for linkage to chromosome 7q, the location of both the *SPCH1* and *AUTS1* loci (see appendix J). At D7S486, the peak of linkage in the KE family (LOD=6.22), our singlepoint LOD remained below 0.001 for all three phenotypes.

However, whilst we found no evidence for linkage to any major locus previously associated with autism (i.e. chromosomes 2,7 and 15 – see section 3.2iii 'Autism'), a recent genome screen for autistic disorder did implicate a region on 19q which overlaps with that reported here (Liu et al 2001). Nevertheless, this study is the only one to implicate 19q in autism and no evidence of replication has yet been reported. Furthermore, Liu et al found linkage to 19q (MLS=1.70) with a narrow diagnostic criteria which excluded any borderline cases of autism, which may be argued to overlap into the SLI spectrum.

Similarly, we found no evidence for the existence of coincidental loci between SLI and ADHD or dyslexia. Neither chromosome 16q nor 19q have been implicated in ADHD or dyslexia and we found no evidence for linkages to chromosomes 2, 6, 15 or 18 (dyslexia – see section 3.2iv 'Dyslexia') or to the sites of the major ADHD candidate gene loci (*DAT1* (chromosome 5) and *DRD4* (chromosome 11) – see section 3.2v 'Attention Deficit Hyperactivity Disorder (ADHD)').

The main limitation of this screen is the sample sizes used. Whilst work is ongoing to collect a second wave of families, larger sample sizes may be required to identify loci which have more minor effects upon the SLI phenotype. Further studies using larger sample sizes, will not only enable the replication of the chromosome 16 and 19 loci but may also allow the identification of loci which may overlap with those loci previously identified in disorders such as autism, ADHD and dyslexia and thereby allow the elucidation of any common genetic mechanisms underlying these disorders.

CHAPTER 4

Candidate Gene Analyses

4.1 INTRODUCTION

Two genes were studied as candidates for involvement in language impairment. The first — numblike (*NBL*), or numb-related (*NUMB-R*) (MIM 604018) — was selected from the region of linkage on chromosome 19q and encodes a protein involved in neural cell fate decisions in the developing neocortex. The second — Forkhead-bOX domain P2 (*FOXP2*) (MIM 605317) — lies on chromosome 7q and is mutated within a family with a severe speech and language disorder.

The sequence of each gene was acquired from the NCBI database and primers were designed to allow the amplification of all known exonic regions. The entire coding sequence of each gene was then screened for mutations in the Guys cohort probands (43 individuals) using DHPLC (see section 2.8 'Denaturing High Performance Liquid Chromatography (DHPLC)'). Any samples that showed a variant elution pattern were directly sequenced to characterise the changes.

In addition to the mutation scan, association analyses were performed within the QTDT package using a selection of SNPs from across each gene. These included polymorphisms within introns and putative upstream CpG regions and therefore allowed the evaluation of the impact of non-coding variants upon the language phenotype.

4.2 ASSOCIATION

4.2i - Association Analysis

Association refers to a situation in which the presence of a given phenotype or genotype increases the probability of a second trait also being present. Such an association may arise due to chance or population stratification (population association), in which case the two traits are not necessarily linked at the genetic level. In contrast, the presence of an 'allelic association' implies that the relationship between the two traits is caused by the physical proximity of the genetic loci that underlie the phenotypes.

Tests for genetic association essentially rely upon the same principles as those used for linkage analyses in as much as they both test for non-random relationships between genotypic and phenotypic information within large samples of individuals affected by the

disorder under study. However, whilst linkage analysis necessitates the use of family-based cohorts, association analyses can just as easily be performed using unrelated, case/control samples. It is assumed that if these individuals come from the same population then the case cohort will share a common ancestral background haplotype which is in linkage disequilibrium with the disease causing allele. Thus whilst linkage analysis relies upon the investigation of a small number of recombinations within a predetermined pedigree structure, association analysis has the ability to reconstruct the meiotic events of several generations using ancestrally related individuals from the present time. Over this period, recombination events and selection will have acted to strengthen any meaningful genotype-phenotype associations whilst allowing the decay of inconsequential relationships. Thus, association analyses afford more power than linkage-based investigations, especially in the study of loci that have only a small effect upon the overall phenotypic outcome (e.g. in complex disorders) (reviewed by Ardlie et al 2002). In addition, this generation gap means that the distance over which linkage disequilibrium (LD) extends is shortened, and thus the resolution resulting from association analyses tends to be greater than that achieved by linkage studies. However, this increase in resolution is accompanied by the caveat that a greater number of markers are needed to map the disease-causing allele. It has been suggested that between 30,000 and 1 million markers (The International SNP Map Working Group 2001) would be required to build a fully comprehensive association map of the entire genome. Thus genome-wide association analyses are not only accompanied by practical limitations but also by multiple testing issues, and most association studies to date have therefore chosen to focus upon specific chromosome regions chosen by a candidate gene approach or highlighted by genome-wide linkage studies.

4.2ii - Approaches to Association Analysis

In the traditional case/control association study design, allele frequencies are compared between unrelated cases and unrelated controls. This approach assumes that any differences in allele frequencies between the two groups will not have arisen for any reason other than those which are directly correlated with the disorder under study, and can therefore lead to an increase in type I errors. One way around this is to use family-based samples in the study of

association. In contrast to case/control studies, these methods compare the frequency with which a given marker allele is transmitted to an affected offspring against the frequency with which it is not transmitted (Transmission Disequilibrium Test (TDT) - Spielman et al 1993, Spielman et al 1994). It is assumed that deviations from the expected 50:50 transmission ratio will arise either because the allele is either causing the disorder (directly or indirectly) or is linked to a locus which does underlie the disease. In addition to overcoming population stratification issues, family-based association methods also allow the use of the same sample collections for both linkage and association analyses.

The original TDT test was developed for the study of discrete traits, using biallelic markers for which the genotype of both parents is known (Spielman et al 1993, Spielman et al 1994). However more recent extensions now mean that these types of analyses can be applied to incomplete genotype data (TRANSMIT – Clayton 1999, S-TDT - Spielman & Ewens 1998, LRAT – Cervino & Hill 2000), multiallelic markers (G-TDT - Schaid 1996, Sham & Curtis 1995) and continuous traits (Allison 1997, Rabinowitz 1997, Fulker et al 1999, Abecasis et al 2000).

The QTDT test allows the use of quantitative trait-associated phenotypes in the measurement of affection status and was developed for the study of association when linkage to the region has previously been demonstrated. Simulation studies indicate that the QTDT test provides comparable power to the TDT test for the detection of association within family samples (Abecasis et al 2001a).

4.2iii - Linkage Disequilibrium Blocks and Association Studies

The exact number of markers required to map a disease-causing allele can be determined by the level of LD, which is influenced at the genetic level by selection, mutation and recombination events and is therefore unique to the region of the genome under study. Studies indicate that, in Europeans at least, the human genome consists of islands of high LD, which are characterised by a relatively low number of haplotype classes, separated by hotspots of recombination where LD extends over only very short regions (Daly et al 2001, Jefferys et al 2001). If the size of the LD blocks is known then in theory it should be possible to treat each block as an 'allele' and therefore represent the haplotype diversity of that region

by typing only a minimal set of polymorphic markers. However, given the unpredictable nature linkage disequilibrium across the genome (Abecasis et al 2001b, Daly et al 2001, Weiss & Clark 2002), and in the absence of a reliable method by which to identify recombination hotspots, such an approach would require the prior characterisation of these LD blocks within each chromosomal region and population to be studied.

In a study of the *APOE* gene region on chromosome 19q (i.e. close to the *NUMB-R* locus), Martin et al (2000a) found that LD extends around 40kb in this area. To date, no detailed LD studies have been performed in the *FOXP2* region of chromosome 7q.

4.2iv - Choice of Markers for Association Studies

A further complication of association studies arises from the fact that the level of association detected can depend heavily upon the exact combinations of markers selected to be studied.

Martin et al (2000b) studied 60 SNPs in a 1.5Mb region around the *APOE* locus (MIM 104300) on chromosome 19q in both Alzheimer's Disease (AD – MIM 107741) patients and unrelated controls. They found that only 16 of these SNPs gave significant evidence of association ($P<0.05$). Furthermore, even though the *APOE* gene has been demonstrated to underlie AD, only 4 of the 7 SNPs that lay close (within 16Kb) to the *APOE* polymorphism were significantly associated with the AD phenotype. From their study, Martin et al (2000b) were unable to deduce any algorithm that would help to identify those SNPs that were more likely to reflect association. Whilst coding and non-coding SNPs were found to be as efficient as each other (unless the coding SNPs were those which were directly influencing disease susceptibility) there was some evidence that SNPs with moderate allele frequencies may provide more power than those with extreme allele frequencies. Conversely, Abecasis et al (2001a) indicate that the greatest power to detect association is obtained when the trait allele and the marker allele are of a similar frequency and are in phase with each other. Thus, even if a trait and a marker both have allele frequencies of 80%:20%, the power to detect association can still vary dramatically. This will depend on whether the marker's rare allele is associated with the trait's common allele, in which case only minimal power will be obtained, or it's rare allele, in which case optimal power will be achieved.

4.3 NUMBLIKE (*NUMB-R*)

The numblike gene is found on human chromosome 19q13.3 (see figure 4.1) and is expressed in neural precursor cells during the development of the neocortex, where it acts as a negative regulator of the Notch protein.

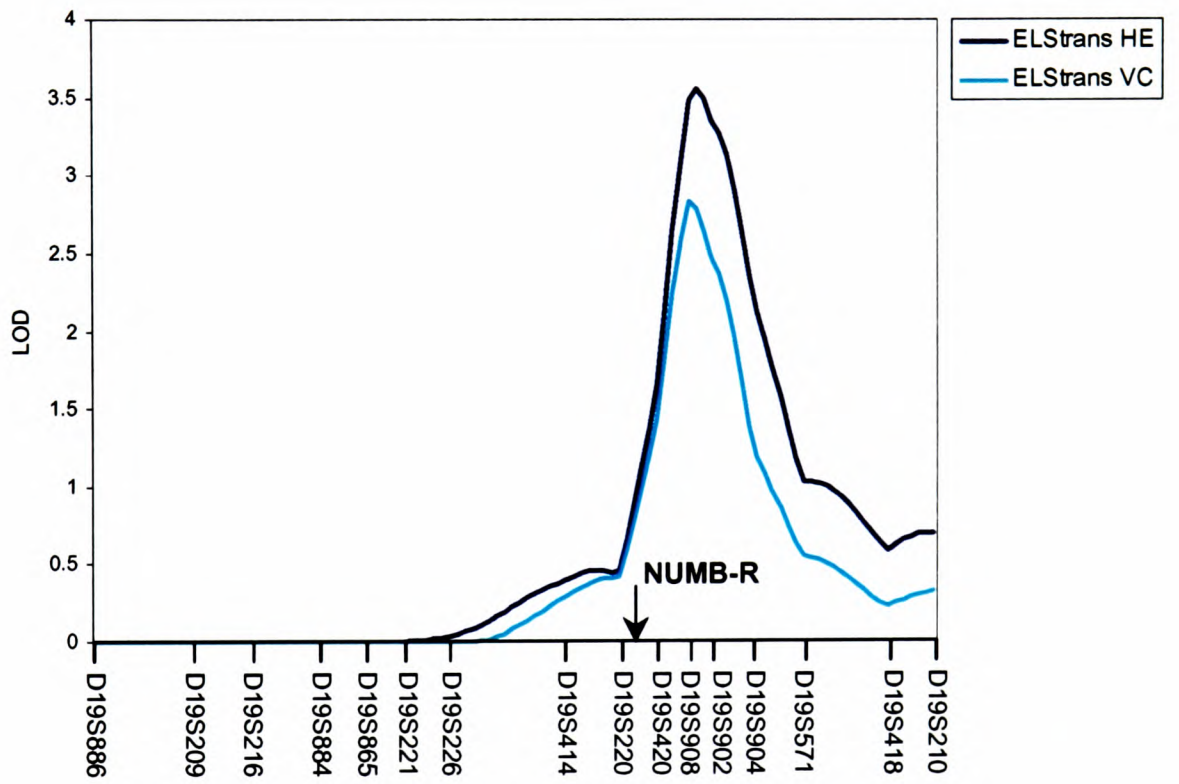


Figure 4.1 - The Position of *NUMB-R* in Relation to the Chromosome 19 Peak : *NUMB-R is positioned at approximately 67.5cM between markers D19S220 (66.0cM) and D19S420 (70.4cM)*

4.3i - Human Neurogenesis

During neurogenesis the cortex is built up as distinct layers (ventricular, subventricular, intermediate, neural plate and marginal zones) within the neural tube. Immature neuronal precursor cells reside in the inner ventricular layer, where they undergo a series of asymmetric divisions. Basal precursor-daughter cells migrate away from the ventricular layer, pass through the intermediate zones and settle in the outer cortical layers where they build up the neural plate in an outside-in pattern. Apical-precursor daughter cells remain within the

ventricular zone and continue to divide asymmetrically providing further daughter cells for the formation of the neural cortex (see figure 4.2).

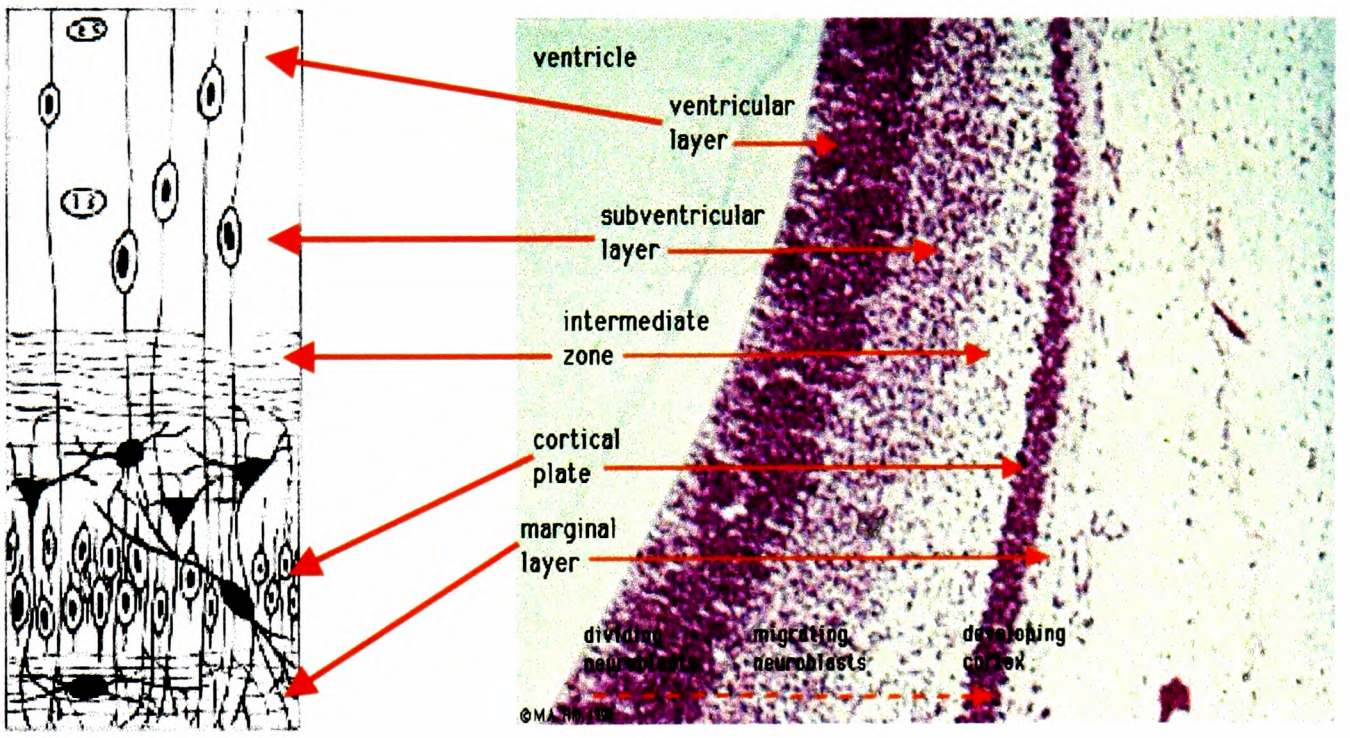


Figure 4.2 - The Development of the Neocortex : *The distinct layers of the developing neocortex. Neuronal precursor cells migrate away from the ventricular layer passing through the intermediate zones and settling at the cortical plate where they differentiate into neurons. Earlier formed neurons form the outer layer of the cortex and later formed neurons are added on the inside*

The process of asymmetric precursor division and neuronal migration is controlled, in part, by the expression of the *NUMB*, *NUMB-R* and *NOTCH1* genes within the neuronal precursor cells. Numb is an asymmetrically expressed membrane-associated protein that directly binds and deactivates the Notch1 transmembrane protein. The interaction between Notch1 and Numb creates opposing gradients of the two proteins within the immature neuronal precursor cell. Thus, upon division, one daughter cell contains a high level of Notch1 and the other contains an excess of Numb. The Numb daughter cell invariably forms the apical cell and remains at the ventricular layer whereas the Notch daughter cell forms the basal descendant, which travels through the cortical layers. During migration, the *NUMB-R* gene is switched on within the basal daughter cell. Rising levels of the Numb-R protein, together with residual amounts of the Numb protein, act to slowly decrease Notch1 levels and, together with cell

contact signals, lead to the termination of cell movement at the cortical plate and allow the differentiation of the basal daughter cell into a neuron (Zhong et al 1997) (see figure 4.3).

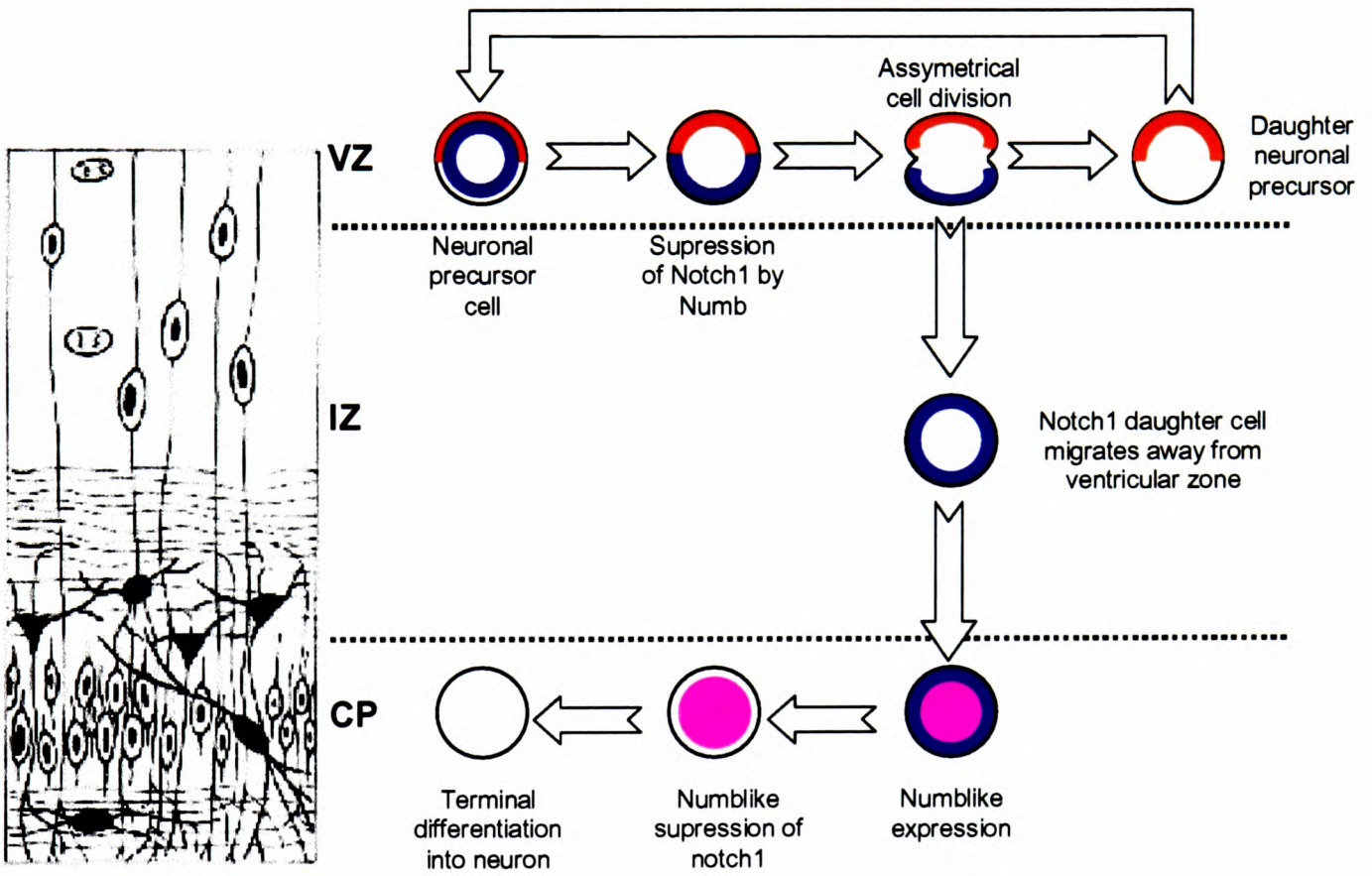


Figure 4.3 - *NUMB*, *NOTCH1* and *NUMB-R* Expression in the Developing Neocortex : *In the developing cortex, neuronal precursor cells divide asymmetrically to produce an apical daughter cell which remains at the ventricular layer and continues to divide mitotically, and a basal daughter cell which migrates through the developing neocortex to form a neuronal cell in the outer cortical plate. The asymmetrical cell division is driven by interactions between Numb, Notch1 and Numb-R. The red shading represents NUMB expression, the blue shading represents NOTCH1 expression and the pink shading represents NUMB-R expression*

VZ = ventricular zone
IZ = intermediate zone
CP = cortical plate

4.3ii - The *NUMB-R* gene

The human *NUMB-R* gene is found on chromosome 19q13.3 and consists of 10 exons encoding a 609 amino acid (65KDa) protein (Salcini et al 1997) (see figure 4.4). As its name suggests, the *NUMB-R* gene sequence is highly similar to that of *NUMB* and the resultant proteins share a 57% homology. Both genes contain two characteristic protein-protein interaction domains; a phosphotyrosine binding (PTB) domain (Zhong et al 1997) and an Eps15 homology (EH) domain (Salcini et al 1997). In addition, exon 10 of the *NUMB-R* gene

encodes a polyglutamine tract that has no counterpart in the *NUMB* gene (Margolis et al 1997). This polyglutamine stretch is encoded by a mixture of CAA and CAG codons and has been reported to be polymorphic in humans with a size range of between 16 and 20 residues (Margolis et al 1997). Expansion of polyglutamine repeats has been well documented in the progression of neurodegenerative disorders such as Huntington's Disease (Huntington's Disease Collaborative Research Group 1993). However, such cases typically involve longer polyglutamine stretches made up of pure CAG triplet repeat sequences (Orr 2001).

The human *NUMB* gene is found on chromosome 14q24 and has been associated with a severe form of early-onset familial Alzheimer's disease (Sherrington et al 1995).

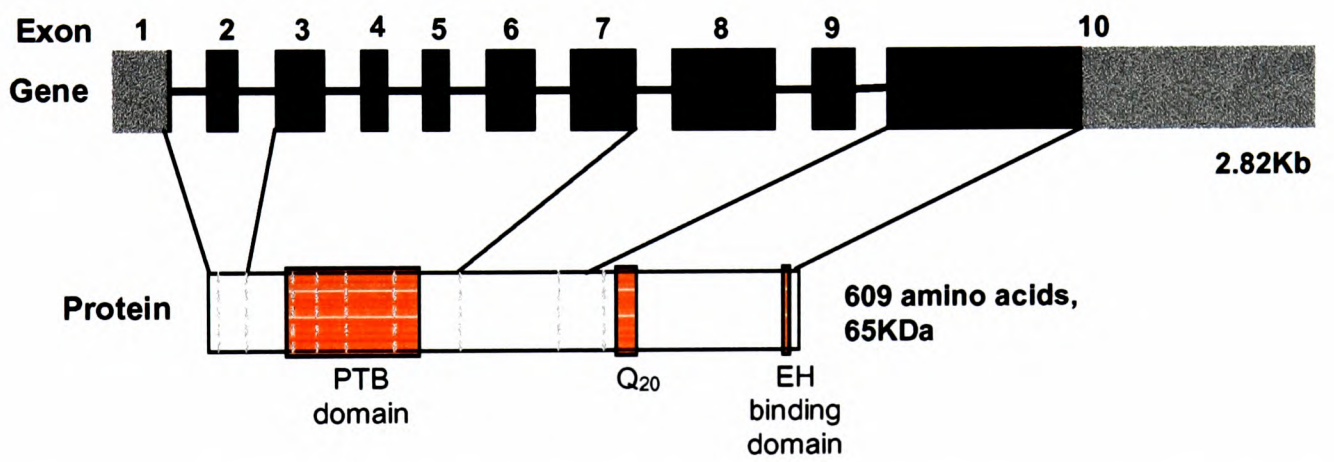


Figure 4.4 - *NUMB-R* Gene Structure : *The structure of the NUMB-R gene and protein. All exons and introns are given to scale. Black regions represent coding regions, grey areas represent known untranslated regions. For the complete gene sequence please see appendices K and L*

4.3iii - *NUMB-R* Gene Expression

Expression studies indicate mouse Numblike (*nbl*) is a cytoplasmic protein that is highly expressed throughout the nervous system from E12.5 onwards. *nbl* mRNA can be detected throughout the neural tube and within the cranial sensory ganglia, the dorsal root ganglia, the sympathetic ganglia and developing sensory organs such as nasal epithelia and the retina (Zhong et al 1997). Closer inspection of the neural tube indicates that the gene is expressed within the postmitotic neural tube zones outside the ventricular layer with the highest levels found at the cortical plate (Zhong et al 1997). In contrast, mouse Notch1 immunoreactivity is

highest in cells in the ventricular and intermediate zones and lowest at the cortical plate whilst m-*numb* is found to be expressed ubiquitously throughout the developing cortex, from the ventricular zone through to the cortical plate (Zhong et al 1997).

4.3iv - *NUMB-R* Associated Knockouts

To date, no *NUMB-R* knockouts have been created within animal model systems. However, both *Notch1* and *Numb* knockouts have been created for both mice and Drosophila. As may be expected, given the roles of Numb and Notch1 in cell-fate decisions, the Drosophila knockouts often display an excess of a given cell type at the cost of it's alternative.

Homozygous *m-numb* knockouts are lethal and result in severe defects in cranial neural tube closure and an under-developed forebrain in the unborn foetus (Zhong et al 2000). Death is thought to occur as a result of the abnormal development of intermediate blood vessels and capillaries, which result in extensive bleeding in the foetus (Zilian et al 2001). Heterozygous mutants are appear to be normal and are indistinguishable from their littermates at a phenotypic level.

Mouse *Notch1* knockouts also die prematurely (between E10 and E12) and display a retarded development rate when compared to their heterozygous and wild type littermates. Embryos have gross defects in the size and number of somites and a defective yolk sac vascular system (Huppert et al 2000, Conlon et al 1995). These mice die before any real neural defects can be studied in detail.

4.3v - *NUMB-R* Analyses

In an initial exploration of the *NUMB-R* gene, the polyglutamine repeat region and checked for any signs of expansion within the entire SLIC cohort (98 families, 473 individuals — see section 2.1 'The genome screen cohort'). The resultant genotype information was used for linkage analyses within GH2.0 and formed part of a fine-mapping marker set across the chromosome 19q linkage region (see section 3.3ii 'Fine mapping of chromosome 19' and appendix H). In addition, the repeat genotypes were tested for association to the ELStrans, RLStrans and NWRtrans phenotypes within QTDT (see sections 2.2 'The genome screen phenotypes' and 2.10 'Single Nucleotide Polymorphisms (SNPs)').

This was followed by a DHPLC mutation screen of the *NUMB-R* coding sequence within 43 probands selected from the Guys cohort. Exons 4 and 5 were amplified as a single fragment whilst exon 10 was amplified in 5 overlapping sections. The sequences of each fragment and details of all primers and amplification reactions can be found in appendices K and N. Finally, 5 SNPs (see figure 4.5), spread across the *NUMB-R* coding and associated 5' and 3' sequences, were chosen from dbSNP, typed via a restriction enzyme assay, and tested for association to the ELStrans, RLStrans and NWRtrans phenotypes (see sections 2.2 'The genome screen phenotypes' and 2.10 'Single Nucleotide Polymorphisms (SNPs)'). The sequences of each SNP region and details of the primers and restriction enzymes used can be found in appendices K, N and Q.

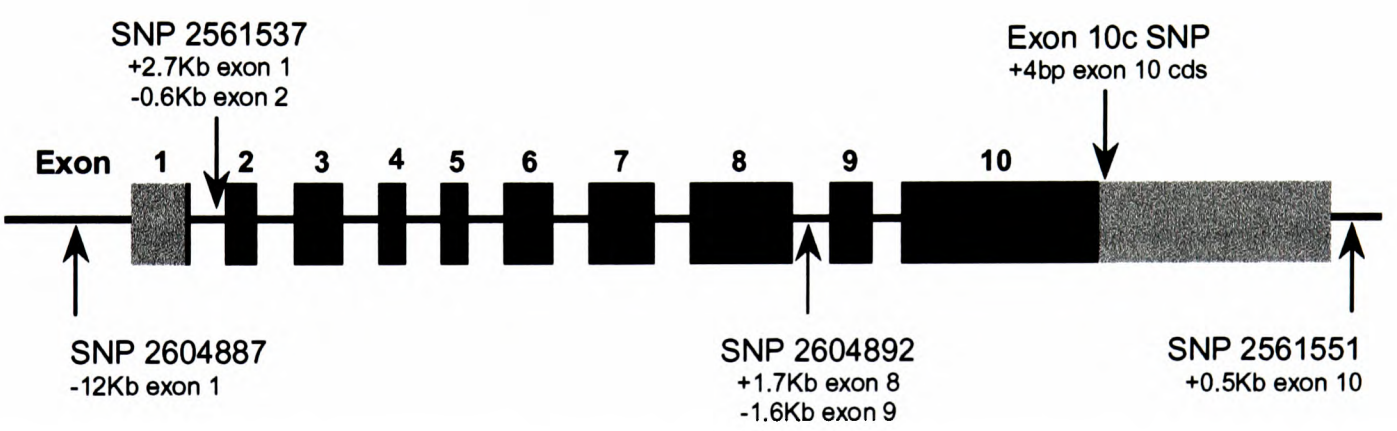


Figure 4.5 – SNPs Analysed for *NUMB-R* Association : *Five SNPs within the NUMB-R gene sequence were analysed for association with three language-related traits. Positions of SNPs are given in relation to flanking exons*

4.4 FORKHEAD-BOX DOMAIN P2 (*FOXP2*)

4.4i - The KE Family

The KE family is a large three-generation pedigree in which approximately half the members (both male and female) are affected by a severe speech and language disorder caused by a single autosomal dominant mutation (see figure 4.6) (Hurst et al 1990).

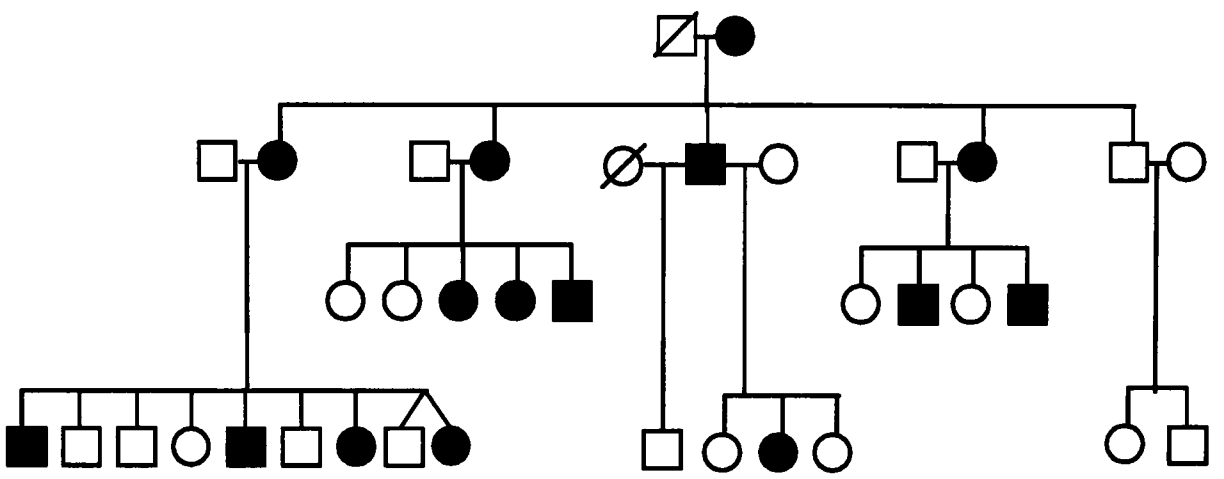


Figure 4.6 - The KE Pedigree : *The KE Pedigree. Circles are females, squares are males. Affected individuals are shaded, unaffected individuals are white. Crossed individuals are deceased*

Affected members of the KE family present with an orofacial dyspraxia that impedes all complex oral movement and leads to a gross deficit in the production of speech sounds. In some individuals this praxis is so severe that their speech is virtually incomprehensible to the naïve listener and a signing system is employed to augment spoken language (Vargha-Khadem et al 1995).

An initial investigation of the language deficits in the KE family suggested that affected members suffer from a grammatical form of language impairment that specifically affects their ability to generate morphosyntactic inflections (Gopnik & Crago 1991, Gopnik 1994). However, subsequent assessments have revealed widespread impairments in both the expressive and receptive language domains and additional limitations in non-verbal oral movement and performance IQ tasks (Vargha-Khadem et al 1995, Watkins et al 2002a). In an investigation which employed a series of 13 tasks, providing a comprehensive screen of language abilities and including tests of language production and comprehension, phonology, grammar, reading and writing, the affected individuals showed significant impairments on all but two tests (object naming and picture vocabulary) (Vargha-Khadem et al 1995). Throughout a series of similar psycholinguistic investigations, three tasks have been established to allow the unambiguous identification of impaired individuals. These are single

word repetition, non-word repetition and complex orofacial movements (Vargha-Khadem et al 1995, Vargha-Khadem et al 1998, Watkins et al 1999, Watkins et al 2002a).

In addition to their language problems, several affected members of the KE family are reported to experience significant ancillary non-verbal deficits. These include problems with non-verbal coding tasks (Watkins et al 2002a), the production of complex (e.g. stick out your tongue) and sequential (e.g. closing the lips, then opening the mouth, then protruding the tongue) oral movements (Vargha-Khadem et al 1995, Alcock et al 2000a) and the discrimination and production of rhythm sequences (Alcock et al 2000b). Note however, that all affected members perform within the normal range on tasks which involve the perception and production of pitch and melody sequences, albeit with long inter-stimulus intervals (Alcock et al 2000b), and those which require the production of simple orofacial movements involving just one muscle group (e.g. make an 'ah' sound) (Alcock et al 2000a).

In addition, studies indicate that the average non-verbal IQ of affected individuals is significantly lower than that of unaffected individuals and several of the affected family members have performance IQ scores extreme enough to place them below the commonly-applied 'normal' threshold of 85 (Vargha-Khadem et al 1995, Watkins et al 2002a). However, it should be noted that these non-verbal deficits do not form an integral part of the phenotype and some affected members have performance IQs within the normal range whilst some unaffected individuals perform below the one standard deviation cut-off on tests of non-verbal IQ (Vargha-Khadem et al 1995, Watkins et al 2002a). In addition, longitudinal investigations of the non-verbal abilities of family members indicate that, like many individuals affected by language impairments, the PIQ scores of some affected individuals is subject to a sharp decline over time. This observation advocates an increasingly common view that disorders of speech and language may limit the development of 'non-verbal intelligence' as measured by current tests (Tallal et al 1991, Watkins et al 2002a).

Thus, although speech and language problems form the central feature of the deficit in the KE family, several aspects of their phenotype would preclude a traditional diagnosis of SLI, and the exact nature of their disorder remains a matter of debate. Whilst many researchers claim that the linguistic difficulties seen in the KE family may have arisen as a direct result of either

their restricted orofacial movement or as part of a more global learning delay, others argue that these aspects of the phenotype are corollaries of the more palpable language deficits. In conclusion therefore, the KE family has a disorder that is characterised by an orofacial dyspraxia and a severe speech and language impairment. Although affected individuals may show some non-verbal deficits, as in more general cases of SLI, the relationship between the onset of disorder and these additional deficits remains unclear.

4.4ii - Brain imaging in the KE Family

Although neuroanatomical studies have failed to identify any gross pathology which can consistently be linked to the speech and language phenotype in the KE family, two recent brain imaging investigations have revealed more subtle brain anomalies which may be postulated to underlie some aspects of the phenotype.

Prior to these studies it was hypothesised that the underlying neuropathology would involve one or more components of the motor system (because of the oral musculature problems) and would be bilateral in nature (because of the persistence of the speech and language difficulties in the KE family).

Studies by Vargha-Khadem et al (1998) indicated that the lentiform nucleus (putamen and globus pallidus) and the angular gyrus both had increased amounts of grey matter bilaterally, whereas the caudate nucleus showed a bilateral decrease in the amount of grey matter in affected individuals when compared to that of unaffected family members. Interestingly, of these three motor-related structures, the left caudate nucleus, which extends into Brocas area, also showed altered levels of activity during a word/non-word repetition task in affected individuals when compared to that of normal control subjects (Vargha-Khadem et al 1998).

Watkins et al (2002b) also found a bilateral reduction in the average amount of grey matter in the caudate nucleus in affected family members. In addition they demonstrated that, of the six affected individuals studied, three had a volume asymmetry that was reversed from that normally expected (i.e. left was significantly larger than right). Furthermore, the absolute volume of the caudate nucleus was significantly correlated with the performance of affected individuals on tests of oral praxis, non-word repetition and the coding subtest of the Wechsler Intelligence Scale (WISC-III/WAIS-R). Note however, that although the average size of the

caudate nucleus was reduced in affected members, some still had caudate nucleus volumes that were equal to or greater than those found in unaffected individuals.

The caudate nucleus, along with the putamen and globus pallidus, forms the basal ganglia, a mass of grey matter deep in the cerebrum. These structures are important in the initiation and co-ordination of movement and may play a particular role in the control of smooth and slow movements. With regards to language, the basal ganglia probably control the movement of the oral muscles, so that the appropriate speed, range and direction of movement can be maintained for clear and coherent speech. Lesions in the basal ganglia tend to result in akinesia (poverty of movement) or dyskinesia (uncontrolled movement) and are linked to disorders such as Parkinson's disease and Huntington's chorea. More specifically, damage to the caudate nucleus has been reported to lead to uncoordinated movements of the limbs and/or facial muscles.

Lesions in the basal ganglia are often linked to aphasic phenotypes (Robin & Schienberg 1990, Fabbro et al 1996, Warren et al 2000), whilst neuronal loss in the subcortical grey matter is a common age-related phenomenon, even in healthy individuals (McDonald et al 1991, Murphy et al 1992, Raz et al 1995, Jernigan et al 2001). Furthermore decreased caudate nucleus volumes have been reported in patients with depression (Krishan et al 1992) and language/learning impairment (Jernigan et al 1991a, Tallal et al 1994), whilst increases in subcortical matter are often seen in children with schizophrenia (Jeste et al 1998, Jernigan et al 1991b) and Downs syndrome (Pinter et al 2001).

4.4iii - *SPCH1* and Chromosome 7q

In 1998 a genome screen was completed involving 27 members of the KE family (Fisher et al 1998). This scan revealed a region of linkage on the long arm of chromosome 7 with marker D7S486 giving a Maximum LOD Score (MLS) of 6.22. Haplotype analyses revealed critical recombinants that placed the *SPCH1* locus (MIM605317) between markers D7S257 and D7S530. The addition of further markers within this region defined a 5.6cM critical region between markers D7S2459 (proximal) and D7S643 (distal) (Fisher et al 1998) (see figure 4.7).

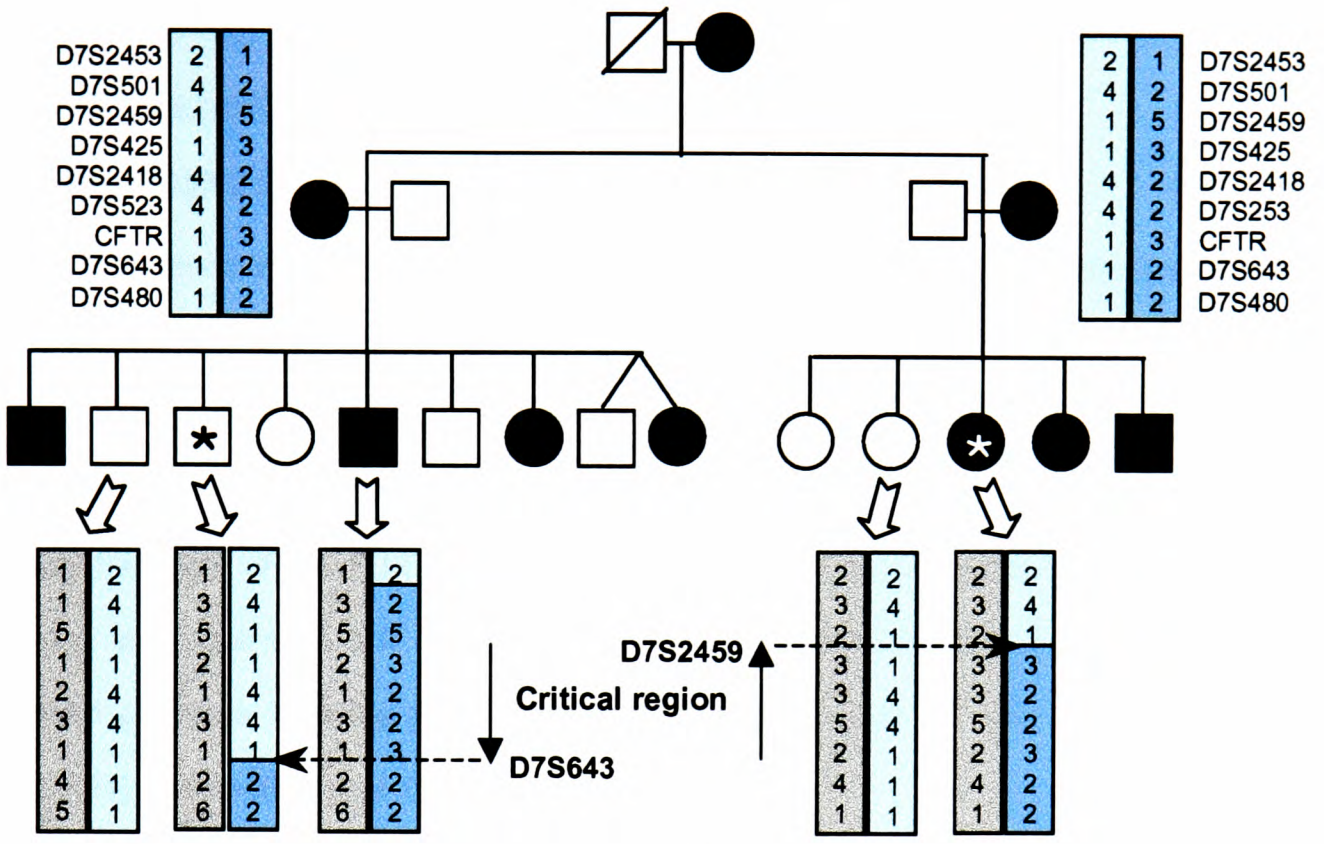


Figure 4.7 - Critical Region of *SPCH1* : *Haplotypes of a section of the KE family within the SPCH1 area of linkage on 7q31. Paternal haplotypes are on the left and maternal haplotypes are on the right. Individuals marked by a star define the critical region between markers D7S2459 and D7S643*

The publication of the *SPCH1* findings resulted in the identification of a second patient, known as CS, with a *de novo* translocation involving a breakpoint within the *SPCH1* region (t(5;7)(q22;q31.2)) (Lai et al 2000). CS was described as having an oral dyspraxia with language impairments in both the receptive and expressive domains. Linguistically, he was considered to be very similar to the affected members of the KE family. Note, however, that his non-verbal skills were reported to be within the normal range. Mapping of the CS translocation revealed the chromosome 7q breakpoint to lie within a partial brain-expressed transcript known as *CAGH44* (Lai et al 2000). Subsequent computer-based investigations allowed the compilation of the full coding sequence of *CAGH44* and revealed the presence of a forkhead box or winged-helix (fox) domain within the gene. Accordingly the gene was renamed *FOXP2* (Lai et al 2001). Mutation screening of the complete *FOXP2* coding sequence within the KE family identified a point mutation that co-segregated perfectly with the speech and language phenotype (Lai et al 2001). The KE mutation involves a single base

transition (G→A), which results in an arginine to histidine substitution within exon 14 of the *FOXP2* gene (R553H) (Lai et al 2001).

4.4iv - The *FOXP2* Gene

The *FOX* genes are a large family of transcription factors all of which possess a winged-helix, or forkhead box (fox), DNA binding domain. The family consists of over 100 members, which are phylogenetically clustered into 15 sub-classes identified by a letter (e.g. *FOXP2*). Members of subclasses are distinguished by a number (e.g. *FOXP2*) (Kaestner et al 2000). Many members of the *FOX* gene family are involved in important cellular regulation events during embryogenesis, and mutations in *FOX* genes have been implicated in a range of human disorders including glaucoma (*FOXC1*) (Nishimura et al 1998) and hypothyroidism (*FOXE1*) (Clifton-Bligh et al 1998).

The forkhead domain typically consists of a stretch of 100 conserved amino acids that adopt a characteristic structure containing three α-helices and two large loops, or wings. In this structure, the third α-helix makes contact with the major groove of the target DNA (Kaestner et al 2000). The KE mutation involves a change within this third helix, at a residue that lies adjacent to the base that makes direct contact with the DNA (Lai et al 2001). This residue has been demonstrated to be invariable in a sample of 182 independent Caucasian controls, and is highly conserved, not only in humans, but also across species such as mice, fruit flies and yeast.

The sequence of *FOXP2* is organised into 19 exons, two of which are alternatively spliced to create three known alternative transcripts (see figure 4.8). The major splice form encodes a 715 amino acid protein that contains a characteristic carboxy-fox domain (exons 12-14), and two adjacent polyglutamine tracts, one of 40 residues and a second of 10, separated by a stretch of eight amino acids (see figure 4.8). Both polyglutamine repeat tracts are encoded by a mixture of CAG and CAA codons and are therefore postulated to be stable in normal individuals. Note that, although a study has since reported the existence of additional exons and splice variants (Bruce & Margolis 2002 — see discussion), this investigation is limited to the initial *FOXP2* sequence as described by Lai et al (2001).

The mouse *Foxp2* gene has three different transcripts and is expressed in adult lung, spleen, small intestine, skeletal muscle, brain and kidney. In the embryonic mouse, *Foxp2* expression is found mainly in the lungs but is also seen to a lesser extent in the neural, gastrointestinal and cardiovascular tissues. Expression in the nervous system is limited to the neopallial cortex, which matures into the cerebral cortex, and the developing cerebral hemispheres, where the gene product acts as a negative regulator of transcription. In humans, expression has been demonstrated across a broad range of adult-derived tissues. Lai et al (2001) postulate that the KE mutation may result in an insufficient dosage of *FOXP2* during embryogenesis that, in turn may lead to the under-development of brain areas critical for speech and language development.

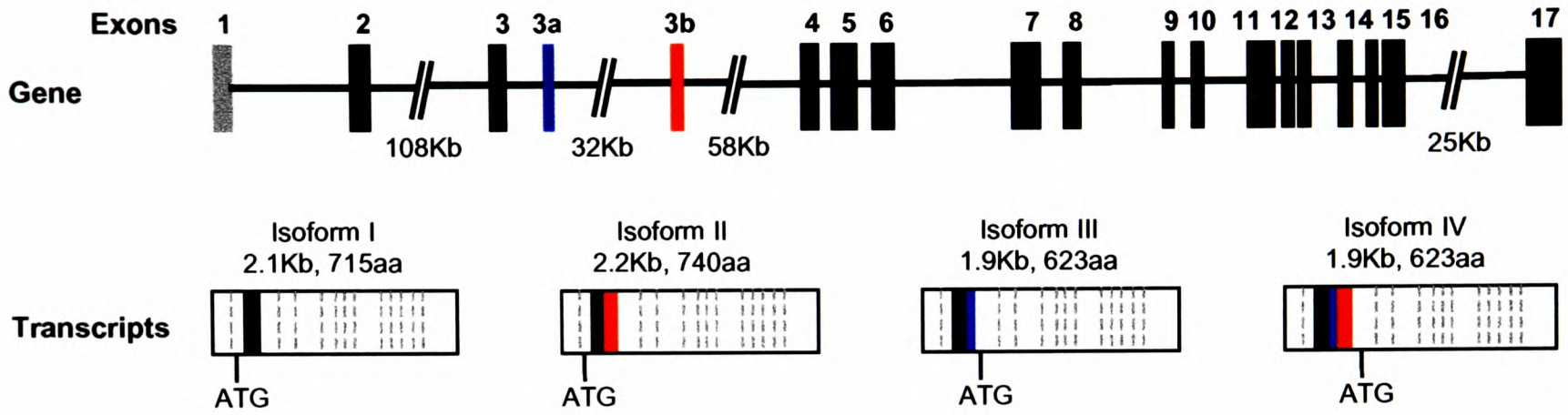


Figure 4.8 - The *FOXP2* Coding Sequence : *The FOXP2 gene is found on chromosome 7q and consists of 19 exons, 2 of which are alternatively spliced. The forkhead domain is encoded by exons 12 to 14 and the polyglutamine tract spans exons 5 and 6. For the complete gene sequence please see appendices N and O. All exons are given to scale. Larger introns are shown with slashes and the sizes given above. Grey regions represent untranslated regions. Alternative splicing of exons 3, 3a and 3b results in 4 known transcripts which give rise to 3 different isoforms of the foxp2 protein. The inclusion of exon 3a results in a frame shift that moves the transcript start codon to exon 4. Isoforms III and IV therefore encode the same protein*

4.4v - *FOXP2* in SLI

Although the orofacial dyspraxia of the KE family would exclude them from a traditional diagnosis of SLI, and despite the fact that the SLIC genome scan found no linkage to chromosome 7q, there is growing evidence for the involvement of this region in speech and language disorders. Cytogenetic studies have described individuals with chromosome 7q abnormalities and associate language impairments (Ashley-Koch et al 1999, Warburton et al 2000), and some meeting abstracts have reported positive associations between chromosome 7q and language phenotypes (Tomblin et al 1998, Schick et al 2000). Thus *FOXP2* remains a prime candidate for involvement in more generalised language impairments.

We therefore performed a mutation screen of the *FOXP2* coding sequence within 43 probands selected from the Guys cohort. Primers for this DHPLC study were provided by Lai et al. The sequences of each fragment and details of all primers and amplification reactions can be found in appendices N and P.

In addition, 2 SNPs and four microsatellite markers were chosen from across the *FOXP2* sequence (see figure 4.9) and tested for association to the ELStrans, RLStrans and NWRtrans phenotypes within the QTDT package (see section 2.10iii 'Quantitative Transmission Disequilibrium Test (QTDT)'). Microsatellites were identified using the RepeatMasker software (BCM Search Launcher Website) and were all intronic to the *FOXP2* sequence (JA Lamb). One SNP was identified through the DHPLC analysis and the second, which lay in a putative promoter region, was chosen from dbSNP.

As previously mentioned (see section 4.4iv 'The *FOXP2* gene'), the transcription of *FOXP2* is believed to be complex and the gene is likely to encode several alternative transcripts (Bruce & Margolis 2002). It is therefore likely that additional *FOXP2* 5' and/or 3' sequences exist which remain uncharacterised. In order to localise the 5' end of the gene, we therefore initiated a search for CpG islands in the genomic sequence immediately upstream of *FOXP2*.

In the human genome, cytosine bases are often found to be methylated (5meC), especially when they form part of CG dinucleotides or CpGs. This methylation has been linked to gene regulation. In tissue specific genes, the level of cytosine methylation is seen to fall prior to

transcription. Thus demethylation is seen as a commitment to gene expression. In the inactive X chromosome all the CpG sites are highly methylated. The human genome can be divided into two fractions with regards to CpGs — the major fraction consisting of methylated CpGs every 50 to 100bp (~98% of the genome) and the minor fraction where the CpGs are undermethylated and occur approximately every 10bp (~2% of the genome). Minor CpG islands are often found to co-occur with the 5' end of house keeping genes and may cover the promoter and one or more exons. In humans, all known housekeeping genes and ~40% of tissue specific genes colocalise with CpGs. It is therefore likely that the promoter regions of the *FOXP2* gene is associated with upstream CpG sequences.

Putative CpG islands were identified using a range of complementary bioinformatic analyses (Nix, Promoter inspector, PromoterScan and Ensembl), all of which indicated that the closest CpG island lay 340Kb upstream of the present *FOXP2* coding sequence. Similarity searches demonstrated that this sequence was highly homologous (83% identity) to a CpG region upstream of the mouse *Foxp2* gene (CS Lai, SE Fisher). Furthermore this murine sequence was directly linked to the *Foxp2* coding sequence by three independent ESTs (AW490098, BB660527 and BB656124), indicating that it is transcribed as part of the mouse *Foxp2* mRNA. SNP 923875 (see figure 4.9) was chosen from the SNP database to lie close to this CpG sequence, and therefore allow an investigation of any associations which may exist between the *FOXP2* promoter sequences and the language impairment phenotype within the Guys cohort. The sequences of each SNP and microsatellite region and details of the primers and restriction enzymes used can be found in appendices N, P and Q.

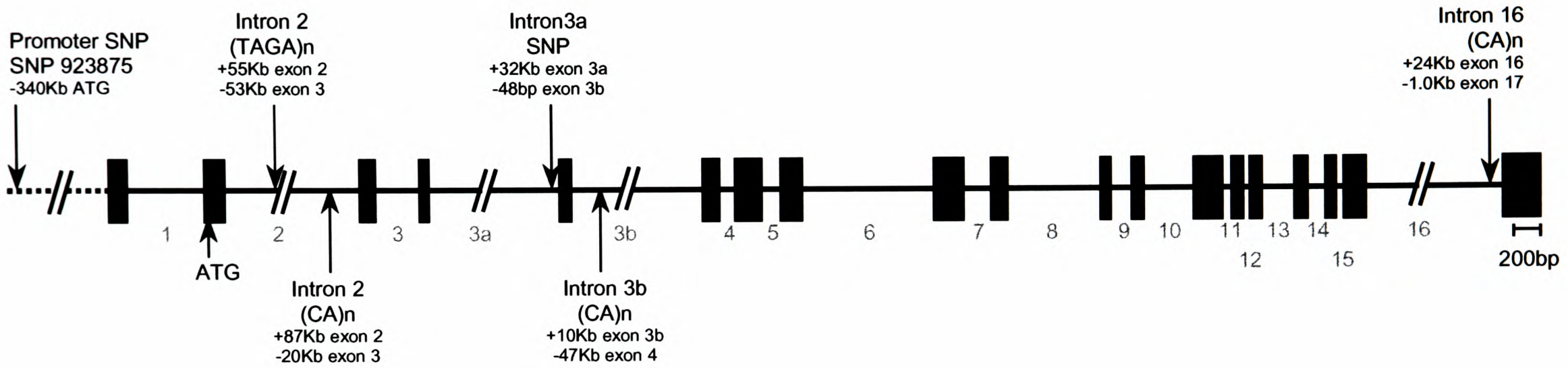


Figure 4.9 - SNPs and Microsatellites Analysed for *FOXP2* Association : *Positions of markers are given in relation to flanking exons*

4.5 CANDIDATE GENE ANALYSIS RESULTS

4.5i - *NUMB-R* Polyglutamine Typing

Genotyping of the polyglutamine tract within the SLIC genome screen sample indicated that this region shows only a low level of polymorphism and failed to reveal any signs of polyglutamine expansion (see table 4.1).

No. of Glutamine Repeats	Allele Frequency
14	0.004
15	0.001
18	0.639
20	0.356

Table 4.1 - Size of Polyglutamine Repeat Tracts Within the Genome Screen Sample

No evidence of linkage or association was found between the polyglutamine tract and the three language-related traits analysed. The association results are shown in table 4.4 and the linkage results can be found in section 3.3ii 'Fine mapping of chromosome 19'.

4.5ii – DHPLC of *NUMB-R*

DHPLC analyses of the *NUMB-R* fragments revealed three sequence variants, two of which lay in the 3' untranslated region of exon 10 (see table 4.2 and figure 4.13). The position of these sequence variants can also be found in appendix K.

Position of Change	Base Change	Amino Acid Change	Frequency
Exon 4			
+27bp exon 4	G→C	Glu→Asp	1/43
Exon 10			
+670bp exon 10	C→T	NC 3'UTR	12/43
+1354bp exon 10	G→A	NC 3'UTR	5/43

Table 4.2 - *NUMB-R* Sequence Variants Detected by DHPLC : *NC = No Change in amino acid sequence*

The heterozygous change in exon 4 was found in a single proband (proband 27) from the Guys sample, and involved a Guanine to Cytosine nucleotide transition which is predicted to result in a glutamic acid to aspartic acid substitution within the PTB binding domain of *NUMB-R* (A92D).

Although a change in a single family would not be sufficient to account for the high level of linkage found on chromosome 19, it remained possible that other, as yet uncharacterised, intronic or promoter variants may contribute to this region of linkage. Exon 4 was therefore sequenced for all individuals of family 27 to allow the tracking of the change through the pedigree. This sequence analysis revealed that the Guanine to Cytosine transition was passed from the mother, who has no history of language problems, to two of her children (child 2 and child 3), both of whom do present with language impairments. Child 2 was diagnosed with 'SLI and associated perceptual and motor problems' and child 4 has 'SLI and triple X syndrome'. The father reports no history of language impairment and child 1 had a language delay when younger but has experienced no further language problems.' Neither the father nor child 1 carry the variant (see figure 4.10).

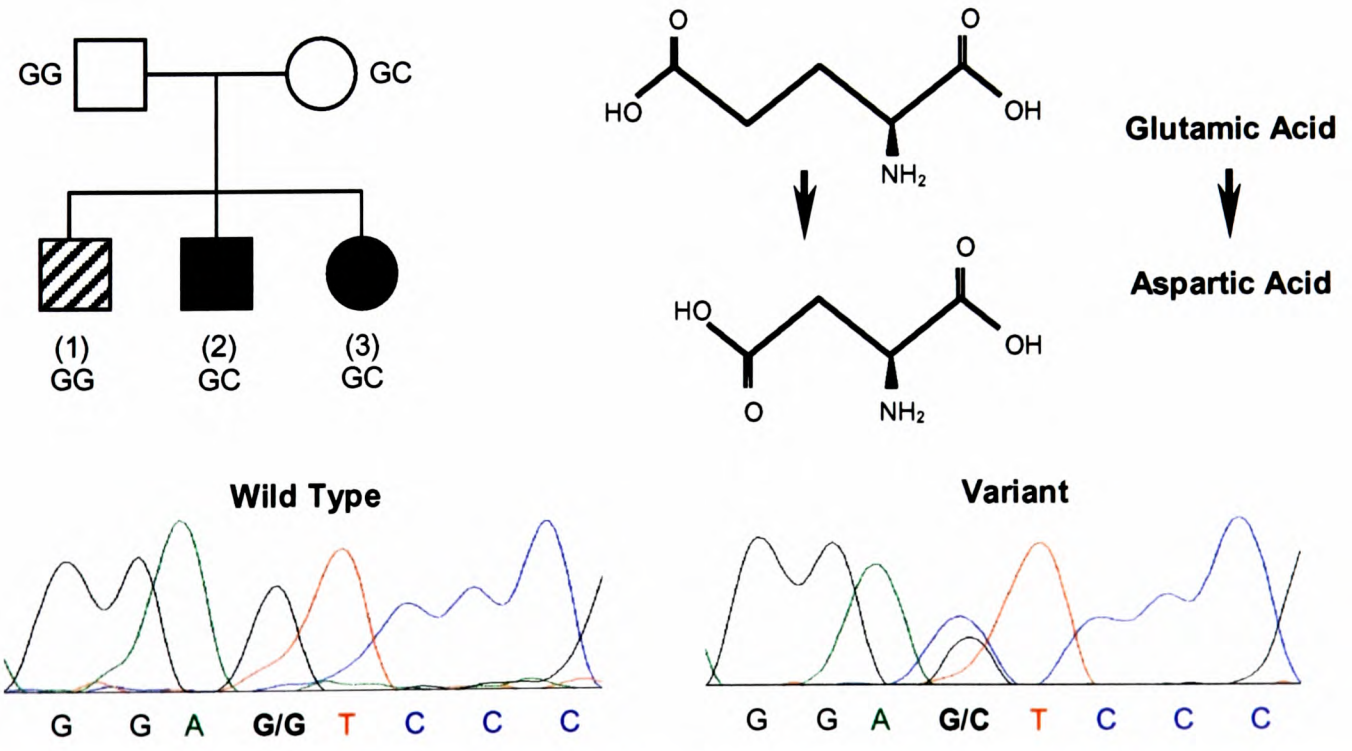


Figure 4.10 - The Exon 4 Polymorphism in Family 27 : *The mother of family 27 passes a heterozygous change in exon 4 of the NUMB-R gene onto the proband (child 2) and his sibling (child 3). This guanine to cytosine nucleotide transition results in a glutamic acid to aspartic acid substitution in the NUMB-R amino acid sequence. Shaded individuals are affected by SLI. Child 1 had a language delay when younger but has experienced no further language problems*

The exon 4 mutation results in a single amino acid change within the *NUMB-R* protein from glutamic acid to aspartic acid. Both of these bases are acidic in nature and vary by only a single methyl group (see figure 4.10). In addition, although this change occurs in the PTB binding domain, it does not involve a base that is highly conserved within this motif (Zhong et al 1997).

In conclusion it is likely that the change in exon 4 represents a rare polymorphism which is not linked to the language phenotype in family 27.

4.5iii – DHPLC of *FOXP2*

DHPLC analyses revealed a total of nine sequence variants (see table 4.3 and figure 4.13), the majority of which were single-base substitutions within intronic regions.

Position of Change	Base Change	Amino Acid Change	Frequency
Intron 3			
-102bp from exon 3a	T→A	NC Intron	11/43
Intron 3a			
-32bp from exon 3b	A→G	NC Intron	1/43
-48bp from exon 3b	T→C	NC Intron	19/43
Intron 5			
+17bp from exon 5	T→G	NC Intron	10/43
Exon 6			
Intron 5/exon 6 border	Ins CAGCAG	+QQ (?)	1/43
Intron 11			
+9bp from exon 11	T→C	NC Intron	1/43
-80bp from exon 12	A→G	NC Intron	1/43
Intron 13			
+30bp from exon 13	C→G	NC Intron	2/43
Intron 14			
+24bp from exon 14	T→C	NC Intron	1/43

Table 4.3 - FOXP2 Sequence Variants Detected by DHPLC : *NC = No Change in amino acid sequence. The CAGCAG insertion was in a stretch of CAGCAG on the intron/exon border and so may or may not result in an amino acid change (see text)*

One change was detected which may result in a coding change (exon 6). This variant was found in a single proband (family 43) and involved a 6 base pair insertion (+CAGCAG) within the shorter of the two *FOXP2* polyglutamine tracts, at the intron/exon border of exon 6. This region represents the longest stretch of pure CAG repeats within *FOXP2* gene and thus is most likely position for an expansion to occur. All available members of this family were therefore sequenced for this change (see figure 4.11). Sequence analysis revealed that the mother, who has no history of language difficulties, passes the insertion variant onto two of her children (child 1 and child 3), both of whom have language impairments. The remaining child (child 2) was also diagnosed with SLI, but yet does not carry the variant allele. The father, who has no history of language impairment, also carries two 'wild-type' alleles.

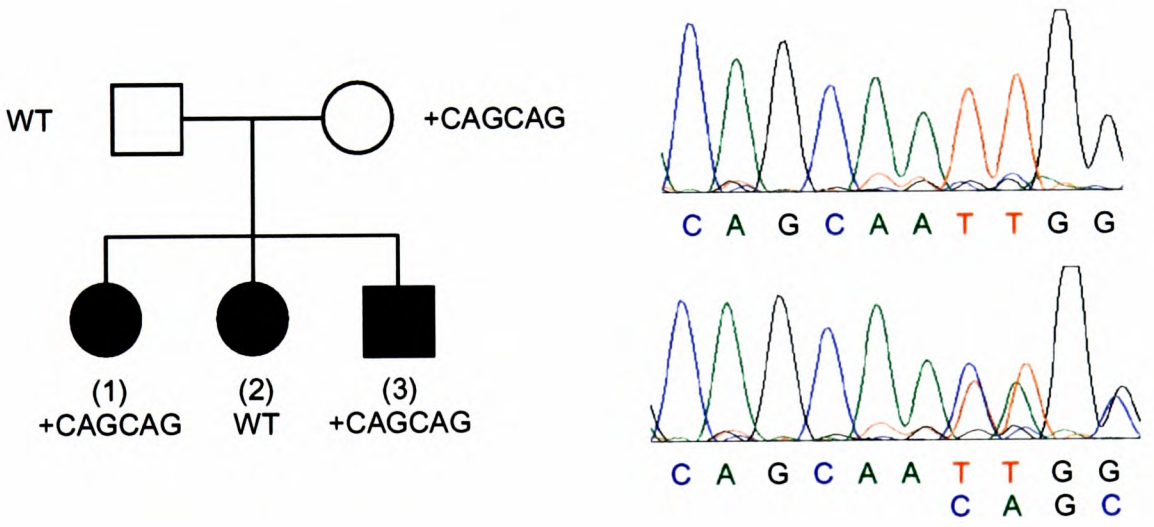


Figure 4.11 - The Exon 6 Polymorphism in Family 43 : *The mother of family 43 passes a heterozygous change in exon 6 of the FOXP2 gene onto the proband (child 1) and her sibling (child 3). This CAGCAG nucleotide insertion may result in an addition of two glutamine residues to the shorter polyglutamine tract of the FOXP2 gene*

Although the exon 6 variant may result in an addition of two glutamine residues to the shorter of the two polyglutamine stretches in *FOXP2*, the repetitive nature of the region means that the exact position of the insertion cannot be determined. Thus the insertion may actually occur within the intronic sequence preceding exon 6 (see figure 4.12).

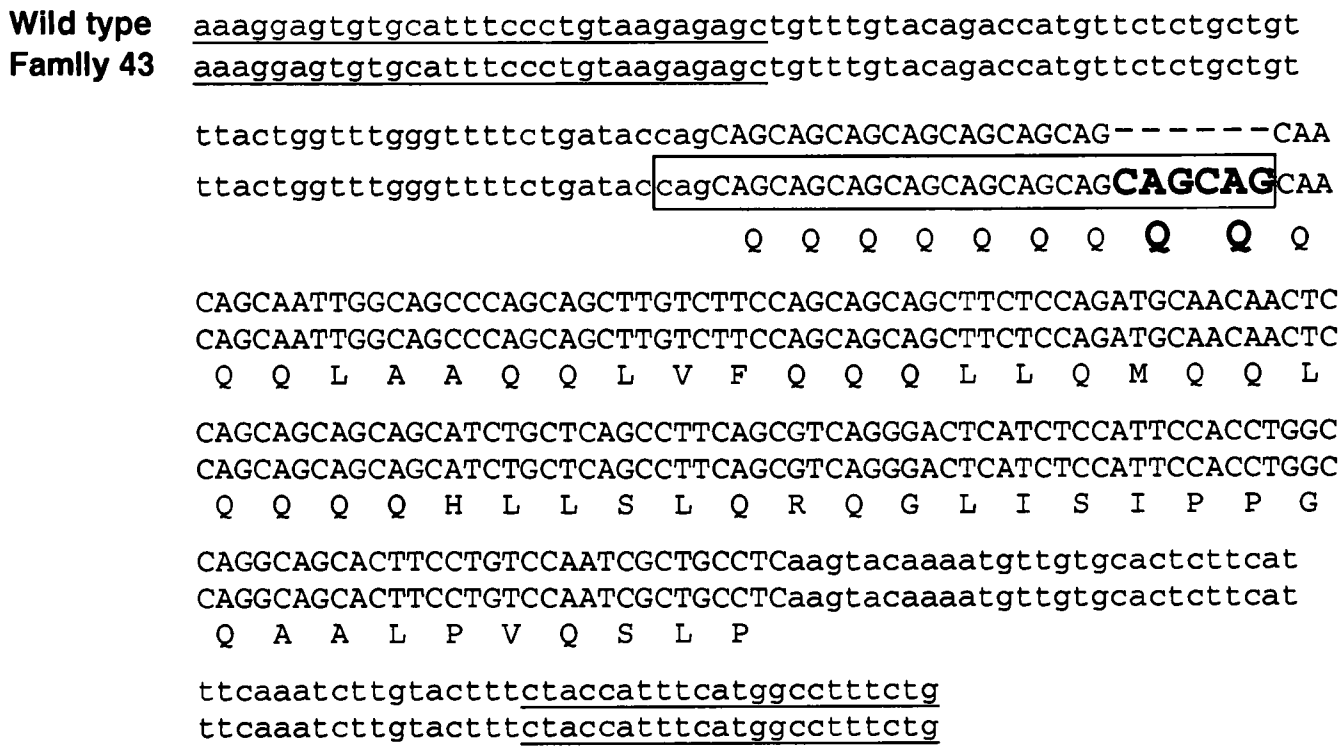


Figure 4.12 - Possible Positions of the CAGCAG Insertion in *FOXP2* Exon 6 : *Because of the repetitive nature of the sequence around the 5' end of exon 6, the exact position of the insertion in family 43 is not clear. The insertion may fall anywhere within the boxed area and therefore may, or may not, result in an addition of two glutamine residues to the FOXP2 polyglutamine tract*

In conclusion, the exon 6 polymorphism found within family 43 does not co-segregate with the language phenotype, may actually fall within intron 5 and would not alter the reading frame of the *FOXP2* sequence even if it was a coding variant. It is therefore likely that this expansion represents a rare polymorphism that is not relevant to the SLI phenotype.

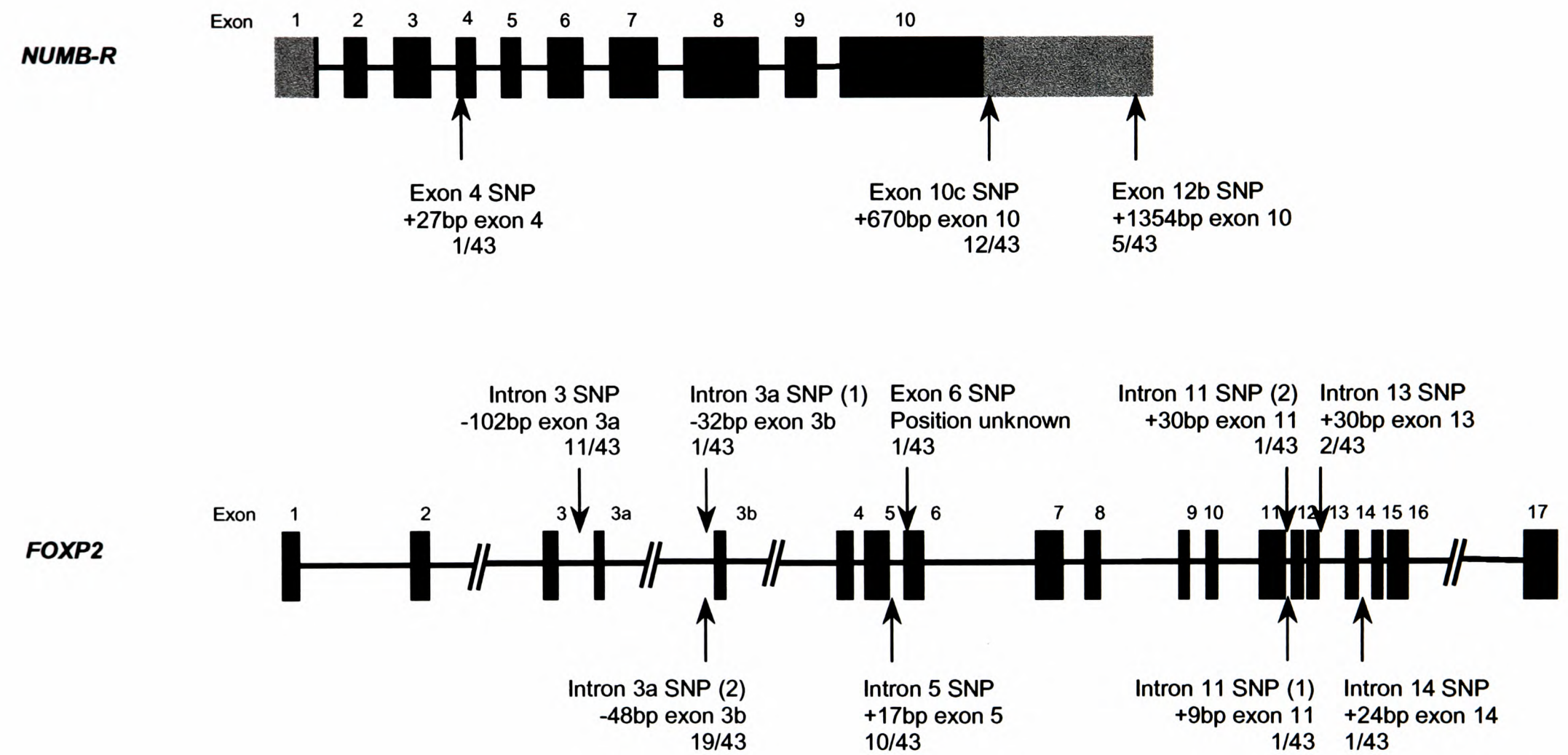


Figure 4.13 - SNPs Detected in *NUMB-R* and *FOXP2* : *Exons are shown as squares with non-coding sequences in grey. Positions are given in relation to nearest exon and the frequency of each SNP in the proband sample is shown*

4.5iv - Association Analyses of *NUMB-R*

No positive associations (P<0.01) were found between any of the five SNPs typed from the *NUMB-R* sequence (see figure 4.5) and the three quantitative measures of language abilities tested (ELStrans, RLStrans and NWRtrans — see section 2.2 'The genome screen phenotypes') (see table 4.4).

		SNP					
		2604887	**2561537**	**2604892**	**Exon 10c SNP**	**2561551**	**PolyQ**
No. Individuals Typed		174	181	166	183	180	347
Heterozygosity (%)		46.6%	52.5%	44.0%	24.6%	42.2%	44.1%
ELStrans	χ^2	0.15	2.42	0.30	0.05	0.00	1.33
	p	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
RLStrans	χ^2	3.15	0.72	0.69	0.00	0.04	0.01
	p	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
NWRtrans	χ^2	0.47	4.48	0.07	0.53	0.98	0.07
	p	>0.1	0.06	>0.1	>0.1	>0.1	>0.1

Table 4.4 - Association Analyses of *NUMB-R* SNPs : *P-values given are empirical p-values derived from 1000 simulations of all markers typed For more information on the polyglutamine tract please see section 4.5i*

4.5v - Association Analyses of *FOXP2*

No positive associations (P<0.1) were found between any of the microsatellites or SNPs typed from the *FOXP2* sequence (see figure 4.9) and the three quantitative measures of language abilities tested (ELStrans, RLStrans and NWRtrans — see section 2.2 'The genome screen phenotypes') (see table 4.5).

		Microsatellites				SNPs	
		Intron 2 (TAGA)n	Intron 2 (CA)n	Intron 3b (CA)n	Intron 16 (CA)n	Intron 3a SNP	923875
No. Individuals Typed		—	—	—	—	210	202
Heterozygosity (%)		—	—	—	—	51.0%	40.6%
ELS trans	χ^2	2.38	7.14	8.23	2.58	0.07	0.28
	p	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
RLS trans	χ^2	0.95	9.60	10.18	2.58	0.81	2.03
	p	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
NWR trans	χ^2	7.88	8.89	7.69	3.11	0.65	0.03
	p	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1

Table 4.5 - Association Analyses of *FOXP2* SNPs and Microsatellites : *P-values given are empirical p-values derived from 1000 simulations of all markers typed*

4.6 DISCUSSION

4.6i - *NUMB-R*

NUMB-R represents a good candidate gene for involvement in language impairments. It is within a region previously linked to SLI (see section 3.3 'Genome screen results'), contains a polyglutamine tract and is expressed within the developing brain. The present study therefore performed a detailed characterisation of *NUMB-R* within 43 individuals affected by SLI. These investigations included genotyping of the polyglutamine stretch, DHPLC screening and variant sequencing of the coding sequence and association analyses of localised SNPs.

In general, polyglutamine tracts that have a susceptibility to expansion, show a high level of polymorphism within the general population with a typical heterozygosity of greater than 70% (Rubinsztein et al 1996, Zoghbi & Orr 2000). The *NUMB-R* polyglutamine region was found to be only slightly polymorphic in length, with an allele size range consistent with that previously reported (14-20 glutamines) (Margolis et al 1997) and showed no evidence of expansion within the SLIC genome screen sample (98 probands, 473 individuals). Thus it is unlikely that the expansion of this area represents a common cause of SLI.

DHPLC screening and variant sequencing revealed only a single change within the *NUMB-R* coding sequence. Although non-conservative, this change maintained the acidic nature of the base and occurred at a position that is only weakly conserved across species (see figure

4.14). It is therefore unlikely that this coding change plays a major role in the onset of SLI within the family studied.

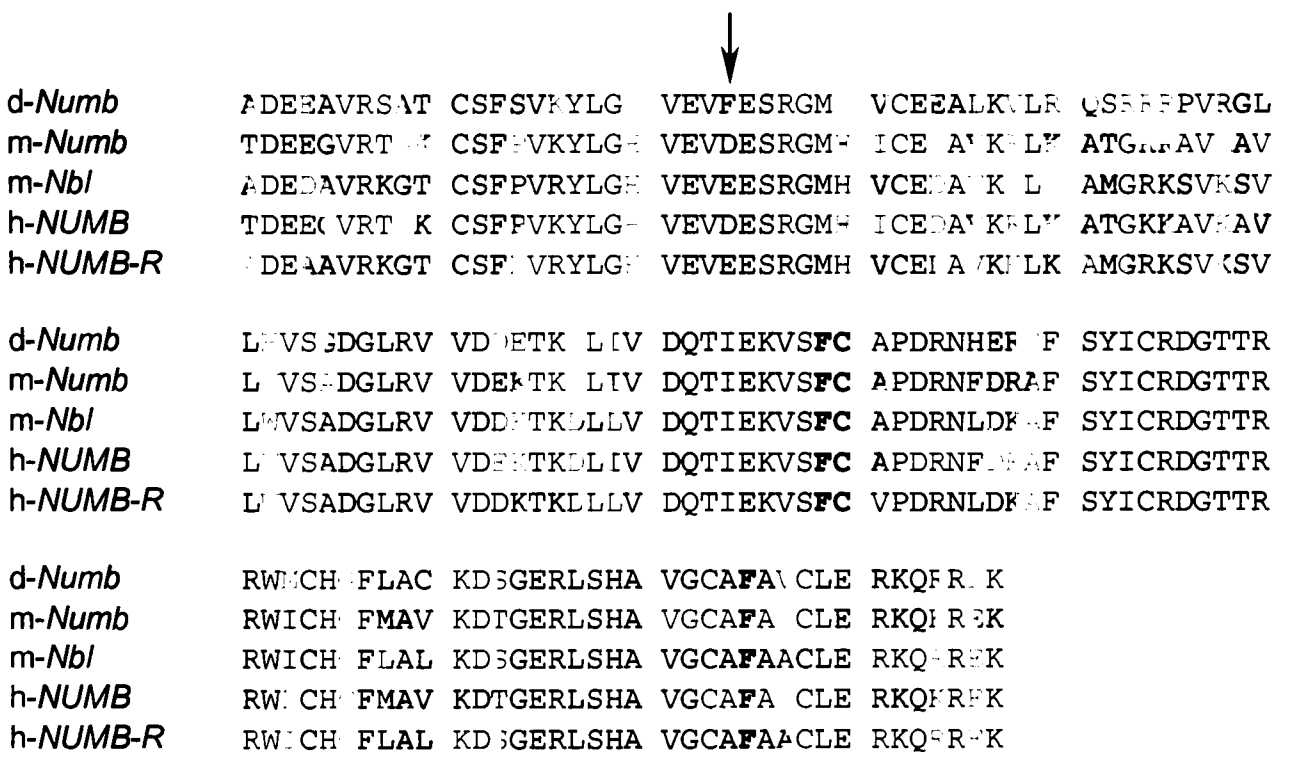


Figure 4.14 - Sequence alignment of the *NUMB* and *NUMB-R* PTB regions across human, mouse and Drosophila : *Alignments of sequences across the PTB region of the NUMB family of genes in Drosophila (d-Numb), mice (m-Numb and m-Nbl) and human (h-NUMB and h-NUMB-R). Black bases represent those conserved across all sequences, red bases are conserved in at least three of the five sequences and blue bases are conserved in two or less sequences. The level of conservation across these sequences is estimated at approximately 65% (Zhong et al 1997). Bases in bold form the recognition sequence of the PTB domain. The arrow represents the position of the NUMB-R amino acid change found in family 27 (see section 4.5ii 'DHPLC of NUMB-R')*

Finally, association analyses of a selection of intronic, 3' and 5' SNPs failed to reveal any signs of association between the *NUMB-R* gene and the three language-related phenotypes analysed.

In conclusion, it is unlikely that the *NUMB-R* gene plays a major role in the onset of SLI.

4.6ii - *FOXP2*

FOXP2 is the first gene to be implicated in a speech and language disorder and, as such, represents the primary candidate gene for SLI. This study therefore undertook a mutation and association screen of this gene within 43 individuals with generalised language impairments.

DHPLC analyses identified one putative coding variant, which involved a +CAGCAG insertion within the smaller of the two *FOXP2* polyglutamine tracts. However, the repetitive nature of the sequence within this tract meant that the exact position of the insertion could not be

determined and it may not actually form part of the transcribed unit. Furthermore the variant sequence was not seen to co-segregate with the language phenotype within the family studied. Following this investigation, two further studies have also reported a low level of polymorphism in the length of the polyglutamine tracts of *FOXP2*. Wassink et al (2002) performed a mutation screen of the *FOXP2* coding sequence within 135 autistic individuals and observed two cases (1.5%) in which the size of the polyglutamine region was altered. Unlike the current study, these changes were found in the larger of the two polyglutamine tracts and involved a reduction in the number of glutamines present (one heterozygous deletion of five glutamine residues and another of six). Neither of these variants was reported to be associated with the autistic phenotype. Bruce & Margolis (2002) studied the *FOXP2* polyglutamine region within 142 individuals with progressive movement disorders. Whilst they found no evidence for polyglutamine expansion within these individuals, they did report that the area showed a low level of polymorphism and estimated the heterozygosity of the polyglutamine tract to be 1%.

Although only one coding variant was identified through DHPLC, the present study identified several intronic variants. The impact of these variants upon the SLI phenotype was therefore determined by association analyses of a selection of SNPs and microsatellites spread across the *FOXP2* region. These analyses were all non-significant.

In conclusion, the role of *FOXP2* in the KE family does not generalise to more common and genetically complex forms of language impairment within the SLIC cohort, and it is therefore unlikely that the *FOXP2* gene plays a major role in the onset of SLI. However, it is worth noting that whilst the SLIC probands tested here were chosen to represent a diverse range of impairments spread over many linguistic domains, they did not show any evidence of verbal or orofacial dyspraxia. Thus, although it is unlikely that *FOXP2* represents a major gene locus in the onset of SLI, it remains possible that it may play a role in a small number of specific cases of language impairment that are characterised by dyspraxic features. Furthermore, it is feasible that another gene within the *FOXP2* pathway may be involved in the SLI phenotype.

In support of this conclusion, a second study has also failed to find any links between the *FOXP2* gene and language impairment in a sample of 270 4-year olds with low general

language scores (Meaburn et al 2002). Note, however, that this group limited their investigation to a search for the specific mutation described in the KE family. Thus, since the individuals selected for the study presented with global language delays and lacked any severe dyspraxic phenotype, the absence of the R553H mutation may not be surprising. If *FOXP2* does play a role in more general language impairments, the majority of cases would be predicted to be genetically complex and therefore involve common polymorphisms that subtly affect the working of the *FOXP2* gene. Without a mutation scan of the entire coding sequence, it is therefore impossible to fully assess the impact of the *FOXP2* gene within this sample.

Also worthy of note is the fact that the *FOXP2* gene has also been studied as a candidate gene for autism. As discussed in section 3.2iii ('Autism'), many genome scans for autistic disorder have provided evidence for the existence of an autism susceptibility gene on the long arm of chromosome 7 (*AUTS1*). Since the location of the *AUTS1* locus overlaps with that of *SPCH1*, many researchers proposed that a single gene located on chromosome 7q31, may be involved in both language impairments and autism (Folstein & Manoski 2000). This theory is strengthened by the overlaps in the phenotypes of autism and SLI (Folstein & Manoski 2000) and by reports of chromosome 7 translocations involving autistic disorder and/or language disorder (Ashley-Koch et al 1999, Warburton et al 2000, Vincent et al 2000). Furthermore, studies indicate that the linkage between chromosome 7q and autism can be strengthened by the stratification of samples according to the presence of severe language difficulties in the proband (Bradford et al 2001), or by the study of quantitative, language-related endophenotypes in autistic patients (Alarcón et al 2002). The identification of the *FOXP2* gene therefore represented an exciting discovery, not only for researchers of language impairment, but also for investigators in the field of autism. To date, two studies have looked for *FOXP2* mutations within independent samples of autistic families (Newbury et al 2002, Wassink et al 2002). Neither of these investigations has found any support for the involvement of *FOXP2* within autistic disorder. However, given the existence of non-autistic, language impairment cases, involving abnormalities of chromosome 7q, it remains possible

that the gene underlying the *AUTS1* locus may also be involved in language impairments or language aspects of the autism phenotype.

As previously mentioned (see section 4.4iv 'The *FOXP2* gene'), since the completion of these studies, the sequence surrounding the *FOXP2* gene has undergone further characterisation and it is now believed that the gene sequence extends beyond that originally described by Lai et al (2001). In an investigation of human cDNAs and Expressed Sequence Tags (ESTs) containing the *FOXP2* sequence, Bruce & Margolis (2002) identified six previously uncharacterised exons, three in the 5' UTR, 2 between exons 2 and 3, one between exons 4 and 5 and a longer version of exon 10 (see figure 4.15).

In this additional sequence, the 5' exons (exons s1, s2 and s3) are untranslated and therefore do not alter the current reading frame which starts in exon 2. These exons are thought to splice directly into exon 2 as the researchers were unable to identify any EST that contained both exon 1 and the new s exons. Exons 2a and 2b, when read in frame with the current sequence, contain stop codons and are therefore also thought to be untranslated. It is possible that these exons form part of the 5'UTR sequence of the alternative transcript (containing exons 3a and 3b – see figure 4.8 'The *FOXP2* coding sequence'), which uses the start codon in exon 4. Exon 4a does contain an open reading frame and is believed to transcribed routinely. Exon 10+ is an extended form of the previously identified exon 10. This alternative exon is 30 bases longer and contains a stop codon that is associated with a downstream polyadenylation signal. The transcription of exon 10+ therefore results in the production of a truncated form of *FOXP2* (known as *FOXP2-S*), which lacks a FOX domain. Over-expression of this transcript in mammalian cells leads to cytoplasmic aggregation and cellular toxicity. It is therefore unlikely that the truncated version of *FOXP2* acts as a transcription factor. In addition to these extra exons, Bruce & Margolis identified a promoter region that lies adjacent to the s1 exon. This promoter region is CpG rich and coincides with the position of the promoter predicted by our analyses. They also used RT-PCR techniques to demonstrate that both *FOXP2* and *FOXP2-S* transcripts are expressed within adult and foetal brain (including the caudate nucleus) and in the lung.

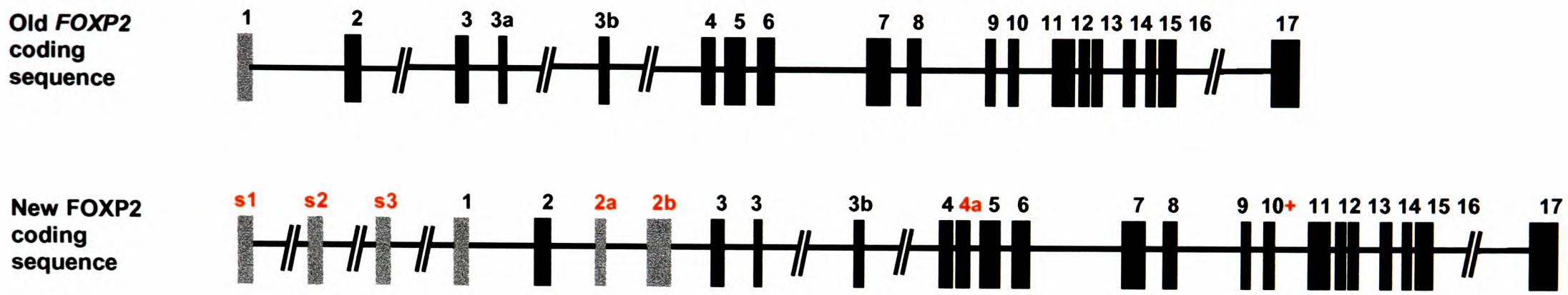


Figure 4.15 – The Revised Genomic Structure of FOXP2 : *Six additional exons have been identified within the FOXP2 sequence. The novel exons are represented by red figures. Coding exons are shown as black boxes and non-coding exons as grey boxes. Introns larger than 50Kbp are shown with hashes*

In conclusion, it appears likely that the *FOXP2* gene encodes a complex range of transcripts that have a variety of cellular functions. It would therefore be prudent to complete extra screening of the *FOXP2* gene before it is excluded as a candidate gene for either SLI or autism.

CHAPTER 5

FISH Mapping of a Translocation in a Child with Language Impairment

5.1 INTRODUCTION

During the collection of the genome screen cohort, metaphase cell spreads were prepared for all clinical (Guys Hospital) samples and cytogenetically tested for chromosome rearrangements. In total, six cases were excluded from the genome screen because of chromosome abnormalities. One child had a rearrangement of chromosome 10, another had a translocation between chromosomes 2 and 11, and the remainder (4 cases) had an abnormal number of sex chromosome (XXX or XXY). Of these cases, the child with the 2;11 translocation was of particular interest. His karyotype was reported to be t(2;11)(q35;p15) and involved a region which had shown linkage to the CELF Receptive Language Score (RLStrans) in the genome screen analysis.

Fluorescence-In-Situ-Hybridisation (FISH) was therefore used to position the chromosome 2q breakpoint within this proband and all available family members (hereafter referred to as the G family). Metaphase cell spreads were prepared for each family member and probed with fluorescently labelled BACs selected from chromosome 2q35 to allow the positioning of the breakpoint. Once the BAC spanning the breakpoint had been identified, SNPs were chosen from across the region, typed via a restriction enzyme assay and tested for association to the ELStrans, RLStrans and NWRtrans phenotypes within the Guys Hospital cohort.

5.2 CHROMOSOME REARRANGEMENTS AND TRANSLOCATIONS

5.2i – Chromosome Rearrangements

There are many different types of chromosome rearrangements. These include duplications, which result in the formation of an additional copy of a given chromosomal segment, deletions, which result in the loss of one copy of a given chromosome segment, and translocations, which result in the transfer of DNA between chromosomes. Rearrangements which do not alter the total amount of DNA present (e.g. translocations and inversions) are known as balanced rearrangements whilst those which do change the complement of the genome (e.g. deletions and duplications) are referred to as imbalanced rearrangements.

In general, chromosome rearrangements usually arise from one of two situations — the erroneous repair of double stranded breaks or the exchange of DNA between non-homologous chromosomes (see figure 5.1).

Within cells, the ends of chromosomes are protected from degradation by complex arrangements known as telomeres. However, occasionally a double-stranded break may occur within the chromosome structure itself resulting in free DNA ends that are susceptible to digestion and are lethal to the cell if not repaired. These double-stranded breaks can be induced by mutagens such as ionising radiation or may just arise by chance. Since double-stranded breaks involve both strands of DNA, they disrupt the sequence that may otherwise be used as a template for the direction of DNA replication. Thus the repair of these types of damage tends to be rather rudimentary and, in essence, simply involves the fusion of the nearest available free chromosome ends. This process tends to be highly inefficient and often results in the loss of several base pairs from around the breakpoints. Furthermore, in cases where more than one double-stranded break has occurred, broken ends may be joined from distinct chromosomes leading to a translocation of DNA from one chromosome to another.

Another method by which chromosome rearrangements may arise is through the pairing of non-homologous chromosomes during meiosis. As discussed in the genome screen chapter (see section 3.ii 'Recombination'), homologous parental chromosomes aligned along the equatorial plane may exchange DNA in a process that drives variability within their offspring (recombination). However, occasionally during meiosis, non-homologous chromosomes may pair up with each other via repetitive elements (e.g. pseudogenes or microsatellites) and recombine resulting in the non-homologous exchange of genetic material. This is known as a reciprocal translocation and results in the generation of two rearranged chromosomes.

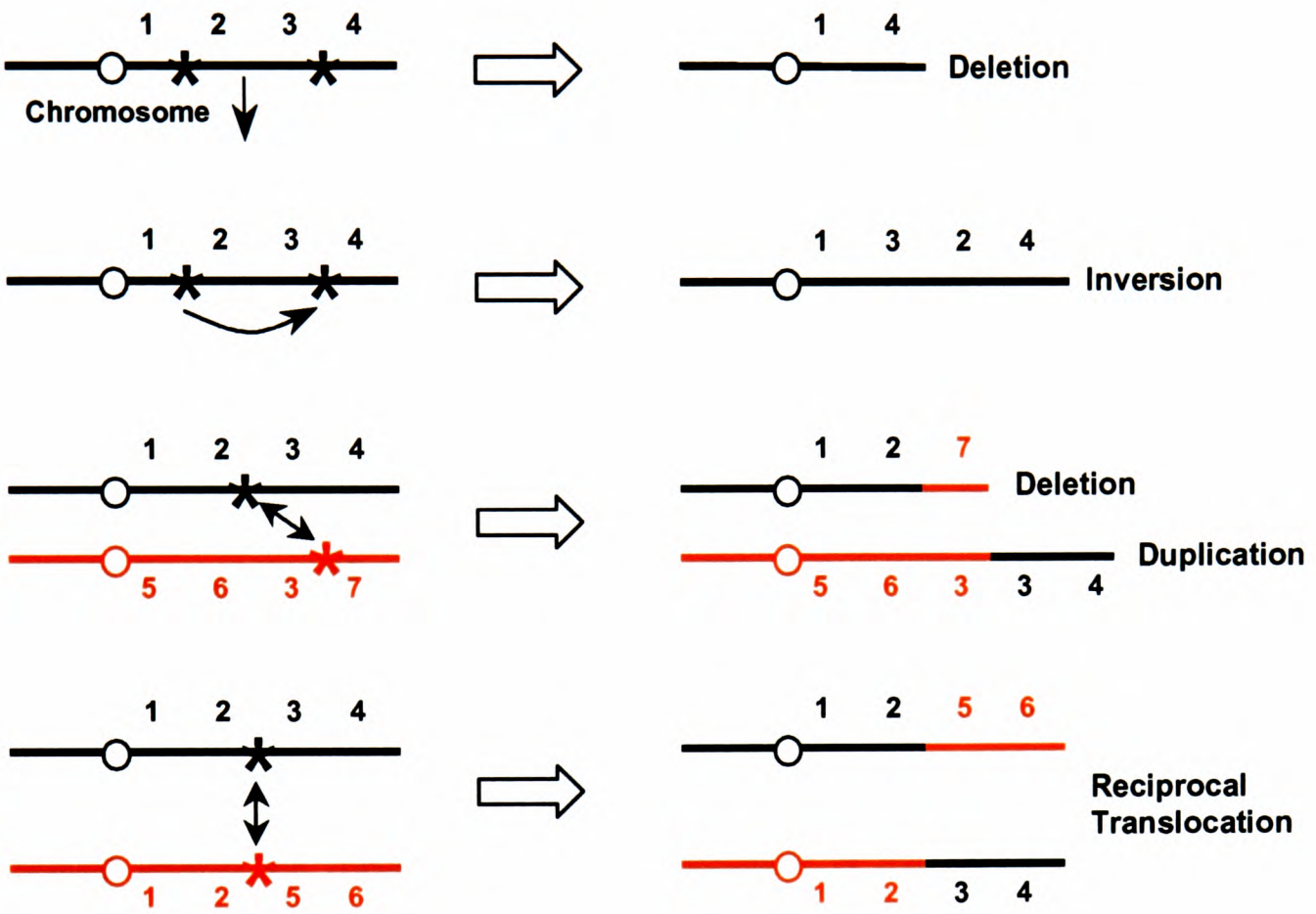


Figure 5.1 - Mechanisms of Chromosome Rearrangements : *Chromosome rearrangements may arise through the repair of double stranded breaks or the exchange of material between non-homologous chromosomes paired at regions containing repetitive sequences. Breakpoints are represented by stars and movement of DNA by thin arrows*

In order to be viable, rearranged chromosomes must contain a centromere and two telomeres. Acentric chromosomes, which lack a centromere, are not able to attach to spindle fibres during meiosis and chromosomes that lack telomeres cannot be properly replicated and are susceptible to digestion by nuclease enzymes within the cell.

5.2ii - Translocations

Although translocations do not always result in the loss or gain of genetic material, they may still result in the alteration of a phenotype. This is because the translocation breakpoints may either directly disrupt a gene, in which case the translocation often acts in an identical manner to a point mutation, or may affect the correct expression of a gene by the alteration of it's position within the genome (position effect). The larger the chromosome region involved in the

rearrangement, the more likely it is that a noticeable phenotype will arise. Translocations in humans are always carried in the heterozygous state.

The disruption of essential genes by translocations can provide a powerful tool for the mapping of disease genes. This is achieved by the positioning of the translocation breakpoints and the characterisation of those genes involved in the chromosome rearrangement (e.g. neurofibromatosis type 1 (NF1)). Neurofibromatosis is an autosomal dominant disorder that is typified by café-au-lait spots and fibromatous skin tumours that have a tendency to become malignant. The *NF1* gene was identified by the concurrent mapping of several chromosome 17 translocations found within patients affected by neurofibromatosis (Menon et al 1989, Fountain et al 1989, O'Connell et al 1989, Wallace et al 1990, Viskochil et al 1990). These studies identified a large transcript that was disrupted by, or near to, the translocation breakpoints in all the above individuals. Subsequent sequencing and cDNA walking allowed the characterisation of the remainder of the transcript (Xu et al 1990, Marchuk et al 1991), which was later demonstrated to be mutated in many neurofibromatosis patients without chromosome abnormalities. The *NF1* gene is found on chromosome 17 and has an open reading frame of 2,818 amino acids. It is a tumour supressor gene that negatively regulates the oncogene *p21* or *RAS*. The large size of the *NF1* gene renders it vulnerable to disruption and the mutation rate of the gene is one of the highest known in humans, with approximately 50% of patients presenting with novel mutations.

A second example of the use of translocation patients to aid gene discovery is the involvement of the CS case in the identification of the *FOXP2* gene (see section 4.3iii '*SPCH1* and chromosome 7q').

5.3 THE G FAMILY AND CHROMOSOME 2q

5.3i – Language Impairment and a Chromosome 2;11 translocation in the G Family

The G family is a nuclear family with four members (see figure 5.2). In this family the father and both children are affected by language deficits of some kind, whilst the mother has no history of language impairment.

The eldest child in the G family (individual G3 – see figure 5.2) was late to talk and presents with an expressive language disorder accompanied by verbal dyspraxia that results in significantly impaired articulation. Whilst his non-verbal IQ is within the normal range (90th percentile), he has a significant verbal/performance IQ discrepancy and his non-word repetition score is below the 10th percentile. His comprehension age is normal but he has significant literacy problems. He has no associated behavioural or social impairments and previously attended a language resource unit attached to a mainstream school. The younger child (individual G4) is the proband and was referred into the genome screen through the special language school that he attends. He presents with a severe expressive and receptive language impairment characterised by acute speech problems, an orofacial dyspraxia, and restricted social communication and interaction. He is clumsy, has limited cognitive and attention skills (IQ~70 on Ravens matrices) and shows a restricted range of interests and repetitive behaviours. As such, his language impairments border on the autistic spectrum and he has since been diagnosed with Pervasive Developmental Disorder (PDD). He has facial differences that distinguish him from the rest of the family and the clinician felt that this might represent a slight facial dysmorphism.

Routine cytogenetic analyses at Guys Hospital revealed a chromosome 2;11 translocation in both the mother (G2) and the proband (G4) of the G family. Their karyotype was reported to be t(2;11)(q35;p15). Neither the father (G1) nor the sibling (G3) carried this chromosome rearrangement (see figure 5.2).

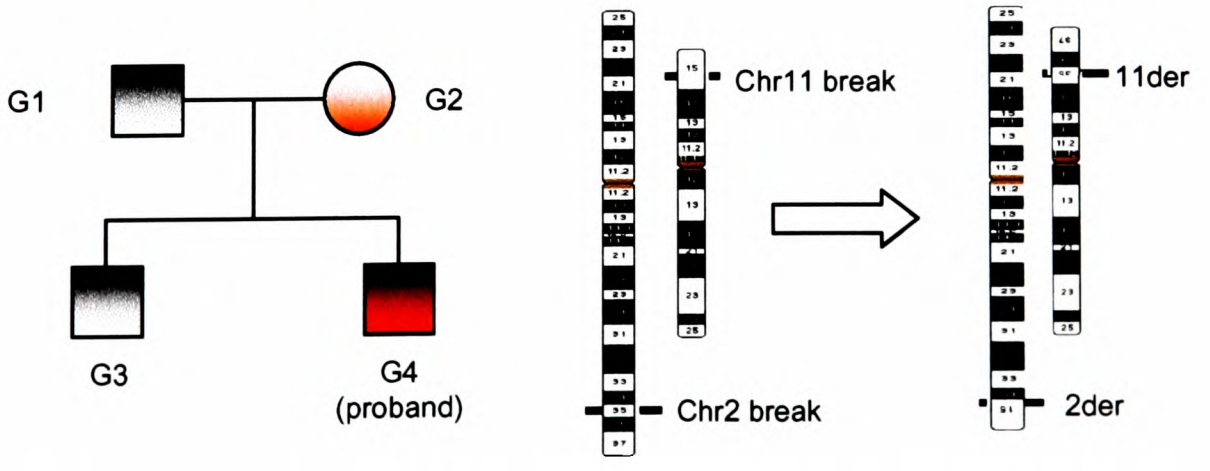


Figure 5.2 - Segregation of the Translocation and Language Impairment in the G Family : *The G family consists of four members. Males are shown as squares females as circles. Those individuals affected by language impairments are shaded in black at the top, those individuals that carry the 2;11 translocation are shaded in red at the bottom. The karyotype of individuals carrying the translocation is t(2;11)(q35;p15)*

5.3ii – Linkage to Chromosome 2q in the SLIC Genome Screen

The reported chromosome 2 breakpoint within individuals G2 and G4 was found to coincide with an area of linkage from the SLIC genome screen (see figure 5.3). The maximum LOD score within this region was 1.52 for the CELF Receptive Language Score (RLStrans) under variance components (VC) analysis (see figure 5.3). Further investigation revealed that the linkage in this region was found only in the clinical sample where it reached a maximum LOD score of 1.75 under HE analysis (see figure 5.3).

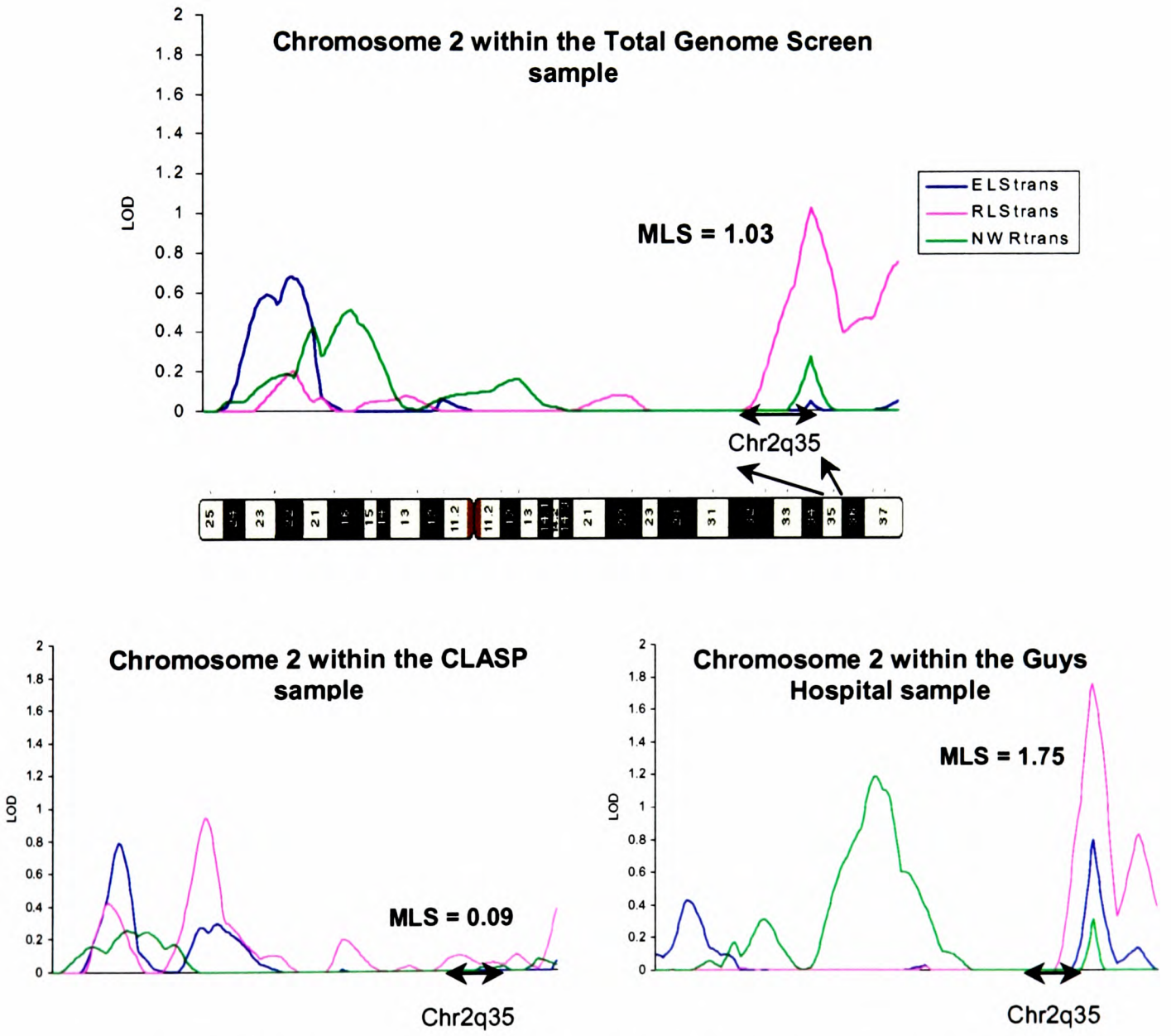


Figure 5.3 - Linkage to Chromosome 2 in the SLIC Genome Screen : *Linkage was observed to the receptive language score (RLStrans) on chromosome 2q35 within the total genome screen sample with a maximum LOD Score (MLS) of 1.52. Separation of the genome screen cohort into its constituent Guys and Cambridge groupings demonstrated that this linkage was driven primarily by the clinical (Guys Hospital) group. Graphs are shown for HE analysis only*

5.3iii – Chromosome 2q35

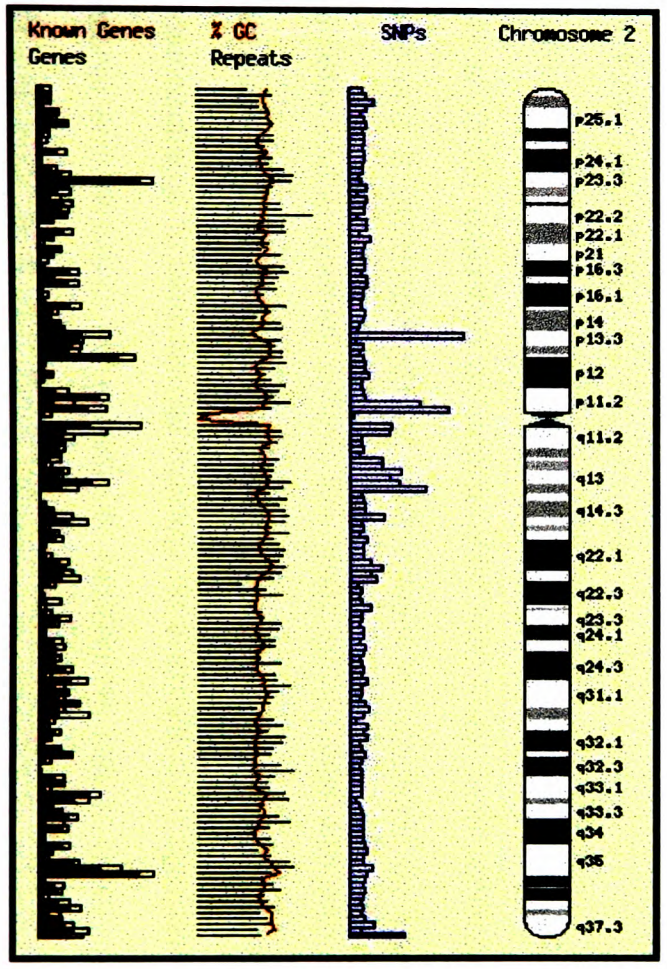
Chromosome 2q35 is moderately gene rich and contains a total of 54 mapped transcripts, of which 41 encode known genes (see figure 5.4 and appendix R). These include many brain-expressed transcripts and several genes that could be considered as ideal candidates for SLI.

5.4 CHROMOSOME 2 TRANSLOCATIONS

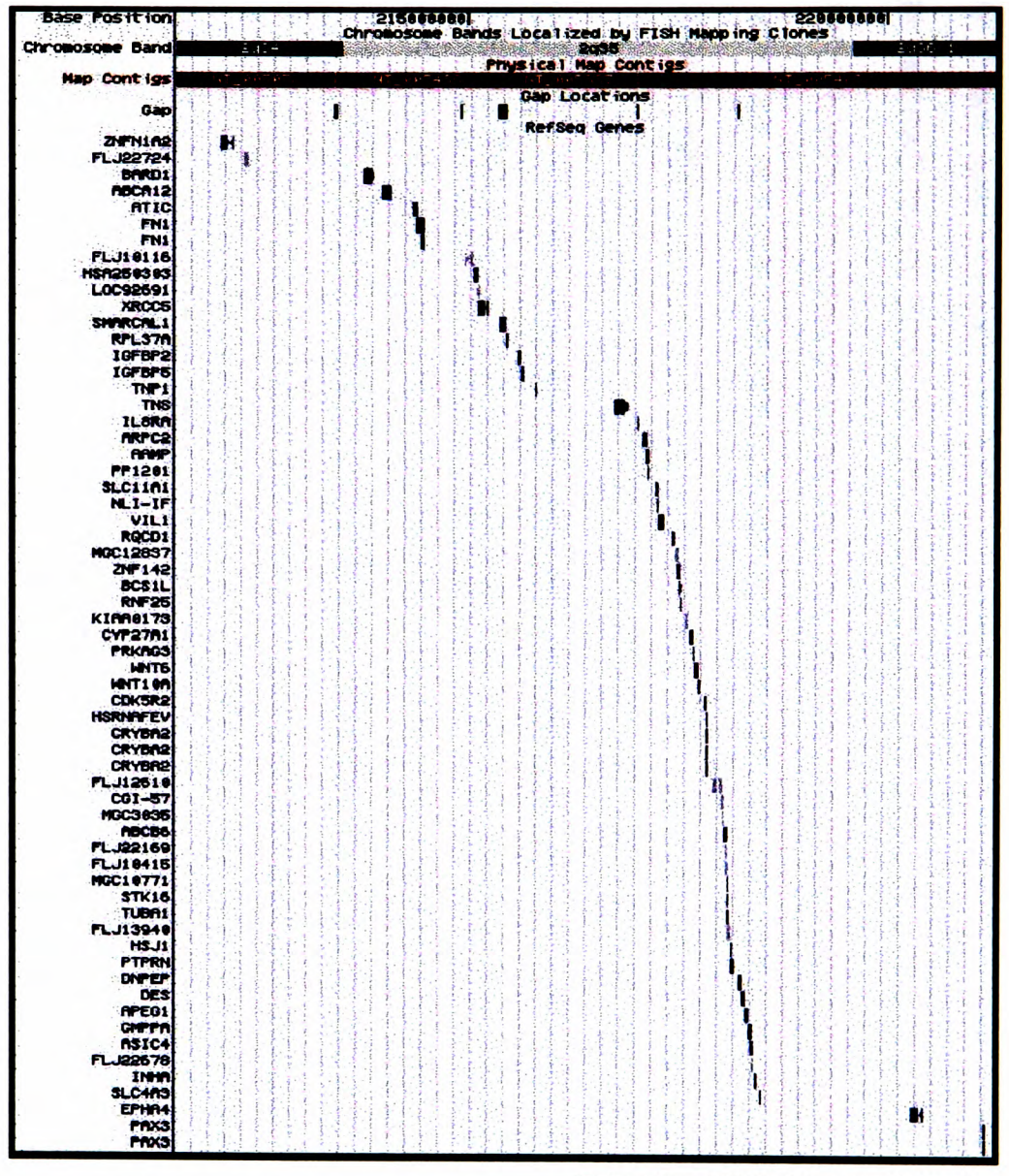
5.4i – Translocations Associated with SLI

Whilst language delays and/or developmental disorders are commonly associated with chromosome abnormalities, these characteristics usually form part of a wider syndrome and, with the exception of 7q31 (see sections 4.4iii '*SPCH1* and chromosome 7q' and 4.4v '*FOXP2* in SLI'), no one chromosome is routinely associated with a language-specific phenotype.

Two translocation cases involving chromosome 2 and specific language impairments have been described in the literature. Froster et al (1993) described a family in which a balanced translocation between chromosome 1 and 2 (t(1;2)(1p22;2q31)) cosegregated with language difficulties and dyslexia. This translocation was seen in the father and two of his sons, all of whom had a severe speech delay, for which they received speech therapy, and a reading and writing disability. All other family members (mother and 7 children) carried a normal karyotype and had normal language development. Weistuch and Schiff-Myers (1996) described a child with a severe verbal apraxia who allegedly carried a balanced translocation between chromosomes 1q and 2q.



Taken from ensembl (see websites)



Taken from UCSC (see websites)

Figure 5.4 - Gene Density and Identified Transcripts Around the Chromosome 2q35 Band

5.4ii - Autism and chromosome 2q

Amongst the several genome screens performed for autism (see section 3.2iii 'Autism'), many have shown some evidence for linkage to chromosome 2q (IMGSAC 1998, Philippe et al 1999, IMGSAC 2001a, Buxbaum et al 2001, Shao et al 2002) (see table 5.1). As with all genome screens, the exact phenotypes of the individuals used in these studies vary, as do the positions of the loci identified. However, the majority of loci appear to be around the 2q32/2q33 region.

Study	2q MLS	Position	Comments
IMGSAC (1998)	1.33	166cM	
Philippe et al (1999)	0.65	231.5cM	
Buxbaum et al (2001)	1.96	186.2cM	Using a strict diagnosis of autism
Buxbaum et al (2001)	2.99	186.2cM	In a subset of families with delayed onset of speech
IMGSAC (2001a)	4.80	206.4cM	
Shao et al (2002)	1.19	236.7cM	Study of 8 selected chromosomal regions

Table 5.1 – Evidence for an Autism Susceptibility Locus on Chromosome 2q

Of particular interest is the Buxbaum et al study (2001). These researchers found that if they restricted their analyses to those families in which two or more individuals were affected by autism and had a delayed onset of phrase speech (> 36 months), their chromosome 2q LOD score increased from 1.96 to 2.99.

In further support of the existence of an autism susceptibility gene in the chromosome 2q region, Borg et al (2002) described an autistic child with a *de novo* translocation between chromosomes 2 and 8 (t(2;8)(q35;q21.2)). The patient is described as having a severe developmental delay and is reported to have little grasp of language with only minimal use of eye contact, although he is capable of using a signing system to augment his communication. He has been demonstrated to have normal hearing. He is sensitive to high-pitched noises and shows obsessional and ritualistic behaviours. In addition he has a slightly enlarged head (macrocephaly) and minor facial dysmorphic features. The mapping of the translocation in this child confirmed the chromosome 2 to 8 translocation and also revealed a cryptic deletion that covered between 4 and 4.8Mb of chromosome 2q35. This deletion was found to include part

of the *PAX3* gene, which encodes a transcription factor involved in melanocyte production (Wantabe et al 1998, Tachibana et al 1996), and the complete coding sequence of four other genes with known functions (phenylalanine tRNA synthetase-like beta subunit (*PheHB* or *FARSLB1*), fatty acid coenzyme A ligase long chain 3 (*FACL3*), secretogranin II (*SECG2*), cullin 3 (*CUL3*) and ribonucleotide reductase M1 polypeptide (*RRM1*)). In addition nine transcription units of unknown function were also removed. Although the translocation breakpoint directly disrupted the *PAX3* gene, it is unlikely that this mutation could account for the autistic-like phenotype displayed by this child. Haploinsufficiency of *PAX3* is commonly associated with cases of Waardenburg syndrome (WS) (Wilcox et al 1992); an autosomal dominant disorder characterised by cochlear deafness, pigment defects and minor but noticeable facial abnormalities. Occasionally WS is associated with mental retardation but, in general, symptoms do not overlap with those described for autistic disorder.

Further links between chromosome 2q and autism come from reports of individuals with chromosome 2q37 telomeric deletions and a characteristic 'autistic-like syndrome' (Conrad et al 1995, Gorski et al 1989, Lin et al 1992, Wiktor et al 2001, Wenger et al 1997, Ghiaziuddin & Burmeister 1999, Young et al 1983, Wolff et al 2002). These cases typically present with severe developmental delays and distinct dysmorphic features involving frontal bossing, depressed nasal bridges and abnormal ears. Other characteristics may involve an enlarged head, cardiac murmurs, hand and feet abnormalities and low-set nipples. Many affected individuals show ritualistic behaviours and social retardation, which can lead to a diagnosis of autistic disorder. However, it is fair to say that the phenotype seen in these individuals is more complex than that usually associated with autistic disorder and they probably would not be classified as autistic under any standard diagnostic schedules.

5.4iii – Other Translocations of Chromosome 2q35

A few other translocations have been described in the literature that involve chromosome 2q35 abnormalities but which do not result in Waardenburg syndrome or autistic disorder. These cases include a range of rearrangement types and a variety of phenotypic outcomes. All individuals described have some kind of developmental delay or mental retardation and commonly present with some level of facial dysmorphism. However, aside from these

features, no clear phenotype is associated with the chromosome rearrangements reported in this region (see table 5.2).

Karyotype	Phenotype	Developmental delay / mental retardation	Microcephaly	Macrocephaly	Heart defect	Low birth weight	Micrognathia (small jaw)	Abnormal ears	Facial dysmorphia	Hand/feet abnormalities	Speech delay
Sanchez & Pantano (1984)											
45XX, rob(13;14), del(2)(q35-qter)		✓	✓	✗	–	✓	✓	✓	✓	✓	–
Dahoun-Hadorn & Bretton-Chappuis (1992)											
46XY, InvDup(2)(q35-qter)		✓	✗	✓	✗	✗	✓	✓	✓	✓	✓
Romain et al (1994)											
46XX, DirDup(2)(q33.1-q35)		✓	✗	✗	✗	✓	✗	✗	✓	✓	✓
Fritz et al (1999)											
46XY, Der(17)(17pter-17q25::2q35-2q37.1::17q25-17qter)		✓	✗	✗	✗	–	✗	✓	✓	✓	✓
Lukusa et al (1999)											
46XY, Der(14)DirIns(14;2)(q22;q32.1-q35)		✓	–	–	–	✗	✗	✓	✓	✓	✓
Borg et al (2002)											
46XY, t(2;8)(q35;q21.2)		✓	✗	✓	–	✓	–	–	✓	✗	✓
G Family											
46XY, t(2;11)(q35;p15)		✓	✗	✗	✗	✗	✗	✗	✓	✗	✓

Table 5.2 – Other Reported Translocations Involving Chromosome 2q35 : *Studies of chromosome 2q35 translocations and associated phenotypes. A tick indicates that phenotype is present, a cross indicates that phenotype is absent and a dash indicates that the presence or absence of the phenotype was not reported*

In summary, there have been cases of language impairment linked with chromosome 2 translocations and reports of 2q abnormalities in both autistic and autistic-like' syndromes. Additional, unrelated, phenotypes, such as facial dysmorphism and abnormal ears, are also found to be common in individuals with chromosome 2q rearrangements.

5.5 – CHARACTERISATION OF THE TRANSLOCATION WITHIN THE G FAMILY

Information from the NCBI, Ensembl and UCSC databases was used to compile an electronic BAC contig across chromosome 2q35. Four BACs were chosen from across the contig, two from the extremes and two from the middle, to use as probes in the first round of FISH analysis (see section 2.11 'Fluorescence in-situ hybridisation' for FISH methods). Subsequent rounds then used the flanking markers from the previous round as anchors, thus allowing a reduction in the size of the critical region with each round (see figure 5.5).

Whenever possible, BACs were chosen which were completely sequenced and positioned within the surrounding sequence. However, as the intervals became smaller this became more problematic and partially sequenced clones were used. All probes were tested on metaphase slides prepared from a normal male adult control and individual G4 (the proband from the G family - see figure 5.2). Any clone that was found to be non-specific or which annealed to a chromosome location other than 2q was not used for subsequent rounds of analyses.

Finally, three SNPs, spread across the clone which covers the breakpoint, were chosen from dbSNP, typed via a restriction enzyme assay and tested for association to the ELStrans, RLStrans and NWRtrans phenotypes within the Guys Hospital cohort (see sections 2.2 'The genome screen phenotypes' and 2.10 'Single Nucleotide Polymorphisms (SNPs)'). The sequences of each SNP region and details of the primers and restriction enzymes used can be found in appendices S and T.

5.6 FISH RESULTS

5.6i - FISH Mapping of the G Family Translocation

In total, four rounds of FISH were performed (see figure 5.5). The fourth, and final, round of these analyses involved four overlapping BACs and allowed the positioning of the chromosome 2 breakpoint to a single clone (RP11-105N14) which lay within the chromosome 2q34 band (see figures 5.5 and 5.6). The translocation was therefore reclassified as t(2;11)(q34;p15).

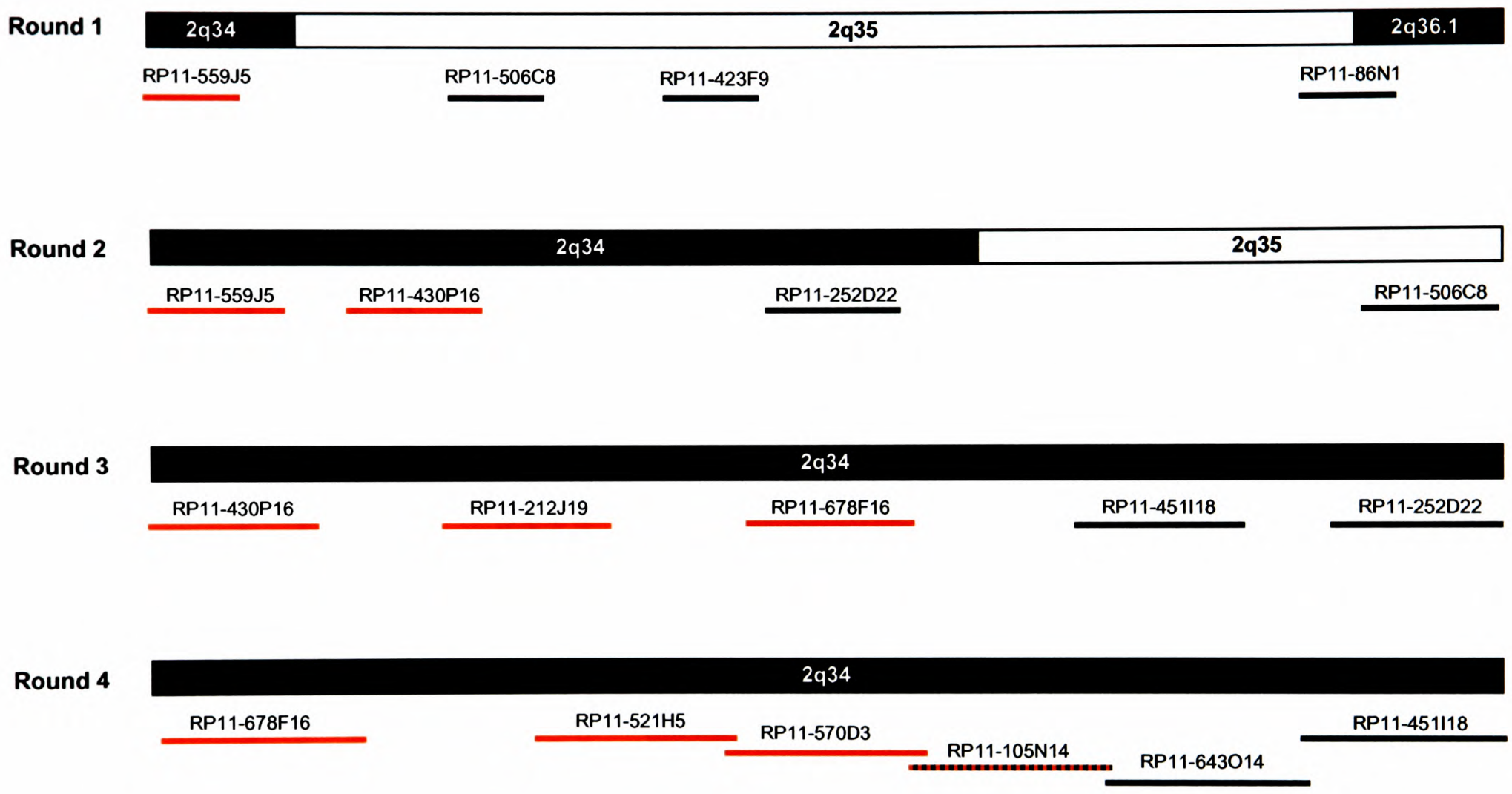


Figure 5.5 - Rounds of FISH Performed on Chromosome 2q35 : *Probes shown in red annealed proximal to the breakpoint in individual G4, whilst those shown in black annealed distal to the breakpoint. All probes annealed to chromosome 2 in a normal male control. Chromosome bands and location of probes are given only as a guide. They are not to scale*

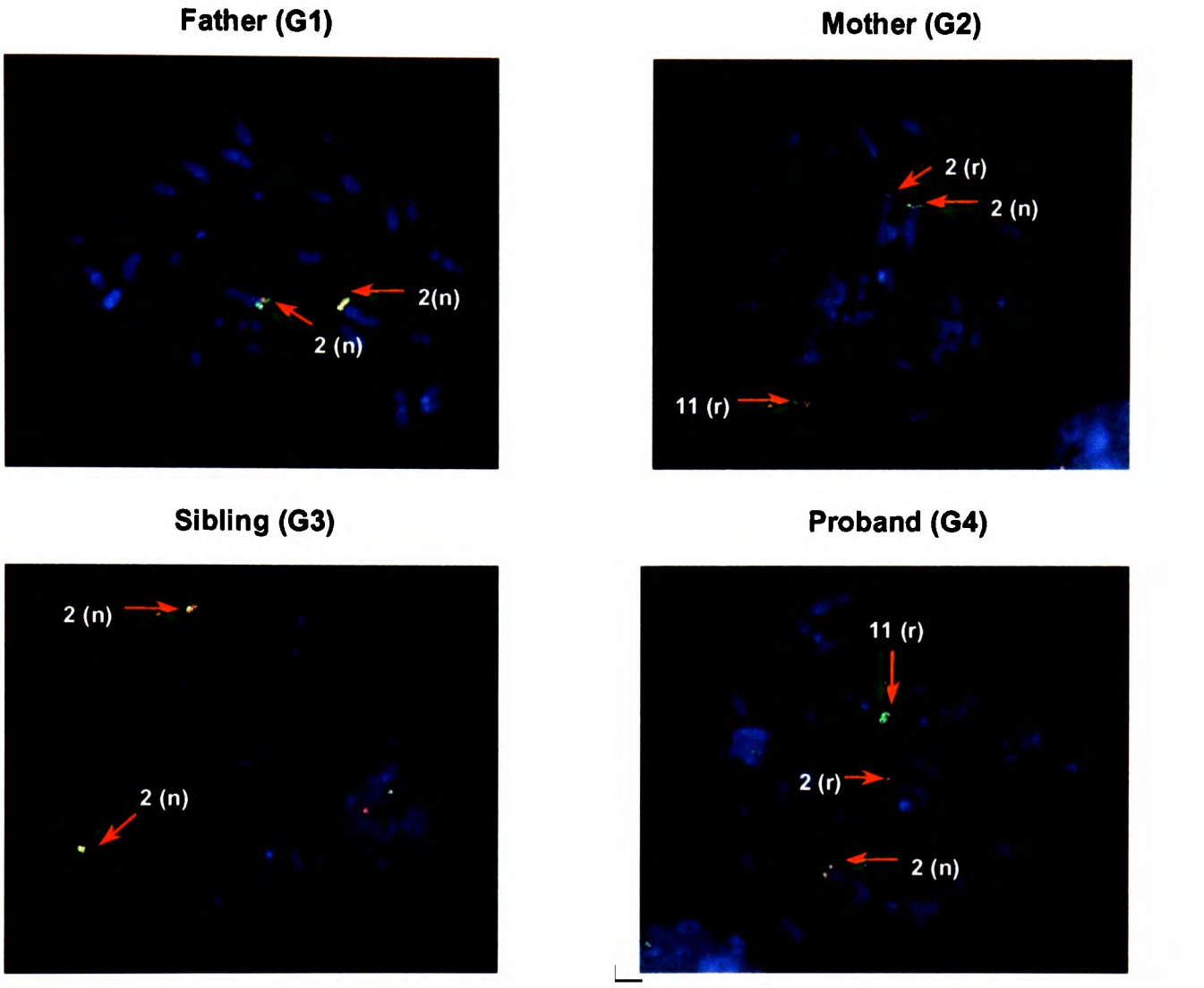


Figure 5.6 - Identification of the Breakpoint within the G Family : *RP11-678N14 anneals to the rearranged chromosome 11 (11 (r)) and the rearranged chromosome 2 (2 (r)) in both the mother and the proband and therefore falls across the breakpoint in these individuals. The father and the sibling carry a pair of normal chromosome 2 (2 (n)), to which both probes anneal*

5.6ii – Segregation of the 2;11 Translocation within the G Family

The FISH analyses confirmed that the translocation was passed from the mother (G2), who has no history of language delay, to the proband (G4). The father (G1) and the sibling (G3), both of whom were reported to have language difficulties, did not carry the translocation (see figure 5.6).

5.6iii – Sequence Analysis of Clone RP11-105N14

Sequence analyses revealed the breakpoint clone, RP11-105N14, to be 161Kb in length and to overlap slightly with two other clones, one proximal (RP11-570D3) and one distal (RP11-

643O14). In addition, it was found that RP11-105N14 contains the 5' end of a gene (*ZNF1A2*) encoding a zinc finger DNA binding protein (Helios) and the first three exons of a putative transcript (*FLJ22724*). The position of both these transcripts can be seen in figure 5.7.

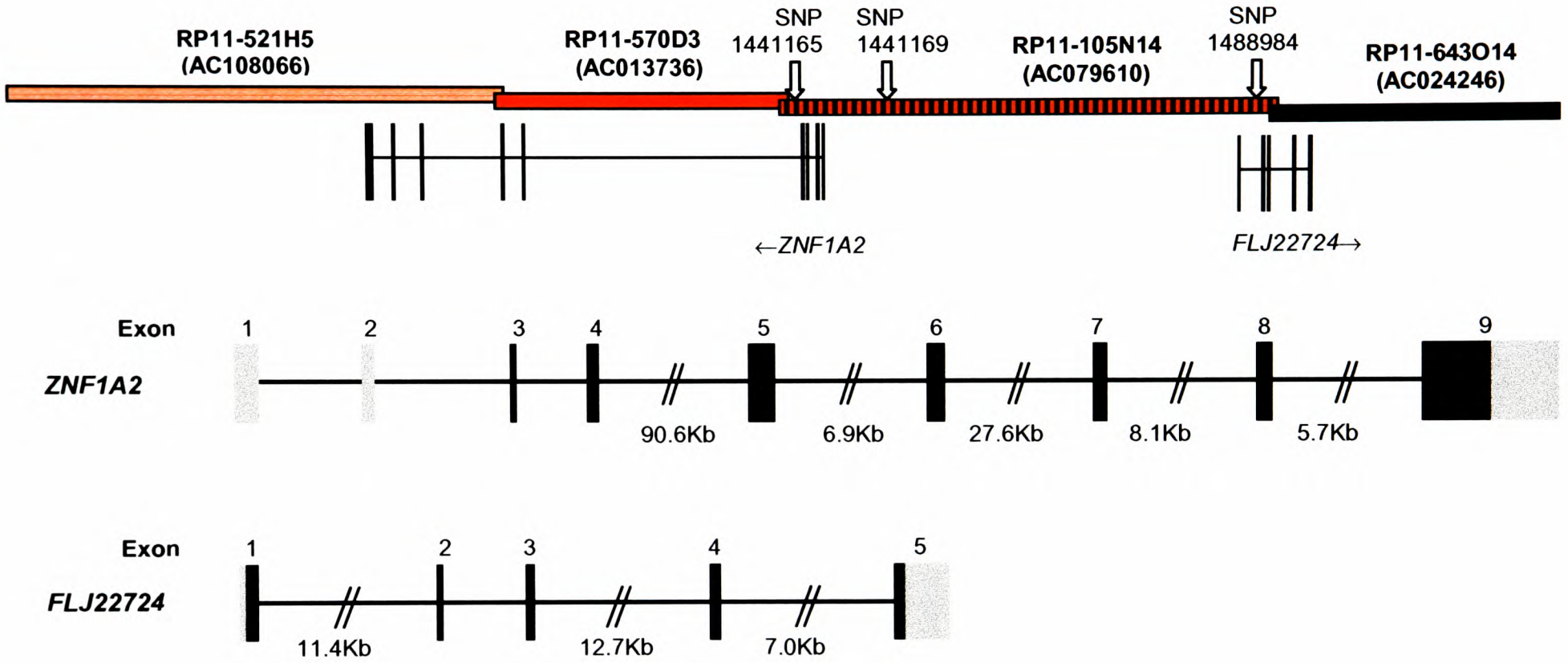


Figure 5.7 - RP11-105N14 and Associated Clones and Transcripts : *RP11-105N14 spans the chromosome 2q34 breakpoint in the G family. Clones shown in red fall proximal to the breakpoint and those shown in black fall distal to the breakpoint. In the clone diagram, all clones, introns and exons are shown to scale. Arrows indicate the positions of the three SNPs used for association analyses within the Guys Hospital cohort. The exact positions, and sequences surrounding these SNPs can be found in appendix S. RP11-105N14 contains the first four exons of the zinc finger gene ZNF1A2, and the first three exons of the putative transcript FLJ22724. In the gene diagrams, all introns larger than 5Kb are not to scale but are marked with slashes and the appropriate intron size. Coding exons are shown as black boxes, non-coding exons as grey boxes*

5.6iv – Association Analyses

No positive associations (P<0.01) were found between any of the three SNPs typed from the RP11-105N14 sequence (see table 5.3) and the three quantitative measures of language abilities tested (ELStrans, RLStrans and NWRtrans — see section 2.2 'The genome screen phenotypes').

		SNP		
		1441165	**1441169**	**1488984**
No. Individuals Typed		186	192	138
Heterozygosity (%)		54.3%	54.7%	43.5%
ELStrans	χ^2	0.30	0.11	1.11
	p	>0.1	>0.1	>0.1
RLStrans	χ^2	4.35	4.84	2.05
	p	0.07	0.05	>0.1
NWRtrans	χ^2	1.51	0.37	0.13
	p	>0.1	0.06	>0.1

Table 5.3 - Association Analyses of RP11-105N14 SNPs : *P-values given are empirical p-values derived from 1000 simulations of all markers typed*

5.7 DISCUSSION

A chromosome 2q translocation was observed within a single proband in the clinical (Guys Hospital) sample collected for the SLI genome screen. This chromosome rearrangement coincided with a region of linkage identified by the genome scan. Furthermore, chromosome abnormalities within this area have been investigated in connection with both language impairments (Froster et al 1993, Weistuch and Schiff-Myers (1996)) and autistic disorder (Borg et al 2002, Wolff et al 2002, Ghaziuddin & Burmeister 1999). The present study therefore chose to map this translocation within the proband and all available family members in the hope that it may allow the identification of a candidate gene for SLI.

FISH analyses allowed the identification of a single BAC clone, RP11-105N14 (AC079610), that spanned the translocation breakpoint and contained two known transcripts, *ZNF1A2* and *FLJ22724* (see figure 5.7).

The *ZNF1A2* gene spans approximately 145Kb of genomic sequence and is expressed mainly in T-cells where it produces a zinc-finger DNA-binding protein known as Helios. In

humans, *ZNF1A2* encodes two transcripts, one with five zinc finger domains (1.50Kb transcript) and a second containing six zinc finger domains (1.58Kb transcript) (Hosokawa et al 1999). The Helios proteins belong to a family of transcription factors that play an important role in lymphocyte differentiation and function.

In haematopoietic cells, Helios forms dimers either with itself or with one of two closely related proteins (Ikaros and Aiolos). Active dimers of these proteins are seen to form ring-like structures within the nuclei of lymphocytes and are thought to recruit genes, which are destined for inactivation, to the centromeric foci (Hahm et al 1998, Kelley et al 1998). In addition it has been suggested that Helios may play a more traditional role as a transcriptional activator of downstream genes which have yet to be identified (Kelley et al 1998).

During mouse embryogenesis, *Helios* expression is found across several tissues including the liver, thymus, kidney, and salivary ducts (Kelley et al 1998). In adult mice, expression is more limited and is found mainly in the thymus with low levels of expression also detectable in the bone marrow and brain (Hahm et al 1998, Kelley et al 1998). In human tissues, *HELIOS* expression is limited mainly to the peripheral blood leukocytes, and is also observed in the skeletal muscle, placenta and lung (Hosokawa et al 1999). No expression has been reported in human brain.

Studies of patients with T-cell acute lymphoblastic leukaemia, indicate that the *HELIOS* gene may be overexpressed in certain forms of carcinomas (Nakase et al 2002).

FLJ22724 is a hypothetical gene whose existence is predicted from cDNA sequencing and EST analysis. The gene covers 33Kb of genomic sequence and is predicted to encode a transcript of 0.5Kb, which is expressed in small intestine, pancreas, brain, prostate, liver and kidney.

Although the function of the *FLJ22724* protein is unknown, the gene sequence is 31% homologous to that of the Drosophila *fuzzy* (*fy*) gene. *Fuzzy* is a transmembrane protein involved in the establishment and maintenance of cell polarity. Mutants of the *fuzzy* gene have abnormalities in the number and polarity of hairs on their abdomen, thorax, wings and legs (Collier & Gubb 1997).

In summary, neither of the genes contained in BAC RP11-105N14 represent ideal candidates for involvement in SLI. The expression of *ZNF1A2* is limited primarily to T cells, and the function of *FLJ22724* remains unknown.

Although it was initially felt that the probarid in the G family was suffering from SLI, further profiling revealed some characteristics which are more typically associated with autism than language impairment (e.g. repetitive and obsesssional behaviours) and the diagnosis was amended to Pervasive Developmental Disorder Not Otherwise Specified (PDDNOS). It therefore remains possible that the translocation identified here may combine with other modifying loci to generate the autistic qualities observed in this child, which are not present in his sibling. However, the translocation breakpoint identified in this family is considerably distant from any chromosome locus previously implicated in autism and lies approximately 8Mb proximal of *PAX3* and the other four genes deleted in a patient with autistic disorder (Borg et al 2002) (see figure 5.4 for relative positions of *ZNF1A2* and *PAX3* genes).

In a final analysis of association across the region containing the translocation breakpoint, three SNPs were typed within the Guys Hospital cohort. However, none of these SNPs were significantly associated with any of the language phenotypes studied.

In conclusion, therefore, it appears unlikely that the chromosome abnormality within the G family has a direct effect upon the onset of language impairment in the proband.

CHAPTER 6

Summary, Appraisal & Future Directions

SUMMARY, APPRAISAL AND FUTURE DIRECTIONS

Approximately 4% of English-speaking children are affected by Specific Language Impairment (SLI); a disorder in the development of language skills despite adequate opportunity and normal intelligence. Several studies have indicated the importance of genetic factors in SLI; a positive family history confers an increased risk of development, and monozygotic concordance consistently exceeds that of dizygotic twins. However, like many behavioural traits, SLI is assumed to be genetically complex with several loci contributing to the overall risk. As such, the mapping of the gene variants underlying this disorder will require the use of large sample sizes alongside model-free non-parametric analyses. This thesis describes the completion of the first-ever, systematic genome-wide screen for loci involved in SLI. A collection of 98 families was drawn from both epidemiological and clinical populations, all with probands whose standard language scores fall at least 1.5 SD below the mean for their age. Genome-wide linkage analyses of three language-related measures (CELF-R receptive and expressive scales and non-word repetition) identified two regions which may harbour susceptibility gene variants for SLI, one on chromosome 16 and one on 19. Simulations suggest that of these two regions, the non-word repetition (NWR) linkage to chromosome 16q is the most significant, with empirical p values reaching 10^{-5} under both Haseman-Elston (HE) (LOD=3.55, p=0.00003) and variance components (VC) (LOD=2.57, p=0.00008) analyses. Singlepoint analyses provided further support for this locus, with three markers under the peak of linkage yielding LOD scores of greater than 1.9. The 19q locus was linked to the CELF-R expressive language score (ELS) and exceeds the threshold for suggestive linkage under all types of analysis performed (multipoint HE (LOD=3.55, empirical p=0.00004) and VC (LOD=2.84, empirical p=0.00027) and singlepoint HE (LOD=2.49) and VC (LOD=2.22)). Furthermore, both the clinical and epidemiological samples showed independent evidence of linkage on both chromosomes 16q and 19q indicating that these may represent universally important loci in SLI and thus general risk factors for language impairments.

The main caveat of the genome screen approach tends to be the reliability of the results obtained. Whilst the evidence supporting the chromosome 16 and 19 loci is strong, this was

the first such screen to be completed for SLI and it should be noted that even linkages of this magnitude might be expected to happen by chance at least once in every full genome screen. Similar studies for other complex disorders indicate that the real test of linkage validity often lies in the replication of the loci detected. In order to address this issue, the SLI Consortium (SLIC) is in the process of collecting further language-impaired families for a second wave of genotyping. Aside from the Guys Hospital and CLASP collections, which are both ongoing, these extra families are being recruited by three further sites distributed around the UK (see appendix A). This wave 2 cohort currently consists of 87 small nuclear families, 393 individuals and 248 sib-pairs. In addition, it is hoped that investigations by other groups will provide support for the chromosome 16 and 19 loci identified here.

A second aim of the SLI Consortium is to develop the phenotypic conceptualisation of SLI and to identify those indices that best measure the underlying impairment in affected individuals. Although the SLIC assessment battery consists of many different measures of language abilities (e.g. each of the CELF subtests), the current study chose to use just three of these for the genome screen analysis. This decision was taken primarily to avoid multiple testing issues, but may have resulted in the loss of a large amount of information with regards to the language abilities of the families involved. A more ideal situation would be presented if it was possible to delineate all the test results into one or two measures whilst maintaining the full variance of the scores obtained in the original assessment. Investigations are therefore underway into the use of multivariate methods for the analysis of the phenotypic and genetic data collected for wave 1 of the SLI genome screen. In a similar manner to the linkage analyses described here, multivariate methods exploit a variance components type model to detect loci underlying QTL measures. However, in addition to the trait variances considered in a univariate system, multivariate techniques also aim to model covariations between traits and thus afford more power than the traditional QTL mapping systems. Furthermore, because multivariate techniques consider many measures at once they are more robust to stochastic variations within a single measure than univariate analyses. Initial investigations of the SLIC QTL data indicate that some of the CELF sub-tests cannot be easily mapped onto a simple expressive/receptive language model and probably reflect different fine-grain aspects of the

processes underlying language development. It is hoped that multivariate analyses will not only allow the further evaluation of the genetic effects at chromosome 16 and 19 loci but will also clarify the relationships between the factors underlying SLI and perhaps language development in general.

A second problem inherent to the genome screen approach is that the size of the regions identified by linkage analyses can be considerable. The larger of the peaks on chromosome 16 covers approximately 40cM and the chromosome 19 locus is approximately 30cM in size. Attempts to fine map the chromosome 19 locus using a high-density map of microsatellite markers failed to significantly reduce the size of this region. As a consequence of the large size of the regions, both the chromosome 16 and chromosome 19 loci contain a large number of genes, a fair proportion of which could be argued to represent good candidates for SLI. The mutation screening of these genes, such as that described for *NUMB-R*, requires a lot of DNA and is time consuming. An alternative approach that may allow the better characterisation of loci would be the association mapping of each region using a high density of SNPs. However, high-throughput SNP typing techniques are relatively new and many of the polymorphisms currently described in the SNP database still require validation (see table 2.10 'SNP validation'). In addition, as discussed in chapter 4 ('Candidate gene analyses'), studies indicate that, prior to the initiation of association-based investigations, it is a good idea to first examine the level of linkage disequilibrium in the region. Such investigations typically involve the sequencing of the chromosomal region in a number of controls and the construction of haplotype and linkage disequilibrium (LD) maps across the area. Since the physical size of the chromosome 19 locus is approximately 17Mb, this approach is restricted by practical limitations. However, given that this strategy has proven successful for other common and complex disorders (Crohn's Disease and asthma – see below), some LD mapping in a restricted section of the locus may be worth consideration.

The identification of a major susceptibility gene for Crohn's Disease (CD) in 2001 represented the end-point of a long process that started with a genome screen and illustrated the first such breakthrough in complex genetics. Crohn's Disease is a common medical condition that involves a chronic inflammation of the gastrointestinal tract. Clinical symptoms include

abdominal pain, diarrhoea, rectal bleeding and weight loss. The symptoms of CD are very similar to those of a closely related disorder known as Ulcerative Colitis (UC), and together these two disorders form the Inflammatory Bowel Diseases (IBD). The evidence for a genetic aetiology in CD has been clearly demonstrated with familial clustering and an increased concordance in monozygotic twins. However, as for many common disorders, the genetics underlying the disease were predicted to be complex in nature. Between 1996 and 2001, several genome screens were performed for both IBD and CD (Hugot et al 1996, Satsangi et al 1996, Cho et al 1998, Hampe et al 1999, Ma et al 1999, Duerr et al 2000, Rioux et al 2000). Between them, these screens implicated two main loci, one on chromosome 12q (*IBD2*) and a second in the centromeric region of chromosome 16 (*IBD1*). Targeted replication studies of these two linkages indicated that whilst the *IBD1* locus was strongest within CD patients, individuals with UC mainly contributed to the *IBD2* locus (Lesage et al 2000, The IBD International Genetics Consortium 2001). The region of chromosome 16 containing the *IBD1* locus was large (~20Mb) and contained many good candidate genes. Different groups therefore employed complementary approaches in an attempt to identify the gene underlying the linkage. In 2001, two of these groups simultaneously published papers implicating the *NOD2* (or *CARD2*) gene in the onset of CD. One team used a candidate gene approach to the identification of *NOD2* (Ogura et al 2001), whilst the others employed fine mapping linkage and association based techniques (Hugot et al 2001). The *NOD2* gene encodes a protein involved in apoptosis and activation of the nuclear factor NF-κB, which is involved in inflammatory pathways.

Similarly, a genome screen/association study has recently identified a gene that may be involved in susceptibility to the common disorder of asthma (Van Eerdewegh et al 2002). The genome screen completed by these researchers yielded a maximum LOD score of 3.93 to a 30cM region of chromosome 20p13. Interestingly, this locus did not coincide with any of the other major loci previously implicated in asthma. However, association studies of polymorphisms within a 1 LOD-interval of the peak revealed a region of positive association surrounding the *ADAM33* gene (p=0.03-0.006). Haplotype and family studies further supported the role of this gene in the onset of asthma. In total, 55 polymorphisms were

identified in *ADAM33*, many of which were significantly associated with an increased susceptibility to asthma within the populations studied (American and English). In general these 'susceptibility polymorphisms' were found to be common variants, with allele frequencies ranging from 20% to 95%. Note, however, that this study has yet to be confirmed in other samples. The *ADAM* genes are a large family of membrane-bound metalloproteases that have a diverse range of functions. The exact function of *ADAM33* remains unclear but it has been shown to be highly expressed in the lung (Van Eerdewegh et al 2002).

The above success stories provide encouragement for researchers studying the genetics of complex disease and validate the genome screen/association mapping approach as a worthwhile strategy in the identification of the genes underlying such disorders.

An alternative approach to the identification of the genes underlying the chromosome 16 and 19 loci may be the study of translocation cases within these regions. Although the translocation case described here (see chapter 5 'FISH mapping of a translocation in a child with language impairment') did not enable the selection of any genes for further investigation, this is a proven means of gene identification in monogenic disorders (e.g. the CS case and neurofibromatosis) that may also prove to be useful for more complex diseases. The cytogenetic department at Guys Hospital maintains a database of individuals with chromosomal rearrangements and has agreed to refer to us any language-impaired individuals who carry chromosome 16 or 19 rearrangements for further study.

Finally, it is possible that the study of candidate genes chosen because of their involvement in related disorders, rather than by their position within the genome, may allow the identification of the polymorphisms underlying specific language impairment. Although the mutation screening of *FOXP2*, described in chapter 4 ('Candidate gene analyses'), did not reveal any polymorphisms which may account for susceptibility to SLI within our sample, recent studies have characterised some of the promoter regions and additional exons within this gene and these will also require screening. Furthermore, it is likely that the study of the *FOXP2* pathway will also identify downstream targets and associated genes whose role in specific language impairment will need to be characterised.

The ultimate goal of any genome screen project is the identification of functional polymorphisms underlying the disorder of interest. Understandably, the course of the work following the detection of these gene variants can only be speculated and, to a large extent, will depend upon the role of the proteins encoded by those genes and the amount of existing knowledge with regards to the gene function. Aside from the SLI Consortium, we currently collaborate with two epidemiological studies, both of which hold DNA samples and language information for a large sample of children. The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal investigation which aims to 'understand the way in which physical and social environment interact, over time, with the genetic inheritance to affect the child's health, behaviour and development'. The ALSPAC cohort consists of 14,000 children born in the Avon area between the 1st of April 1991 and the 31st of December 1992. All of these individuals were invited to partake in tests of reading, spelling and language abilities at various time points between the ages of 7 and 9. In addition a sub-set of approximately 1000 children were selected from this study to complete more intensive language tests at the 25-month and 61-month time points (see website for further information). The Twins Early Development Sample (TEDs) is a sample of over 7000 twin-pairs born in England and Wales in 1994 (Dale et al 1998). Approximately 4000 of these twin-pairs were assessed at 4-years of age for signs of language impairment and it is proposed that all available children will be further screened for language and behavioural problems at 8-years of age. Access to the ALSPAC and TEDs samples will allow the use of larger language-impaired samples for the validation of any putative polymorphisms identified. Furthermore, since both these cohorts are epidemiological in nature, they will enable the evaluation of the frequency and impact of such variants within a range of children spread across the entire spectrum of language abilities.

Once the genes underlying language impairment have been identified, their expression patterns can be studied, both within a range of tissues and across a series of time points in development, using in-situ hybridisation techniques. It is hoped that such studies may highlight the brain regions and neural structures that are crucial for the proper development of language. The role of the gene within these given areas can then be further characterised by the use of expression constructs and protein-specific antibodies. In addition, it will be possible

to identify upstream regulators and downstream targets of those gene products and thus build a pathway of gene expression elucidating those cellular mechanisms which play an important role in the language acquisition process.

Although it may be naïve to expect animal models to yield a realistic prototype for the language acquisition process in humans, such animals will certainly facilitate a relatively straightforward and comprehensive study of the expression of relevant genes within a model system. Furthermore the generation of transgenic or knockout animals may aid the development of, and endorse any assumptions made by, the genetic model.

Lastly, it will be possible to use these gene discoveries to study the evolutionary development of language. Sequence comparisons of gene products both between humans and other primates, and within conserved regions across more diverse species, may increase our understanding of how human language evolved and provide some interesting insights into the processes which underlie language acquisition.

In conclusion, the genome screen described in this thesis provides an overview of the whole genome with respect to language-related phenotypes and highlights two loci that appear to have a significant effect on the development of SLI. This work represents the first step in the clarification of the genetic mechanisms behind SLI, which, it is hoped, may eventually lead to the identification of the gene variants which render some people susceptible to the onset of language impairments. In time, the study of these genes will provide a fascinating insight into the mechanics of language and should facilitate better diagnosis and treatment for individuals with SLI.

ATTRIBUTIONS

This project involved a whole team of individuals, to all of whom I am extremely grateful.

The Guys Hospital families were ascertained and phenotyped by Gillian Baird, Leila Jannoun and Vicky Slonims. The CLASP sample was ascertained and phenotyped by Ian Goodyear, Patrick Bolton, Melanie Merricks and Carol Stott.

Whilst James Cleak and Yumiko Ishikawa-Brush helped with the DNA extraction, preparation and genotyping, I performed all the error checking, binning, data management and analyses for the genome screen.

Simon Fisher and Cecilia Lai characterised *FOXP2* and designed and optimised the primers used for the DHPLC analysis of this gene.

Janine Lamb performed all the microsatellite detection, typing and association analyses for *FOXP2*.

Cecilia Lai, Andrew Jeffersen and Elaine Levy aided and instructed me in all the techniques used for FISH.

WEBSITES

Website	Reference	Website Address
AB Gene	Technical products	http://www.abgene/com
Alpha Innotech Corporation	UV transilluminator	http://www.alphainnotech.com
Applied Biosystems (ABI)	Technical products	http://www.appliedbiosystems/com
The Avon Longitudinal Study of Parents and Children (ALSPAC)	ALSPAC	http://www.alspac.bris.ac.uk
BCM Search Launcher Website	RepeatMasker software	http://searchlauncher.bcm.tmc.edu:9331/seq-util/seq-util.html
Beckman	UV spectrophotometer	http://www.beckman.com
Center for Statistical Genetics, University of Michigan	SIBMED	http://www.sph.umich.edu/statgen/software
Children's Hospital Oakland Research Institute (CHORI)	BACs	http://www.chori.org
Corning	Technical products	http:www.corning.com
Division of Statistical Genetics, University of Pittsburgh	MEGA2	http://watson.hgen.pitt.edu/mega2.html
Ensembl	Genetic information	http://www.ensembl.org
European Collection of Cell Cultures (ECACC)	Cell culture and transformation	http://www.ecacc.org.uk
Généthon.	Genetic information	http://www.genethon.fr/
Genomatrix	Promoter inspector	http://arithea.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl
Genome Database (GDB)	Genetic information	http://www.gdb.org
Genomica Corporation	Discovery Manager™ system	http://www.genomica.com
Human Genome Mapping Project Resource Centre	Genetic information	http://www.hgmp.mrc.ac.uk
Marshfield Center for Medical Genetics	Genetic information	http:/research.marshfieldclinic.org/genetics
Millipore	Technical products	http://www.millipore.com
MJ Research	Technical products	http://www.mjr.com
MRC Human Genome Mapping Project Resource Centre	Nix	http://www.hgmp/mrc.ac.uk

Website	Reference	Website Address
MWG Biotech	Primers	http://www.mwg-biotech.com
National Center for Biotechnology Information (NCBI)	Genetic information	http://www.ncbi.nlm.nih.gov
New England Biolabs (NEB)	Restriction enzymes	http://www.neb.com
Nucleon Biosciences	Technical products	http://www.tepnel.com
Online Mendelian Inheritance in Man (OMIM)	Genetic information	http://www.ncbi.nlm.nih.gov/Omim
Oxford University Bioinformatics centre	PromoterScan	http://www.molbiol.ox.ac.uk/promoterscan.htm
Robbins Scientific	Hydra96	http://www.robsci.com/
Rockfeller University	SIMULATE	ftp://linkage.rockefeller.edu/software/simulate
Sigma	Technical products	http://www.sigmaaldrich.com
SNP Database	SNPs database	http://www.ncbi.nlm.nih.gov/SNP/
Stratagene	Technical products	http://www.stratagene.com
The Cooperative Human Linkage Center (CHLC)	Marker maps	http://gai.nci.nih.gov/CHLC/
The Wellcome Trust Centre for Human Genetics	QTDT	http://www.well.ox.ac.uk/asthma/QTDT/index.html
Transgenomic	WAVE	http://www.transgenomic.com
University of California, Santa Cruz	USCS	http://genome.ucsc.edu/
Vector Labs	Technical products	http://www.vectorlabs.com
Vysis	Technical products	http://www.vysis.com
Weizmann Institute of Science Genome and Bioinformatics	RECODE (version 1.4)	ftp://bioinformatics.weizmann.ac.il/pub/software/linkage_and_mapping/dweeks/
Whitehead Institute for Biomedical Research, MIT	GENEHUNTER (version 2.0)	http://www.genome.wi.mit.edu/ftp/pub/software/genehunter
Whitehead Institute for Biomedical Research, MIT	MAPMAKER/SIBS	ftp://ftp-genome.wi.mit.edu/distribution/software/sibs

BIBLIOGRAPHY

Abecasis, GR, Cardon, LR & Cookson WO (2000) A General Test of Association for Quantitative Traits in Nuclear Families. American Journal of Human Genetics, 66, 279-292

Abecasis, GR, Cookson, WOC & Cardon, LR (2001a) The power to detect linkage disequilibrium with quantitative traits in selected samples. American Journal of Human Genetics, 68, 1463-1474

Abecasis, GR, Noguchi, E, Heinzmann A, Traherne JA, Bhattacharyya, S, Leaves, NI, Anderson, GG, Zhang, Y, Lench, NJ, Carey, A, Cardon, LR, Moffatt, MF & Cookson, WO (2001b) Extent and distribution of linkage disequilibrium in three genomic regions. American Journal of Human Genetics, 68, 191-197

Accili, D, Fishburn, CS, Drago, J, Steiner, H, Lachowicz, JE, Park, BH, Gauda, EB, Lee, EJ, Cool, MH, Sibley, DR, Gerfen, CR, Westphal, H & Fuchs, S (1996) A targeted mutation of the D3 dopamine receptor gene is associated with hyperactivity in mice. Proceedings of the National Academy of Science USA, 93, 1945-1949

Alarcón, M, Cantor, RM, Liu, J, Gillam, TC & Geschwind, DH (2002) Evidence for a quantitative trait locus on chromosome 7q in multiplex autism families. American Journal of Human Genetics, 70, 60-71

Alcock, KJ, Passingham, RE, Watkins, KE & Vargha-Khadem F (2000a) Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. Brain and Language, 75, 17-33

Alcock, KJ, Passingham, RE, Watkins, KE & Vargha-Khadem F (2000b) Pitch and timing abilities in inherited speech and language impairment. Brain and Language, 75, 34-46

Allison, DB (1997) Transmission-Disequilibrium tests for quantitative traits. American Journal of Human Genetics, 60, 676-690

Allison, DB, Neale, MC, Zannolli, R, Schork, NJ, Amos, CI & Blanjero, J (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. American Journal of Human Genetics, 65, 531-544

Almasy, L & Blanjero, J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. The American Journal of Human Genetics, 62, 1198-1211

Altmüller, J, Palmer, LJ, Fischer, G Scherb, H & Wjst, M (2001) Genomewide scans of complex human disease: true linkage is hard to find. American Journal of Human Genetics, 69, 936-950

American Psychiatric Association (1994) diagnostic and statistical manual of mental disorders (DSM-IV). Washington DC

Amos, CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. American Journal of Human Genetics, 54, 535-543

Aram, D & Nation, J (1975) Patterns of language behaviour in children with developmental language disorders. Journal of Speech and Hearing Research, 18, 229-241

Ardlie, KG, Kruglyak, L & Seielstad, M (2002) Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics, 3, 299-309

Ashley-Koch, A, Wolpert, CM, Menold, MM, Zaeem, L, Basu, S, Donnelly, SL, Ravan, SA, Powell, CM, Qumsiyeh, MB, Aylsworth, AS, Vance, JM, Gibert, JR, Wright, HH, Abramson, RK, DeLong, GR, Cuccaro, ML & Pericak-Vance, MA (1999) Genetic studies of autistic disorder and chromosome 7. Genomics, 61, 227-236

Auranen, M, Nieminen, T, Majuri, S, Vanhala, R, Peltonen, L & Järvelä (2000) Analysis of autism susceptibility gene loci on chromosomes 1p, 4p, 6q, 7q, 13q, 15q, 16p, 17q, 19q and 22q in Finish multiplex families. Molecular Psychiatry, 5, 320-322

Baddeley, AD & Hitch, G (1974) Working memory. In GA Bower (Ed) The psychology of learning and motivation (pp47-89) New York: Academic Press

Baddeley, AD (1992) Working memory. Science, 255, 556-559

Baddeley A & Wilson BA (1993) A developmental deficit in short-term phonological memory: implications for language and reading. Memory 1:65-78

Baddeley, AD, Gathercole, SE & Papagno, C (1998) The phonological loop as a language learning device. Psychological Review, 105, 158-173

Baddeley, AD (2000) The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4, 417-423

Badner, JA & Gershon, ES (2002) Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. Molecular Psychiatry, 7. 56-66

Baik, JH, Picetti, R, Saiardi, A, Thiriet, G, Dierich, A, Depaulis, A, Le Meur M & Borrelli, E (1995) Parkinsonian-like locomotor impairment in mice lacking dopamine D2 receptors. Nature, 377, 424-428

Barr, CL, Wigg, K, Malone, M, Schachar, R, Tannock, R, Roberts, W & Kennedy, JL (1999) Linkage study of catechol-O-methyltransferase and attention-deficit hyperactivity disorder. American Journal of Medical Genetics, 88, 710-713

Barr, CL, Wigg, KG, Wu, J, Zai, C, Bloom, S, Tannock, R, Roberts, W, Malone, M, Schachar, R & Kennedy, JL (2000a) Linkage study of two polymorphisms at the dopamine D3 receptor gene and attention-deficit hyperactivity disorder. American Journal of Medical Genetics, 96, 114-117

Barr, CL, Wigg, KG, Feng, Y, Zai, G, Malone, M, Roberts, W, Schachar, R, Tannock, R & Kennedy, JL (2000b) Attention-deficit hyperactivity disorder and the gene for the dopamine D5 receptor. Molecular Psychiatry, 5, 548-51

Barr, CL, Wigg, KG, Bloom, S, Schachar, R, Tannock, R, Roberts, W, Malone, M & Kennedy, JL (2000c) Further evidence from haplotype analysis for linkage of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. American Journal of Medical Genetics, 96, 262-267

Barr CL, Feng Y, Wigg K, Bloom S, Roberts W, Malone M, Schachar R, Tannock R, Kennedy JL. (2000d) Identification of DNA variants in the SNAP-25 gene and linkage study of these polymorphisms and attention-deficit hyperactivity disorder. Molecular Psychiatry, 5, 405-409

Barr, CL, Xu, C, Kroft, J, Feng, Y, Wigg, K, Zai, G, Tannock, R, Schachar, R, Malone, M, Roberts, W, Nothen, MM, Grunhage, F, Vandenbergh, DJ, Uhl, G, Sunohara, G, King, N & Kennedy, JL (2001a) Haplotype study of three polymorphisms at the dopamine transporter locus confirm linkage to attention-deficit/hyperactivity disorder. Biological Psychiatry, 49, 333-339

Barr, CL, Feng, Y, Wigg, KG, Schachar, R, Tannock, R, Roberts, W, Malone, M & Kennedy, JL (2001b) 5'-untranslated region of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. American Journal of Medical Genetics, 105, 84-90

Barr, CL, Kroft, J, Feng, Y, Wigg, K, Roberts, W, Malone, M, Ickowicz, A, Schachar, R, Tannock, R & Kennedy, JL (2002) The norepinephrine transporter gene and attention deficit hyperactivity disorder. American Journal of Medical Genetics, 114, 255-259

Bartlett, CW, Flax, JF, Logue, MW, Vieland, VJ, Bassett, AS, Tallal, P & Brzustowicz (2002) A major susceptibility locus for Specific Language Impairment is located on 13q21. American Journal of Human Genetics, 71, 45-55

Beitchman, JH, Nair, R, Clegg, M & Patel, PG (1986) Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. Journal of Speech and Hearing Disorders, 51, 98-110

Beitchman, JH, Hood, J & Inglis, A (1990) Psychiatric risk in children with speech and language disorders. Journal of Abnormal Child Psychology, 18, 283-296

Beitchman, JH, Hood, J & Inglis, A (1992) Familial transmission of speech and language impairment: A preliminary investigation. Canadian Journal of Psychiatry, 37, 151-156

Beitchman, JH, Brownlie, EB, Inglis, A, Wild, J, Matthews, R, Schachter, D, Kroll, R, Martin, S, Ferguson, B & Lancee, W (1994) Seven-year follow-up of speech/language-impaired and control children: Speech/language stability and outcome. Journal of the American Academy of Child Adolescent Psychiatry, 33, 1322-1330

Beitchman, JH, Wilson, B, Brownlie, EB, Walters, H & Lancee, W (1996a) Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. Journal of the American Academy of Child Adolescent Psychiatry, 35, 804-814

Beitchman, JH, Brownlie, GB & Wilson, B (1996b) Linguistic impairment and psychiatric disorder: Pathways to outcome in JH Beitchman, NJ Cohen, MM Konstantareas & R Tannock (Eds) Language, Learning and behaviour disorders: Developmental, biological and clinical perspectives (pp493-514). New York: Cambridge University Press

Beitchman, JH, Wilson, B, Brownlie, EB, Walters, H, Inglis, A & Lancee, W (1996c) Long-term consistency in speech/language profiles: II. Behavioural, emotional and social outcomes. Journal of the American Academy of Child Adolescent Psychiatry, 35, 815-825

Beitchman, JH, Brownlie, EB, Inglis, A, Wild, J, Ferguson, B, Schachter, D, Lancee, W, Wilson, B & Matthews, R (1996d) Seven-year follow-up of speech/language impaired and control children: Psychiatric outcome. Journal of Child Psychology and Psychiatry, 37, 961-970

Beitchman, JH, Wilson, B, Johnson, CJ, Atkinson, L, Young, A, Adlaf, E, Escobar, M & Douglas, L (2001a) Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. Journal of the American Academy of Child Adolescent Psychiatry, 40, 75-82

Beitchman, JH, Adlaf, EM, Douglas, L, Atkinson, L, Young, A, Johnson, CJ, Escobar, M & Wilson, B (2001b) Comorbidity of psychiatric and substance abuse disorders in late adolescence: A cluster analytic approach. American Journal of Drug and Alcohol Abuse, 27, 421-440

Bellaire, S, Plante, E & Swisher, L (1994) Bound morpheme skills in the oral language of school-age language-impaired children. Journal of Communication Disorders, 27, 265-279

Benton, A (1964) Developmental aphasia and brain damage. Cortex, 1, 40-52

Bisgaard, ML, Eiberg, H, Moller, N, Niebuhr, E & Mohr, J (1987) Dyslexia and chromosome 15 heteromorphism: negative lod score in a Danish material. Clinical Genetics, 32, 118-119

Bishop, DVM & Edmundson, A (1986) Is otitis media a major cause of specific developmental language disorders? British Journal of Disorders of Communication, 21, 321-338

Bishop, DVM & Edmundson, A (1987) Specific language impairment as a maturational lag: Evidence from longitudinal data in language and motor development. Developmental Medicine and Child Neurology, 29, 442-459

Bishop, DVM & Rosenbloom, L (1987) Classification of childhood language disorders. In W. Yule & M. Rutter (Eds.), Language development and disorders (pp. 16-41). London : MacKeith Press

Bishop, DVM (1994a) Is specific language impairment a valid diagnostic category? Genetic and psycholinguistic evidence. Philosophical Transactions of the Royal Society, B, 346, 105-111

Bishop, DVM (1994b) Grammatical errors in specific language impairment: Competence or performance limitation? Applied Psycholinguistics, 15, 507-549

Bishop, DVM, North, T & Donlan, C (1995) Genetic basis of specific language impairment; Evidence from a twin study. Developmental Medicine and Child Neurology, 37, 56-71

Bishop, DVM, North, T & Donlan, C (1996) Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. Journal of Child Psychology and Psychiatry, 37, 391-403

Bishop, DVM (1997a) Uncommon Understanding. Psychology Press Limited

Bishop, DVM (1997b) Pre- and perinatal hazards and family background in children with specific language impairments: A study of twins. Brain and Language, 56, 1-26

Bishop, DVM & Bishop, SJ (1998) Twin language: A risk factor for language impairment? Journal of Speech, Language and Hearing Research, 41, 150-160

Bishop, DVM, Bishop, SJ, Bright, P, James, C, Delaney, T & Tallal, P (1999a) Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. Journal of Speech, Language and Hearing Research, 42, 155-168

Bishop, DVM, Carlyon, RP, Deeks, JM & Bishop, SJ (1999b) Auditory temporal processing: Neither necessary nor sufficient for causing language impairment in children. Journal of Speech, Hearing and Language Research, 42, 1295-1310

Bishop, DVM (2001a) Genetic and environmental risks for specific language impairment in children. Philosophical Transactions of the Royal Society of London Series B, 356, 369-380

Bishop, DVM (2001b) Genetic influences on language impairment and literacy problems in children: Same or different? Journal of Child Psychology and Psychiatry, 42, 189-198

Bishop, DVM (2002) Motor immaturity and Specific Speech and Language Impairment: Evidence for a common genetic basis. American Journal of Medical Genetics, 114, 56-63

Bohannon, JN & Stanowicz, L (1988) The issue of negative evidence: Adult responses to children's language errors. Developmental psychology, 24, 684-689

Bonora, E, Bacchelli, E, Levy, ER, Blasi, F, Marlow, A, Monaco, AP, Maestrini, E & the International Molecular Genetic Study of Autism Consortium (2002) Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region. Molecular Psychiatry, 7, 289-301

Borg, I, Squire, M, Menzel, C, Stout, K, Morgan, D, Willatt, L, O'Brien, PCM, Ferguson-Smith, MA, Ropers, HH, Tommerup, N, Kalscheuer, VM & Sargan, DR (2002) A cryptic deletion of 2q35 including part of the PAX3 gene detected by breakpoint mapping in a child with autism and a de novo 2;8 translocation. Journal of Medical Genetics, 39, 391-399

Botstein, D, White, RL, Skolnick, MH & Davies, RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314

Bountress, N & Richards, J (1979) Speech, language and hearing disorders in an adult penal institution. Journal of Speech and Hearing Disorders, 44, 293-300

Bradford, Y, Haines, J, Hutcheson, H, Gardiner, M, Braun, T, Sheffield, V, Cassavant, T, Huang, W, Wang, K, Vielsnd, V, Folstein, S, Santangelo, S & Piven, J (2001) Incorporating language phenotypes strengthens evidence of linkage to autism. American Journal of Medical Genetics, 105, 539-547

Brownstein, MJ, Carpten, JD & Smith, JRH (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. Biotechniques, 20, 1008-1010

Bruce, HA & Margolis, RL (2002) *FOXP2*: Novel exons, splice variants, and CAG repeat length stability. Human Genetics, DOI10.1007/s00439-002-0768-5

Burden, V, Stott, CM, Forge, J & Goodyer, I (1996) The Cambridge Language and Speech Project (CLASP). I. Detection of language difficulties at 36 to 39 months. Developmental Medicine and Child Neurology, 38, 613-631

Butler, CC & MacMillan, H (2001) Does early detection of otitis media with effusion prevent delayed language development? Archives of Disease in Childhood, 85, 96-103

Buxbaum, JD, Silverman, JM, Smith, CJ, Kilifarski, M, Reichert, J, Hollander, E, Lawlor, BA, Fitzgerald, M, Greenberg, DA & Davis, KL (2001) Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. American Journal of Human Genetics, 68, 1514-1520

Cantwell, DP & Baker, L (1987) Prevalence and type of psychiatric disorder and developmental disorders in three speech and language groups. Journal of Communication Disorders, 20, 151-160

Cardon, LR & Fulker, DW (1994) The power of interval mapping of Quantitative Trait Loci, using selected sib pairs. The American Journal of Human Genetics, 55, 825-833

Cardon, LR, Smith, SD, Fulker, DW, Kimberling, WJ, Pennington, BF & DeFries, JC (1994) Quantitative Trait Locus for reading disability on chromosome 6. Science, 266, 276-279

Cardon, LR, Smith, SD, Fulker, DW, Kimberling, WJ, Pennington, BF & DeFries, JC (1995) Quantitative Trait Locus for reading disability: Correction. Science, 268, 1553

Castellanos, FX, Lau, E, Tayebi, N, Lee, P, Long, RE, Giedd, JN, Sharp, W, Marsh, WL, Walter, JM, Hamburger, SD, Ginns, EI, Rapoport, JL & Sidransky, E (1998) Lack of an association between a dopamine-4 receptor polymorphism and attention-deficit/hyperactivity disorder: genetic and brain morphometric analyses Molecular Psychiatry, 3, 431-434

Cervino, ACL & Hill, AVS (2000) Comparison of tests for association and linkage in incomplete families. American Journal of Human Genetics, 67, 120-132

Cheung, J, Petek, E, Nakabayashi, K, Tsui, LC, Vincent, JB & Scherer, SW (2001) Identification of the human cortactin-binding protein-2 gene from the autism candidate region at 7q31. Genomics, 78, 7-11

Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM, Pickles MR, Qin L, Fu Y, Mann JS, Kirschner BS, Jabs EW, Weber J, Hanauer SB, Bayless TM, Brant SR (1998) Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. Proceedings of the National Academy of Science USA, 95, 7502-7507

Chomsky, N (1965) Aspects of the theory of syntax. Cambridge, Mass: MIT Press

Chomsky, N (1981) Lectures on government and binding. Dordrecht: Foris

Chomsky, N (1986) Knowledge of language. New York: Praeger Special Studies

Clahsen, H (1989) The grammatical characterisation of developmental dysphasia. Linguistics, 27, 897-920

Clahsen, H, Rothweiler, M, Woest, A & Marcus, G (1992) Regular vs irregular inflection in the acquisition of German noun plurals. Cognition, 45, 225-255

Clark, PM & Dickman, Z (1984) Features of interaction in infant twins. Acta Geneticae Medicae et Gemellologiae: Twin Research, 33, 165-171

Clayton, D (1999) A generalisation of the transmission/disequilibrium test for uncertain haplotype transmission. American Journal of Human Genetics, 65, 950-961

Clerget-Darpoux, F, Bonaiti-Pellie, C & Hochez, J (1986) Effects of misspecifying genetic parameters in lod score analysis. Biometrics, 42, 393-399

Clifton-Bligh, RJ, Wentworth, JM, Heinz, P, Crisp, MS, John, R, Lazarus, JH, Ludgate, M & Chatterjee, VK (1998) Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nature Genetics, 19, 140-147 1998

Coën, R (1866) Pathologie und therapie der sprachanomalien. Vienna: Urban & Schwarzenberg

Cohen, NJ, Davine, M Horodezky, N, Lipsett, L & Isaacson, L (1993) Unsuspected language impairment in psychiatrically disturbed children: Prevalence and language and behavioural characteristics. Journal of the American Academy of Child and Adolescent Psychiatry, 32, 595-603

Cohen, NJ, Vallance, DD, Barwick, M, Im, N, Menna, R, Horodezky, NB & Isaacson, L (2000) The interface between ADHD and language impairment: An examination of language achievement and cognitive processing. Journal of Child Psychology and Psychiatry, 41, 353-362

Cohen, NJ (2001) Language impairment and psychopathology in infants, children, and adolescents. Sage publications: London

Cole, KN, Dale, PS & Mills, PE (1992) Stability of the intelligence quotient-language quotient relation: Is discrepancy modelling based on a myth? American Journal on Mental Retardation, 97, 131-143

Collaborative Linkage Study of Autism: Barrett, S, Beck, JC, Bernier, R, Bisson, E, Braun, TA, Casavant, TL, Childress, D, Folstein, S, Garcia, M, Gardiner, M, Gilman, S, Haines, JL, Hopkins, K, Landa, R, Meyer, NH, Mullane, JA, Nishimura, DY, Palmer, P, Piven, J, Purdy, J, Santangelo, SL, Searby, C, Sheffield, V, Singleton, J, Slager, S, Struchen, T, Svenson, S, Vieland, V, Wang, K & Winklosky, B (1999) An autosomal genomic screen for autism. American Journal of Medical Genetics, 88, 609-615

Collier, S & Gubb, D (1997) Drosophila tissue polarity requires the cell-autonomous activity of the fuzzy gene, which encodes a novel transmembrane protein. Development, 124, 4029-4037

Comings, DE, Gade, R, Muhleman, D & Sverd, J (1995) No association of a tyrosine hydroxylase gene tetranucleotide repeat polymorphism in autism, tourette syndrome, or ADHD. Biological Psychiatry, 37, 484-486

Conlon, RA, Reaume, AG & Rossant, J (1995) *Notch1* is required for the co-ordinate segmentation of somites. Development, 121, 1533-1545

Connell, PJ & Stone, CA (1992) Morpheme learning of children with specific language impairment under controlled instructional conditions. Journal of Speech and Hearing Research, 35, 844-852

Conrad, B, Dewald, G, Christensen, E, Lopez, M, Higgins, J & Pierpoint, ME (1995) Clinical phenotype associated with terminal 2q37 deletion. Clinical Genetics, 48, 134-139

Cook, EH Jr, Stein, MA, Krasowski, MD, Cox, NJ, Olkon, DM, Kieffer, JE & Leventhal, BL (1995) Association of attention-deficit disorder and the dopamine transporter gene. American Journal of Human Genetics, 56, 993-998

Cook, EH Jr, Courchesne, R, Lord, C, Cox, NJ, Yan, S, Lincoln, A, Haas, R, Courchesne, E & Leventhal BL (1997) Evidence of linkage between the serotonin transporter and autistic disorder. Molecular Psychiatry, 2, 247-250

Cook, EH, Courchesne, RY, Cox, NJ, Lord, C, Gonen, D, Guter, SJ, Lincoln, A, Nix, K, Haas, R, Leventhal, BL & Courchesne E (1998) Linkage-Disequilibrium Mapping of Autistic Disorder, with 15q11-13 Markers American Journal of Human Genetics, 62, 1077-1083

Crago, MB, Shanley, EM, Hough-Eyamie, A & Hough-Eyamie, WP (1997) Exploring innateness through cultural and linguistic variation (pp. 70-90) In M Gopnik (Ed) The inheritance and innateness of grammars. Oxford University Press: New York

Curran, S, Mill, J, Tahir, E, Kent, L, Richards, S, Gould, A, Huckett, L, Sharp, J, Batten, C, Fernando, S, Ozbay, F, Yazgan, Y, Siminoff, E, Thompson, M, Taylor, E & Asherson, P (2001) Association study of a dopamine transporter polymorphism and attention deficit hyperactivity disorder in UK and Turkish samples. Molecular Psychiatry, 6, 425-428

Dahoun-Hadorn S & ,Bretton-Chappuis, B (1992) de novo inversion-duplication of 2q35-2qter without growth retardation. Annales de Génétique, 35, 55-57

Dale, PS, Siminoff, E, Bishop, DVM, Eley, TC, Oliver, B, Price, TS, Purcell, S, Stevenson, J & Plomin, R (1998) Genetic influence on language delay in two year old children. Nature Neuroscience, 1, 324-328

Daly, G, Hawi, Z, Fitzgerald, M, & Gill M (1999) Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. Molecular Psychiatry, 4, 192-196

Daly, MJ, Rioux, JD, Schaffner, SF, Hudson, TJ & Lander, ES (2001) High-resolution haplotype structure in the human genome. Nature Genetics, 29, 229-232

DeCasper, AJ & Fifer, WP (1980) Of human voices: Newborns prefer their mothers' voices. Science, 208, 1174-1176

DeFries, JC & Fulker, DW (1988) Multiple regression analysis of twin data: Aetiology of deviant scores vs individual differences. Acta Geneticae Medicae et Gemellologiae (Roma), 37, 205-216

Demetras, MJ, Post, KN & Snow, CE (1986) Feedback to first language learners: The role of repetitions and clarification questions. Journal of Child Language, 13, 275-292

Dib, C, Faure, S, Fizames, C, Samson, D, Drouot, N, Vignal, A, Millasseau, P, Marc, S, Hazan, J, Seboun, E, Lathrop, M, Gyapay, G, Morisette, J & Weisssenbach, J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature, 380, 152-154

Douglas, JA, Boehnke, M & Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. American Journal of Human Genetics, 66, 1287-1297

Duerr, RH, Barmada, MM, Zhang, L, Pfützer, R & Weeks, DE (2000) High-density genome scan in Crohn Disease shows confirmed linkage to chromosome 14q11-12. American Journal of Human Genetics, 66, 1857-1862

Duffy, RJ (1975) The myth of the dichotomy of expressive and receptive aphasia. Conneticut Medicine, 39, 383-385

Eisenberg, J, Mei-Tal, G, Steinberg, A, Tartakovsky, E, Zohar, A, Gritsenko, I, Nemanov, L & Ebstein, RP (1999) Haplotype relative risk study of catechol-O-methyltransferase (COMT) and attention deficit hyperactivity disorder (ADHD): association of the high-enzyme activity Val allele with ADHD impulsive-hyperactive phenotype. American Journal of Medical Genetics, 88, 497-502

Eisenberg, J, Zohar, A, Mei-Tal, G, Steinberg, A, Tartakovsky, E, Gritsenko, I, Nemanov, L & Ebstein, RP (2000) A haplotype relative risk study of the dopamine D4 receptor (DRD4) exon III repeat polymorphism and attention deficit hyperactivity disorder (ADHD). American Journal of Medical Genetics, 96, 258-261

Elman, JL (1990) Finding structure in time. Cognitive Science, 14, 179-211

Elman, JL (1993) Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99

Elston, RC (1998) Methods of linkage analysis - and the assumptions underlying them. American Journal of Human Genetics, 63, 931-934

Ewing, A (1930) Aphasia in children. New York: Oxford University Press

Fabbro, F, Clarici, A & Bava, A (1996) Effects of left basal ganglia lesions on language production. Perceptual and Motor Skills, 82, 1291-1298

Fagerheim, T, Raeymaekers, P, Tønnessen, FE, Pedersen, M, Tranebjærg, L & Lubs, HA (1999) A new gene (DYX3) for dyslexia is located on chromosome 2. Journal of Medical Genetics, 36, 664-669

Faraone, SV, Biederman, J, Weiffenbach, B, Keith, T, Chu, MP, Weaver, A, Spencer, TJ, Wilens, TE, Frazier, J, Cleves, M & Sakai J (1999) Dopamine D4 gene 7-repeat allele and attention deficit hyperactivity disorder. American Journal of Psychiatry, 156, 768-770.

Faraone, SV, Doyle, AE, Mick, E & Biederman, J (2001) Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. American Journal of Psychiatry, 158, 1052-1057

Feinberg, AP & Vogelstein, B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analytical Biochemistry, 132, 6-13

Feldman, HM, Dollaghan, CA, Campbell, TF, Colburn, DK, Kurs-Lasky, M, Janosky, JE & Paradise, JL (1999) Parent-reported language and communication skills at one and two years of age in relation to otitis media in the first two years of life. Pediatrics, 104, e52

Felsenfeld, S & Plomin, R (1997) Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. Journal of Speech and Hearing Research, 40, 778-791

Fernald, A (1991) Prosody and focus in speech to infants and adults. Annals of Child Development, 8, 43-80

Field, LL & Kaplan BJ (1998) Absence of linkage of phonological coding dyslexia to chromosome 6p23-p21.3 in a large family data set. American Journal of Human Genetics, 63, 1448-1456.

Fisher, SE, Vargha-Khadem, F, Watkins, KE, Monaco, AP, Pembrey, ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nature Genetics, 18, 168-170

Fisher, SE, Marlow, AJ, Lamb, J, Maestrini, E, Williams, DF, Richardson, AJ, Weeks, DE, Stein JF & Monaco, AP (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. American Journal of Human Genetics, 64, 146-156

Fisher, SE, Francks, C, Marlow, AJ, MacPhie, L, Newbury, DF, Cardon, LR, Ishikawa-Brush, Y, Richardson, AJ, Talcott, JB, Gayán, J, Olson, RK, Pennington, BF, Smith, SD, DeFries, JC, Stein, JF & Monaco, AP (2002a) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nature Genetics, 30, 86-91

Fisher, SE, Francks, C, McCracken, JT, McGough, JJ, Marlow, AJ, MacPhie, IL, Newbury, DF, Crawford, LR, Palmer, CGS, Woodward, JA, Del'Homme, M, Cantwell, DP, Nelson, SF, Monaco, AP & Smalley, SL (2002b) A genomewide scan for loci involved in attention-deficit/ hyperactivity disorder. American Journal of Human Genetics, 70, 1183-1196

Flejter, WL, Bennett-Baker, PE, Ghaziuddin, M, McDonald, M, Sheldon, S & Gorski, JL (1996) Cytogenetic and molecular analysis of inv dup (15) chromosomes observed in two patients with autistic disorder and mental retardation. American Journal of Medical Genetics, 61, 182-187

Folstein, SE & Mankoski, RE (2000) Chromosome 7q : Where autism meets language disorder? American Journal of Human Genetics, 67, 278-281

Fountain, JW, Wallace, MR, Bruce, MA, Seizinger, BR, Menon, AG, Gusella, JF, Michels, VV, Schmidt, MA, Dewald, GW & Collins, FS (1989) Physical mapping of a translocation breakpoint in neurofibromatosis. Science, 244, 1085-1087

Frengen, E, Weichenhan, D, Zhao, B, Osoegawa, K, van Gell, M & de Jong, PJ (1999) A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics, 58, 250-253

Fritz, B, Müller-Navia, J, Hillig, U, Köhler, M, Aslan, M & Rehder, H (1999) Trisomy 2q35-q37 due to insertion of 2q material into 17q25: Clinical, cytogentic and molecular cytogenetic characterisation. American Journal of Medical Genetics, 87, 297-301

Froster, U, Schulte-Körne, G, Hebebrand, J & Remschmidt, H (1993) Cosegregation of balanced translocation (1;2) with retarded speech development and dyslexia. Lancet, 342, 178-179

Fulker, DW, Cherny, SS, Sham, PC & Hewitt, JK (1999) Combined linkage and association analysis for quantitative traits. American Journal of Human Genetics, 64, 259-267

Galaburda, AM, Sherman, GF, Rosen, GD, Aboitiz, F & Geschwind, N (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. Annals of Neurology, 18, 222-233

Gali, F (1835) The function of the brain and each of its parts, 5, Organology. Boston: Marsh, Capen & Lyon

Gallagher, A, Frith, U & Snowling, MJ (2000) Precursors of literacy delay among children at genetic risk of dyslexia. Journal of Child Psychology and Psychiatry, 41, 203-13

Gathercole SE & Baddeley AD (1990) Phonological memory deficits in language disordered children: Is there a causal connection? Journal of Memory and Language, 29, 336-360

Gathercole, SE, Willis, C, Baddeley, AD & Emslie, H (1994) The children's test of nonword repetition: A test of phonological working memory. Memory, 2, 103-127

Gathercole, SE & Baddeley, AD (1996) The children's test of non-word repetition. The Psychological Corporation: UK

Gathercole SE (1999) Cognitive approaches to the development of short-term memory. Trends in Cognitive Sciences, 3, 410-419

Gathercole, SE, Service, E, Hitch, GJ, Adams, AM & Martin, AJ (1999) Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. Applied Cognitive Psychology, 13, 65-77

Gathercole, SE & Pickering, SJ (2000) Working memory deficits in children with low achievements in the national curriculum at 7 years of age. British Journal of Educationbal Psychology, 70, 177-194

Gayán, J, Smith, SD, Cherny, SS, Cardon, LR, Fulker, DW, Brower, AM, Olson, RK, Pennington, BF & DeFries, JC (1999) Quantitative trait locus for specific language and reading deficits on chromosome 6p. American Journal of Human Genetics, 64, 157-164

Gertner, B, Rice, M & Hadley, P (1994) The influence of communicative competence on peer preferences in a preschool classroom. Journal of Speech and Hearing Research, 37, 913-923

Gessell, A & Amatruda, C (1947) Developmental diagnosis (second edition). New York: Hoeber

Ghiaziuddin, M & Burmeister, M (1999) Deletion of chromosome 2q37 and autism,: A distinct subtype? Journal of Autism and Developmental Disorders, 29, 259-263

Gill, M, Daly, G, Heron, S, Hawi, Z & Fitzgerald, M (1997) Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism. Molecular Psychiatry, 2, 311-313

Gillam, RB, Cowan, N & Day, LS (1995) Sequential memory in children with and without language impairment. Journal of Speech and Hearing Research, 38, 393-402

Giros, B, Jaber, M, Jones, SR, Wightman, RM & Caron, MG (1996) Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature, 379, 606-612

Goad, H & Rebellati, C (1994) Pluralisation in familial language impairment. In J Matthews (Ed) Linguistic aspects of familial language impairment. Special issue of the McGill Working Papers in Linguistics, vol 10. Montréal, Canada: McGill University

Gopnik, M (1990) Feature blindness: A case study. Language acquisition, 1, 139-164

Gopnik, M & Crago, M (1991) Familial aggregation of a developmental language disorder. Cognition, 39, 1-50

Gopnik, M (1994) Impairments of tense in a familial language disorder. Journal of neurolinguistics, 8, 109-133

Gopnik, M (1997) Language deficits and genetic factors. Trends in Cognitive Sciences, 1, 5-9

Gordon, CT, State, RC, Nelson, JE, Hamburger, SD & Rapport, JL (1993) A double-blind comparison of clomipramine, desipramine, and placebo in the treatment of autistic disorder. Archives of General Psychiatry, 50, 441-447

Gorski, JL, Cox, BA, Kyine, M, Uhlmann, W & Glover, TW (1989) Terminal deletion of the long arm of chromosome 2 in a mildly dysmorphic hypotonic infant with karyotype 46, XY, del(2)(q37). American Journal of Medical Genetics, 32, 350-352

Grant, J, Karmiloff-Smith, A, Gathercole, SA, Paterson, S, Howlin, P, Davies, M & Udwin, O (1997) Phonological short-term memory and its relationship to language in Williams syndrome. Cognitive Neuropsychiatry, 2, 81-99

Grigorenko, EL, Wood, FB, Meyer, MS, Hart, LA, Speed, WC, Shuster, A & Pauls, DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. American Journal of Human Genetics, 60, 27-39

Habib, M (2000) The neurological basis of developmental dyslexia: an overview and working hypothesis. Brain, 123, 2379-2399

Hahm, K, Cobb, BS, McCarty, AS, Brown, KE, Klug, CA, Lee, R, Akashi, K, Weissman, IL, Fisher, AG & Smale, ST (1998) Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes and Development, 12, 782-796

Hampe, J, Schreiber, S, Shaw, SH, Lau, KF, Bridger, S, Macpherson, AJ, Cardon, LR, Sakul, H, Harris, TJ, Buckler, A, Hall, J, Stokkers, P, van Deventer, SJ, Nurnberg, P, Mirza, MM, Lee, JC, Lennard-Jone,s JE, Mathew, CG, Curran, ME (1999) A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. American Journal of Human Genetics, 64, 808-816

Harley, TA (2001) The psychology of language: From data to theory. Psychology Press: New York

Haseman, JK & Eiston, RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behavior Genetics, 2, 3-19

Hawi, Z, McCarron, M, Kirley, A, Daly, G, Fitzgerald, M & Gill, M (2000a) No association of the dopamine DRD4 receptor (DRD4) gene polymorphism with attention deficit hyperactivity disorder (ADHD) in the Irish population. American Journal of Medical Genetics, 96, 268-272

Hawi, Z, Millar, N, Daly, G, Fitzgerald, M & Gill, M (2000b) No association between catechol-O-methyltransferase (COMT) gene polymorphism and attention deficit hyperactivity disorder (ADHD) in an Irish sample. American Journal of Medical Genetics, 96, 282-284

Hawi, Z, Foley, D, Kirley, A, McCarron, M, Fitzgerald, M & Gill, M (2001) Dopa decarboxylase gene polymorphisms and attention deficit hyperactivity disorder (ADHD): no evidence for association in the Irish population. Molecular Psychiatry, 6, 420-424

Hay, DA, Prior, M, Collett, S & Williams, M (1987) Speech and language development in preschool twins. Acta Geneticae Medicae et Gemellologiae, 36, 213-223

Haynes, C & Naidoo, S (1991) Children with specific speech and language impairment. London, England: MacKeith Press

Herault, J, Petit, E, Martineau, J, Cherpi, C, Perrot, A, Barthelemy, C, Lelord, G & Muh, JP (1996) Serotonin and autism: biochemical and molecular biology features. Psychiatry Research, 65, 33-43

Hill, EL (2001) Non-specific nature of Specific Language Impairment: A review of the literature with regard to concomitant motor impairments. International Journal of Language and Communication Disorders, 36, 149-171

Hirsh-Pasek, K, Treiman, R & Schneiderman, M (1984) Brown and Hanlon revisited: Mothers' sensitivity to ungrammatical forms. Journal of Child Language, 11, 81-88

Hodge, SE (1984) The information contained in multiple sibling pairs Genetic Epidemiology, 1, 109-122

Holm, VA & Kunze, LH (1969) Effect of chronic otitis media on language and speech development. Pediatrics, 43, 833-839

Holmes, J, Payton, A, Barrett, JH, Hever, T, Fitzpatrick, H, Trumper, AL, Harrington, R, McGuffin, P, Owen, M, Ollier, W, Worthington, J & Thapar, A (2000) A family-based and case-control association study of the dopamine D4 receptor gene and dopamine transporter gene in attention deficit hyperactivity disorder. Molecular Psychiatry, 5, 523-530

Holmes, J, Payton, A, Barrett, J, Harrington, R, McGuffin, P, Owen, M, Ollier, W, Worthington, J, Gill, M, Kirley, A, Hawi, Z, Fitzgerald, M, Asherson, P, Curran, S, Mill, J, Gould, A, Taylor, E, Kent, L, Craddock, N & Thapar, A (2002) Association of DRD4 in children with ADHD and comorbid conduct problems. American Journal of Medical Genetics, 114, 150-153

Hooper, JL & Matthews, JD (1982) Extensions to multivariate normal models for pedigree analysis. Annals of Human Genetics, 46, 373-383

Hosokawa, Y, Maeda, Y & Seto, M (1999) Human Helios, an Ikaros-related zinc finger DNA binding protein: cDNA cloning and tissue expression pattern. Immunogenetics, 50, 106-108

Hugdahl, K, Synnevag, B & Satz, P (1990) Immune and autoimmune diseases in dyslexic children Neuropsychologia, 28, 673-679

Hugot, JP, Laurent-Puig, P, Gower-Rousseau, C, Olson, JM, Lee, JC, Beaugerie, L, Noam, I, Dupas, JL, Van Gossum, A, Orholm, M, Bonaiti-Pellie, C, Weissenbach, J, Matthew, CG, Lennard-Jones, JE, Cortot, A, Colombel, JF & Thomas, G (1996) Mapping of a susceptibility locus for Crohn's Disease on chromosome 16. Nature, 29, 821-823

Hugot, JP, Chamaillard, M, Zouali, H, Lesage, S, Cézard, JP, Belaiche, J, Almer, S, Tysk, C, O'Morain, CA, Gassull, M, Binder, V, Finkel, Y, Cortot, A, Modigliani, R, Laurent-Puig, P, Gower-Rousseau, C, Macry, J, Colombel, JF, Sahbatou, M & Thomas, G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohns Disease. Nature, 411, 599-603

Huppert, SS, Le, A, Schroeter, EH, Mumm, JS, Saxena, MT, Milner, LA & Kopan, R (2000) Embryonic lethality in mice homozygous for a processing-deficient allele of Notch1. Nature, 415, 966-970

Hurst, JA, Baraitser, M, Auger, E, Graham, F & Norell, S (1990) An extended family with a dominantly inherited speech disorder. Developmental Medicine and Child Neurology, 32, 347-355

International Molecular Genetic Study of Autism Consortium (IMGSAC) (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. Human Molecular Genetics, 7, 571-578

International Molecular Genetic Study of Autism Consortium (IMGSAC) (2001a) A genomewide screen for autism: Strong evidence for linkage to chromosomes 2q, 7q and 16p. American Journal of Human Genetics, 69, 570-581

International Molecular Genetic Study of Autism Consortium (IMGSAC) (2001b) Further characterisation of the autism susceptibility locus AUTS1 on chromosome 7q. Human Molecular Genetics, 10, 973-982

Isaacs, EB & Vargha-Khadem, F (1989) Differential course of development of spatial and verbal memory span: A normative study. British Journal of Developmental Psychology, 7, 377-380

Jarrold, C & Baddeley, AD (1997) Short-term memory for verbal and visuospatial information in Down's Syndrome. Cognitive Neuropsychiatry, 2, 101-122

Jarrold, C, Baddeley, AD & Hewes, AK (1998) Verbal and nonverbal abilities in the Williams syndrome phenotype: Evidence for diverging developmental trajectories. Journal of Child Psychology and Psychiatry, 39, 511-523

Jarrold, C, Baddeley, AD, Hewes, AK & Phillips, C (2001) A longitudinal assessment of diverging verbal and nonverbal abilities in the Williams syndrome phenotype. Cortex, 37, 423-431

Jefferys, AJ, Kauppi, L & Neumann, R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatability complex. Nature Genetics, 29, 217-222

Jernigan, TL, Hesselink, JR, Sowell, E & Tallal, PA (1991a) Cerebral structure on magnetic resonance imaging in language- and learning-impaired children. Archives of Neurology, 48, 539-545

Jernigan, TL, Zisook, S, Heaton, RK, Moranville, JT, Hesselink, JR & Braff, DL (1991b) Magnetic resonance imaging abnormalities in lenticular nuclei and cerebral cortex in schizophrenia. Archives of General Psychiatry, 48, 881-890

Jernigan, TL, Archibald, SL, Fennema-Notestine, C, Gamst, AC, Stout, JC, Bonner, J & Hesselink, JR (2001) Effects of ageing on tissues and regions of the cerebrum and cerebellum. Neurobiology of Aging, 22, 581-594

Jeste, DV, McAdams, LA, Palmer, BW, Braff, D, Jernigan, TL, Paulsen, JS, Stout, JC, Symonds, LL, Bailey, A & Heaton, RK (1998) Relationship of neuropsychological and MRI measures to age of onset of schizophrenia. Acta psychiatrica Scandinavica, 98, 156-164

Jiang, S, Xin, R, Lin, S, Qian, Y, Tang, G, Wang, D & Wu, X (2001) Linkage studies between attention-deficit hyperactivity disorder and the monoamine oxidase genes. American Journal of Medical Genetics, 105, 783-788

Johnson, CJ, Beitchman, JH, Young, A, Escobar, M Atkinson, L, Wilson, B, Brownlie, EB, Douglas, L, Taback, N, Lam, I & Wang, M (1999) Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. Journal of Speech and Hearing Research, 42, 744-760

Johnston, RB, Stark, RE, Mellits, ED & Tallal, P (1981) Neurological status of language impaired and normal children. Annals of Neurology, 10, 159-163

Jolly, DJ, Okayama, H, Berg, P, Esty, AC, Filpula, D, Bohlen, P, Johnson, GG, Shively, JE, Hunkapillar, T& Friedmann, T (1983) Isolation and characterisation of a full-length expressible cDNA for human hypoxanthine phosphoribosyl transferase. Proceedings of the National Academy of Sciences USA, 80, 477-481

Kaestner, KH, Knöchel, W & Martinez, DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. Genes and Development, 14, 142-146

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

Kaplan, DE, Gayán, J, Ahn, J, Won, T-W, Pauls, D, Olson, RK, DeFries, JC, Wood, F, Pennington, BF, Page, GP, Smith, SD & Gruen, JR (2002) Evidence for linkage and association with reading disability on 6p21.3-22. American Journal of Human Genetics, 70, 1287-1298

Karmiloff-Smith, A, Grant, J, Berthould, I, Davies, M, Howling, P & Udwin, O (1997) Language and Williams syndrome: How intact is intact? Child Development, 68, 246-262

Kelley, CM, Ikeda, T, Koipally, J, Avitahl, N, Wu, L, Georgopoulos, K & Morgan, BA (1998) Helios, a novel dimerisation partner of Ikaros expressed in the earliest haematopoietic progenitors. Current Biology, 8, 508-515

Kent, L, Green, E, Holmes, J, Thapar, A, Gill, M, Hawi, Z, Fitzgerald, M, Asherson, P, Curran, S, Mills, J, Payton, A & Craddock, N (2001) No association between CHRNA7 microsatellite markers and attention-deficit hyperactivity disorder. American Journal of Medical Genetics, 105, 686-689

Klauck, SM, Poutstka, F, Benner, A, Lesch, KP & Poutska, A (1997) Serotonin transporter (5-HTT) gene variants associated with autism. Human Molecular Genetics, 6, 2233-2238

Koenig, M, Hoffman, EP, Bertelson, CJ, Monaco, AP, Feener, C & Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organisation of the DMD gene in normal and affected individuals. Cell, 50, 509-517

Kotler, M, Manor, I, Sever, Y, Eisenberg, J, Cohen, H, Ebstein, RP & Tyano,S (2000) Failure to replicate an excess of the long dopamine D4 exon III repeat polymorphism in ADHD in a family-based study. American Journal of Medical Genetics, 96, 278-281

Kovac, I, Gavabedian, B, DuSouich, C & Palmour, RM (2001) Attention deficit/hyperactivity in SLI children increases risk of speech/language disorders in first-degree relatives: a preliminary report. Journal of Communication Disorders, 34, 339-354

Kraus, N, McGee, TJ, Carrell, TD, Zecker, SG, Nicol, TG & Koch, DB (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science, 273, 971-973

Krishan, KR, McDonald, WM, Escalona, PR, Doraiswamy, PM, Na, C, Husain, MM, Figiel, GS, Boyoko, OB, Elinwood, EH & Nemeroff, CB (1992) Archives of General Psychiatry, 49, 553-557

Krugiyak, L & Lander, ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. American Journal of Human Genetics, 57, 439-454

Kruglyak, L, Daly, MJ, Reeve-Daly, MP & Lander ES (1996) Parametric and nonparametric linkage analysis : a unified multipoint approach. American Journal of Human Genetics, 58, 1347-1363

Lahey, M & Edwards, J (1995) Specific language impairment: Preliminary investigation of factors associated with family history and with patterns of language performance. Journal of Speech and Hearing Research, 38, 643-657

LaHoste, GJ, Swanson, JM, Wigal, SB, Glabe, C, Wigal, T, King, N & Kennedy, JL (1996) Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. Molecular Psychiatry, 1, 121-124

Lai, CSL, Fisher, SE, Hurst, JA, Levy, ER, Hodgson, S, Fox, M, Jeremiah, S, Povey, S, Jamison, DC, Green, ED, Vargha-Khadem, F & Monaco, AP (2000) The SPCH1 region on chromosome 7q31: Genomic characterisation of the critical interval and localisation of translocations associated with speech and language disorder. American Journal of Human Genetics, 67, 357-368

Lai, CSL, Fisher, SE, Hurst, JA, Vargha-Khadem, F & Monaco, AP (2001) A forkhead-domain gene is mutated in a sever speech and language disorder. Nature, 413, 519-523

Landau, WM & Kleffner, FR (1957) Syndrome of acquired aphasia with convulsive disorder in children. Neurology, 7. 523-530

Lander, ES & Schork, NJ (1994) Genetic dissection of complex traits. Science, 265, 2037-2048

Lander, ES & Krugylak, L (1995) Genetic dissection of complex traits: guidelines for interpretting and reporting linkage results. Nature Genetics, 11, 241-247

Lange, K, Westlake, J & Spence, MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. Annals of Human Genetics, 39, 485-491

Lassig, JP, Vachirasomtoon, K, Hartzell, K, Leventhal, M, Courchesne, E, Courchesne, R, Lord, C, Leventhal, BL & Cook, EH Jr (1999) 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. American Journal of Medical Genetics, 88, 472-475

Law, J, Boyle, J, Harris, F, Harkness, A & Nye, C (1998) Screening for speech and language delay: a systematic review of the literature. Health and Technology Assessment 2, 1-184

Laws, G (1998) The use of non-word repetition as a test of phonological memory in children with Down's syndrome. Journal of Child Psychology and Psychiatry, 39, 1119-1130

Lea, J (1968) Language and receptive aphasia. Special Education, 57, 21-25

Leonard, L (1972) What is deviant language? Journal of Speech and Hearing Disorders, 37, 427-446

Leonard, LB, Sabbadini, L, Leonard, JS & Volterra, V (1987) Specific Language Impairment in children: A cross linguistic study. Brain and Language, 32, 233-252

Leonard, LB (1989) Language learnability and specific language impairment in children. Applied Psycholinguistics, 10, 179-202

Leonard, LB, Bortolini, U, Caselli, MC, McGregor, KK & Sabbadini L (1992a) Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. Language Acquisition, 2, 151-179

Leonard, LB, McGregor, KK & Allen, GD (1992b) Grammatical morphology and speech perception in children with specific language impairment. Journal of Speech and Hearing Research, 35, 1076-1085

Leonard, LB (1998) Children with Specific Language Impairment. MIT Press: Massachusetts

Lesage, S, Zouali, H, Colombel, JF, EPIMAD, Belaiche, J, GETAID, Cézard, JP, EPGWGIBD, Tysk, C, Almer, S, Gassull, M, Binder, V, Chamaillard, M, LeGall, I. Thomas, G & Hugot, JP (2000) Genetic analyses of chromosome 12 loci in Crohn's Disease. Gut, 47, 787-791

Lewis, BA, Ekelman, BL & Aram, DM (1989) A familial study of severe phonological disorders. Journal of Speech and Hearing Research, 32, 713-724

Lewis, BA (1992) Pedigree analysis of children with phonology disorders. Journal of learning disabilities, 25, 586-597

Lewis, BA & Thompson, LA (1992) A study of developmental speech and language disorders in twins. Journal of Speech and Hearing Research, 35, 1086-1094

Lewontin, RC, Rose, S & Kamin, L (1984) Not in our genes. New York: Pantheon Books

Ley, J (1929) Un cas d'audi-multitéidiopathique (aphasie congénitale) chez des jumeaux monozygotiques L'Encéphale, 24, 121-165

Liebmann, A (1898) Vorlesungen über sprachstörungen, 3, Hörstummheit. Berlin: Coblentz

Lin, SP, Petty, EM, Gibson, LH, Inserra, J, Seashore, MR & Yang-Feng, TL (1992) Smallest terminal deletion of the long arm of chromosome 2 in a mildly affected boy. American Journal of Medical Genetics, 44, 500-502

Liu, J, Nyholt, DR, Magnussen, P, Parano, E, Pavone, P, Geschwind, D, Lord, C, Iversen, P, Hoh, J, the Autism Genetic Resource Exchange Consortium, Ott, J & Gillam, TC (2001) A genomewide screen for autism susceptibility loci. American Journal of Human Genetics, 69, 327-340

Lord C, Rutter M & LeCouteur A (1994) Autism Diagnostic Interview – Revised : a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders 24:659-685

Loveland KA, Landry SH (1986) Joint attention and language in autism and developmental language delay. Journal of Autism and Developmental Disorders, 16, 335-349

Lovell, K, Hoyle, H, & Sidall, M (1968) A study of some aspects of the play and language of young children with delayed speech. Journal of Child Psychology and Psychiatry, 9, 41-50

Lowe, AD & Campbell RA (1965) Temporal discrimination in aphasoid and normal children. Journal of Speech and Hearing Research, 8, 313-314

Lukusa, T, Devriendt, K, Jaeken, J & Fryns, JP (1999) Mild dysmorphic signs in two male sibs with partial trisomy 2q32.1-->q35 due to maternal ins(14;2) translocation. Clinical Dysmorphology, 8, 47-51

Lytton, H (1980) Parent-child interaction: The socialisation process observed in twin and singleton families. New York: Plenum Press

Ma, Y, Ohmen, JD, Li, Z, Bentley, LG, McElree, C, Pressman, S, Targan, SR, Fischel-Ghodsian, N, Rotter, JI & Yang, H (1999) A genome-wide search identifies potential new susceptibility loci for Crohn's Disease. Inflammatory Bowel Disease, 5, 271-278

Maestrini, E, Lai, C, Marlow, A, Matthews, N, Wallace, S, Bailey, A, Cook, EH, Weeks, DE, Monaco, AP & the International Molecular Genetic Study of Autism Consortium (1999) Serotonin transporter (5-HTT) and γ-aminobutyric acid receptor subunit β3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. American Journal of Medical Genetics, 88, 492-496

Manor, I, Eisenberg, J, Tyano, S, Sever, Y, Cohen, H, Ebstein, RP, Kotler, M (2001) Family-based association study of the serotonin transporter promoter region polymorphism (5-HTTLPR) in attention deficit hyperactivity disorder. American Journal of Medical Genetics, 105, 91-95

Marchuk, DA, Saulino, AM, Tavakkol, R, Swaroop, M, Wallace, M, Andersen, LB, Mitchell, AL, Gutmann, DH, Boguski, M & Collins, FS (1991) cDNA cloning of the type I neurofibromatosis gene: complete sequence of the NF1 gene product. Genomics, 11, 931-940

Margolis, RL, Abraham, MR, Gatchell, SB, Li, SH, Kidawi, AS, Breschel, TS, Stine, OC, Callahan, C, McInnis, MG & Ross, CA (1997) cDNAs with long CAG trinucleotide repeats from human brain. Human Genetics, 100, 114-122

Martin, ER, Gilbert, JR, Lai, EH, Riley, J, Rogala, AR, Slotterbeck, BD, Sipe, CA, Grubber, JM, Warren, LL, Conneally, PM, Saunders, AM, Schmechel, DE, Purvis, I, Pericak-Vance, MA, Roses, AD & Vance, JM (2000a) Analysis of association at single nucleotide polymorphisms in the APOE region. Genomics, 63, 7-12

Martin, ER, Lai, EH, Gilbert, JR, Rogala, AR, Afshari, AJ, Riley, J, Finch, KL, Stevens, JF, Livak, KJ, Slotterback, BD, Slifer, SH, Warren, LL, Conneally, PM, Schmechel, DE, Purvis, I, Pericak-Vance, MA, Roses, AD & Vance, JM (2000b) SNPing away at complex diseases: Analysis of single-nucleotide polymorphisms around *APOE* in Alzheimer disease. American Journal of Human Genetics, 67, 383-394

Mawhood L, Howlin P & Rutter M (2000) Autism and developmental receptive language disorder – a comparative follow-up in early adult life I : Cognitive and language outcomes. Journal of Child Psychology and Psychiatry 41:547-559

McArthur, D & Adamson, LB (1996) Joint attention in preverbal children: Autism and developmental language disorder. Journal of Autism and Developmental Disorders, 26, 481-496

McDonald, WM, Husain, M, Doraiswamy, PM, Figiel, G, Boyko, O & Krishnan, KR (1991) A magnetic resonance image study of age-related changes in human putamen nuclei. Neuroreport, 2, 57-60

McDougle, CJ, Naylor, ST, Cohen, DJ, Volkmar, FR, Heninger, GR & Price, LH (1996) A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder. Archives in General Psychiatry, 53, 1001-1008

Meaburn, E, Dale, PS, Craig, IW & Plomin, R (2002) Language-impaired children: No sign of the *FOXP2* mutation. Neuroreport, 13, 1075-1077

Menon, AG, Ledbetter, DH, Rich, DC, Seizinger, BR, Rouleau, GA, Michels, VF, Schmidt, MA, Dewald, G, DallaTorre, CM, Haines, JL & Gusella, JF (1989) Characterisation of a translocation within the von Recklinghausen neurofibromatosis region of chromosome 17. Genomics, 5, 245-249

Menyuk, P (1964) Comparison of grammar of children with functionally deviant and normal speech. Journal of Speech and Hearing Research, 7, 109-121

Merzenich, MM, Jenkins, WM, Johnston, P, Schreiner, C, Miller, SL & Tallal, P (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science, 271, 77-81

Michas, IC & Henry, LA (1994) The link between phonological memory and vocabulary acquisition. British Journal of Developmental Psychology, 12, 147-163

Mill, J, Curran, S, Kent, L, Gould, A, Huckett, L, Richards, S, Taylor, E & Asherson, P (2002) Association study of SNAP-25 microsatellite and attention deficit hyperactivity disorder. American Journal of Medical Genetics, 1114, 269-271

Montgomery, J (1995) Sentence comprehension in children with specific language impairment: The role of phonological working memory. Journal of Speech and Hearing Research, 38, 187-199

Muglia, P, Jain, U, Macciardi, F & Kennedy, JL (2000) Adult attention deficit hyperactivity disorder and the dopamine D4 receptor gene. American Journal of Medical Genetics, 96, 273-277

Muglia, P, Jain, U & Kennedy, JL. (2002) A transmission disequilibrium test of the Ser9/Gly dopamine D3 receptor gene polymorphism in adult attention-deficit hyperactivity disorder. Behavioural Brain Research, 130, 91-95

Mukhopadhyay, N, Almasy, L, Schroeder, M, Mulvihil, I WP & Weeks, DE (1999) Mega2, a data-handling program for facilitating genetic linkage and association analyses. American Journal of Human Genetics, 65, A436

Mullis, K, Faloona, F, Scharf, S, Saiki, R, Horn, G & Erlich, H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbour Symposia on Quantitative Biology, 51, 263-273

Murphy, D, DeCarli, C, Schapiro, M, Rapoport, S & Horwitz, B (1992) Age-related differences in volumes of subcortical nuclei, brain matter, and cerebrospinal fluid in healthy men as measured with magnetic resonance imaging. Archives of Neurology, 49, 839-845

Murray, L, Kempton, C, Woolgar, M & Hooper, R (1993) Depressed mothers speech to their infants and it's relation to infant gender and cognitive development. Journal of Child Psychology and Psychiatry, 34, 1083-1102

Murray, L, Hipwell, A, Hooper, R, Stein, A & Cooper, P (1996) The cognitive development of five-year old children of postnatally depressed mothers. Journal of Child Psychology and Psychiatry, 37, 927-935

Myers, WC & Mutch, PJ (1992) Language disorders in disruptive behaviour disordered homicidal youth. Journal of Forensic Sciences, 37, 919-922

Nakase, K, Ishimaru, F, Fujii, K, Tabayashi, T, Kozuka, T, Sezaki, N, Matsuo, Y & Harada, M (2002) Overexpression of novel short isoforms of Helios in a patient with T-cell acute lymphoblastic leukaemia. Experimental Hematology, 30, 313-317

Neale, M & Cardon, LR (1992) Methodology for genetic studies of twins and families. Fordrecht, The Netherlands: Kluwer Academic

Nells, J & Aram DM (1986) Family history of children with developmental language disorders. Perceptual and Motor Skills, 63, 655-658

Neville, HJ, Coffey, SA, Holocomb, PJ & Tallal, P (1993) The neurobiology of sensory and language processing in language-impaired children. Journal of Cognitive Neuroscience, 5, 235-253

Newbury, DF, Bonora, E, Lamb, JA, Fisher, SE, Lai, CSL, Baird, G, Jannoun, L, Slonims, V, Stott, CM, Merricks, MJ, Bolton, PF, Bailey, AJ, Monaco, AP & the International Molecular Genetic Study of Autism Consortium (2002) *FOXP2* is not a major susceptibility gene for autism or specific language impairment. American Journal of Human Genetics, 70, 1318-1327

Nishimura, DY Swiderski, RE, Alward, WL, Searby, CC, Patil, SR, Bennet, SR, Kanis, AB, Gastier, JM, Stone, EM & Sheffield, VC (1998) The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. Nature Genetics, 19, 140-147

Nyholt, DR (2000) All LODs are not created equal. American Journal of Human Genetics, 67, 282-288

O'Connell, P, Leach, RJ, Ledbetter, DH, Cawthorn, RM, Culver, M, Elridge, JR, Frej, AK, Holm, TR, Wolff, E, Thayer, MJ, Schafer, AJ, Fountain, JW, Wallace, MR, Collins, FS, Skolnick, MH, Rich, DC, Fournier, REK, Baty, BJ, Carey, JC, Leppert, MF, Lathrop, GM, Lalouel, JM & White, R (1989) Fine structure DNA mapping studies of the chromosomal region harbouring the genetic defect in neurofibromatosis type I. American Journal of Human Genetics, 44, 51-57

Ochs, E & Schieffelin, B (1995) The impact of language socialisation on grammatical development. (pp.73-94) In P Fletcher & B MacWhinney (Eds) Handbook of child language Blackwell: Oxford

Oefner, PJ and Underhill, PA (1995) Comparative DNA sequence by denaturing high performance liquid chromatography (DHPLC). American Journal of Human Genetics, 57, A266

Ogura, Y, Bonen, DK, Inohara, N, Nicolae, DL, Chen, FF, Ramos, R, Britton, H, Moran, T, Karalluskas, R, Duerr, RH, Achkar, JP, Brant, SR, Bayless, TM, Kirschner, BS, Hanauer, SB, Nuñez, G & Cho, JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's Disease. Nature, 411, 603-606

O'Reilly, R, Braver, TS & Cohen, JD (1999) A biologically based computational model of working memory. In A Miyake & P Shah (Eds) Models of working memory (pp375-411). Cambridge University Press

Orr, HT (2001) Beyond the Qs in the polyglutamine disease. Genes and Development, 15, 925-935

Osoegawa, K, Mammose, AG, Wu, C, Frengen, E, Zeng, C, Catanese, JJ & de Jong, PJ (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genomic Research, 11, 483-496

Palmer, CG, Bailey, JN, Ramsey, C, Cantwell, D, Sinsheimer, JS, Del'Homme, M, McGough, J, Woodward, JA, Asarnow, R, Asarnow, J, Nelson, S & Smalley, SL (1999) No evidence of linkage or linkage disequilibrium between DAT1 and attention deficit hyperactivity disorder in a large sample. Psychiatric Genetics, 9, 157-160

Paradis, M & Gopnik, M (1994) Compensatory strategies in genetic dysphasia: Declarative memory. Journal of neurolinguistics, 10, 173-185

Parker, JG & Asher SR (1987) Peer relations and later personal adjustment: Are low-accepted children at risk? Psychological Bulletin, 102, 357-389

Payton, A, Holmes, J, Barrett, Hever, T, Fitzpatrick, H, Trumper, AL, Harrington, R, McGuffin, P, O'Donovan, M, Owen, M, Ollier, W, Worthington, J & Thapar, A (2001) Examining for association between candidate gene polymorphisms in the dopamine pathway and attention deficit hyperactivity disorder: A family based study. American Journal of Medical Genetics, 105, 464-470

Pennington, BF, Smith, SD, Kimberling, WJ, Green, PA & Haith, MM (1987) Left-handedness and immune disorders in familial dyslexics. Archives of Neurology, 44, 634-639

Persico, AM, Militerni, R, Bravaccio, C, Schneider, C, Melmed, R, Trill, S, Montecchi, F, Palermo, M, Pascucci, T, Puglisi-Allegra, S, Reichelt, KL, Conciatori, M & Keller, F (2001) No association between the 4g/5G polymorphism of the plasminogen activator inhibitor-1 gene promoter and autistic disorder. Psychiatric Genetics, 11, 99-103

Petitto, LA (1997) In the beginning: On the genetic and environmental factors that make early language acquisition possible (pp. 45-69) In M Gopnik (Ed) The inheritance and innateness of grammars. Oxford University Press: New York

Philippe, A, Martinez, M, Guilloud-Bataille, M, Gillberg, C, Rastam, M, Sponheim, E, Coleman, M, Zappella, M, Aschauer, H, Van Maldergem, L, Penet, C, Feingold, J, Brice, A, Leboyer, M, & the Paris Autism Research International Sibpair Study (1999) Genome-wide scan for autism susceptibility genes. Human Molecular Genetics, 8, 805-812

Pickles, A, Bolton, P, Macdonald, H, Bailey, A, Le Couteur, A, Sim, CH & Rutter, M (1995) Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error : a twin and family history study of autism. American Journal of Human Genetics, 57, 717-726

Pinker, S (1994) The language instinct. Penguin Books

Pinter, JD, Eliez, S, Schmitt, JE, Capone, GT & Reiss, AL (2001) Neuroanatomy of Down's syndrome: A high resolution MRI study. American Journal of Psychiatry, 158, 1659-1665

Piven, J, Tsai, GC, Nehme, E, Coyle, JT, Chase, GA & Folstein, SE (1991) Platelet serotonin, a possible marker for familial autism. Journal of Autism and Developmental Disorders, 21, 51-59

Plante, E, Swisher, L, Vance, R & Rapcsak, S (1991) MRI findings in boys with specific language impairment. Brain and Language, 41, 52-66

Powell, R & Bishop, DVM (1992) Clumsiness and perceptual problems in children with specific language impairment. Developmental Medicine and Child Neurology, 34, 755-765

Pratt, SC, Daly, MJ & Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. American Journal of Human Genetics, 66, 1153-1157

Prober, JM, Trainor, GL, Dam, RJ, Hobbs, FW, Robertson, CW, Zagursky, RJ, Cocuzza, AJ, Jensen, MA & Baumeister, K (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science, 238, 336-341

Psarommatis, IM, Goritsa, E, Dimitrios, D, Tsakanikos, MJ, Kontrogianni, D & Apostolopoulos, N (2001) Hearing loss in speech-language delayed children. International Journal of Paediatric Otorhinolaryngology, 58, 205-210

Rabin, M, Wen, XL, Hepburn, M, Lubs, HA, Feldman, E & Duara, R (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. Lancet, 342, 178

Rabinowitz, D (1997) A transmission disequilibrium test for quantitative trait loci. Human Heredity, 47, 342-350

Randell, D, Reynell, J & Curwen, M (1974) A study of development in a sample of 3 year old children. British Journal of Disorders of Communication, 9, 3-16

Rapin I & Allen D (1983) Developmental language disorders : Nosologic considerations. In U. Kirk (Ed.) Neuropsychology of language, reading and spelling. New York : Academic Press

Rapin, I (1998) Understanding childhood language disorders. Current Opinion in Pediatrics, 10, 561-566

Raz, N, Torres, IJ & Acker, JD (1995) Age, gender, and hemispheric differences in human striatum: A quantitative review and new data from in vivo MRI morphometry. Neurobiology of Learning and Memory, 63, 133-142

Reed, PW, Davies, JL, Copeman, JB, Bennet, ST, Palmer, SM, Pritchard, LE, Gough, SC et al. (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nature Genetics, 7, 390-395

Retz, W, Thome, J, Blocher, D, Baader, M & Rosler, M (2002) Association of attention deficit hyperactivity disorder-related psychopathology and personality traits with the serotonin transporter promoter region polymorphism. Neuroscience Letters, 319, 133-136

Rice, ML, Sell, MA & Hadley, PA (1991) Social interactions of speech- and language-impaired children. Journal of Speech and Hearing Research, 34, 1299-1307

Rice, ML & Oetting, JB (1993) Morphological deficits of children with specific language impairment: Evaluation of number marking and agreement. Journal of Speech and Hearing Research, 36, 1249-1257

Rice, ML, Oetting, JB, Marquis, J, Bode, J & Pae, S (1994) Frequency of input effects on word comprehension of children with specific language impairment. Journal of Speech and Hearing Research, 37, 106-122

Rice, M, Wexler, K & Cleave, P (1995) Specific language impairment as a period of extended optional infinitive. Journal of Speech and Hearing Research, 38, 850-863

Rice, ML & Wexler, K (1996) Towards tense as a clinical marker of specific language impairment in English-speaking children. Journal of Speech and Hearing Research, 39, 1239-1257

Rice, ML, Haney, KJ & Wexler, K (1998) Family histories of children with SLI who show extended optional infinitives. Journal of Speech, Language and Hearing Research, 41, 419-432

Rineer, S, Finucane, B & Simon EW (1998) Autistic symptoms among children and young adults with isodicentric chromosome 15. American Journal of Medical Genetics, 81, 428-433

Rioux, JD, Silverberg, MS, Daly, MJ, Steinhart, AH, McLeod, RS, Griffiths, AM, Green, T, Brettin, TS, Stone, V, Bull, SB, Bitton, A, Williams, CN, Greenberg, GR, Cohen, Z, Lander, ES, Hudson, TJ & Siminovitch, KA (2000) Genomewide search in Canadian families with Inflammatory Bowel Disease reveals two novel susceptibility loci. American Journal of Human Genetics, 66, 1863-1870

Risch, N, Spiker, D, Lotspeich, L, Nouri, N, Hinds, D, Hallmayer, J, Kalaydjieva, L,McCague, P, Dimiceli, S, Pitts, T, Nguyen, L, Yang, J, Harper, C, Thorpe, D, Vermeer, S, Younhg, H, Herbert, J, Lin, A, Ferguson, J, Chiotti, C, Wiese-Slater, S, Rogers, T, Salmon, B, Nicholas, P, Petersen, PB, Pingree, C, McMahon, W, Wong, DL, Cavalli-Sforza, LL, Kraemer, HC & Myers, RM (1999) A genomic screen of autism : evidence for a multilocus etiology. American Journal of Human Genetics, 65, 493-507

Roberts, JE, Burchinal, MR, Davis, BP, Collier, AM & Henderson, FW (1991) Otitis media in early childhood and later language. Journal of Speech and Hearing Research, 34, 1158-1168

Roberts, JE, Burchinal, MR, Zeisel, SA, Neebe, EC, Hooper, SR, Roush, J, Bryant, D, Mundy, M & Henderson, FW (1998) Otitis media, the caregiving environment, and language and cognitive outcomes at 2 years. Pediatrics, 102, 346-354

Roberts, JE, Burchinal, MR, Jackson, SXC, Hooper, SR, Roush, J, Mundy, M, Neebe, EC & Zeisel, SA (2000) Otitis media in childhood in relation to preschool language and school readiness skills among black children. Pediatrics, 106, 725-735

Robin, DA & Scheinberg, S (1990) Subcortical lesions and aphasia. Journal of Speech and Hearing Disorders, 55, 90-100

Romain, DR, Mackenzie, NG, Moss, D, Columbano-Green, LM, Smythe, RH, Parfitt, RG & Dixon, JW (1994) Partial trisomy for 2q in a patient with dir dup(2) (q33.1q35). Journal of Medical Genetics 31, 652-653

Roman, T, Schmitz, M, Polanczyk, G, Eizirik, M, Rohde, LA & Hutz, MH (2001) Attention deficit hyperactivity disorder: A study of association with both the dopamine transporter gene and the dopamine D4 receptor gene. American Journal of Medical Genetics, 105, 471-478

Roman, T, Schmitz, M, Polancyzk, GV, Eizirik, M, Rohde, LA & Hutz, MH (2002) Further evidence for the association between attention deficit/hyperactivity disorder and the dopamine-β-hydroxylase gene. American Journal of Medical Genetics, 114, 154-158

Rubinsztein, DC, Leggo, J, Crow, TJ, DeLisi, LE, Walsh, C, Jain, S & Paykel, ES (1996) Analysis of polyglutamine-coding repeats in the TATA-binding protein in different human populations and in patients with schizophrenia and bipolar affective disorder. American Journal of Medical Genetics, 67, 495-498

Sachs, J, Bard, B & Johnson, ML (1981) Language with restricted input: Cases tudies of two hearing children of deaf parents. Applied Psycholinguistics, 2, 33-54

Salcini, AE, Confalonieri, S, Doria, M, Santolini, E, Tassi, E, Minenkova, O, Cesareni, G, Pelicci, PG & DiFiore, PP (1997) Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. Genes and Development, 11, 2239-2249

Sanchez, JM & Pantano, AM (1984) A case of deletion 2q35---qter and a peculiar phenotype. Journal of Medical Genetics, 21, 147-149

Sanger, F, Nickelsen, S & Coulson, AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA, 74, 5463-5467

Sato, S & Dreifuss, FE (1973) Electroencephalographic findings in a patient with developmental expressive aphasia. Neurology, 23, 181-185

Satsangi, J, Parkes, M, Louis, E, Hashimoto, L, Kato, N, Welsh, K, Terwilliger, JD, Lathrop, GM, Bell, Ji, Jewell, DP (1996) Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nature Genetics, 1996, 14, 199-202

Sawyer, DL, Krishnamani, RS, Hannig, VL, Garcia, M, Kim, JK, Haines, JL &Phillips, JA (1998) Genetic analysis of phonologic core deficit dyslexia (PCDD). American Journal of Human Genetics, 63, A307

Saxton, M (1997) The contrast theory of negative input. Journal of Child Language, 24, 139-161

Schald, DJ (1996) General score tests for associations of genetic markers with disease using cases and their parents. Genetic Epidemiology, 13, 423-450

Schick, JH, Kundtz, AM, Taylor, HG, Shriberg, LD, Freebairn, LA, Lewis, BA, Myengar, SK (2000) Evidence of linkage for speech impairment with chromosome 7 markers. American Journal of Human Genetics, 69, A537

Schiff-Myers, N (1988) Hearing children of deaf parents. In D Bishop & K Mogford (Eds) Language development in exceptional circumstances. Edinburgh: Churchill Livingstone

Schulte-Körne, G, Grimm, T, Nöthen, MM, Müller-Myhsok, B, Cichon, S, Vogt, IR, Propping, P & Remschmidt, H (1998) Evidence for linkage of spelling disability to chromosome 15. American Journal of Human Genetics, 63, 279-282

Schwartz, MF (1984) What the classical aphasia categories can't do for us and why. Brain and Language, 21, 3-8

Segman, RH, Meltzer, A, Gross-Tsur, V, Kosov, A, Frisch, A, Inbar, E, Darvasi, A, Levy, S, Goltser, T, Weizman, A & Galili-Weisstub, E (2002) Preferential transmission of interleukin-1 receptor antagonist alleles in attention deficit hyperactivity disorder. Molecular Psychiatry, 7, 72-74

Semel, EM, Wiig, EH & Secord, W (1992) Clinical Evaluation of Language Fundamentals - Revised. Psychological Corporation, San Antonio

Service, E & Kohonen, V (1995) Is the relation between phonological memory and foreign language accounted for by vocabulary acquisition? Applied Psycholinguistics, 16, 155-172

Shafer, VL, Schwartz, RG, Morr, ML, Kessler, KL, Kurtzberg, D & Ruben, RJ (2001) Neurophysiological indicies of language impairment in children. Acta Otolaryngologica, 121, 297-300

Sham, PC & Curtis, D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Annals of Human Genetics, 59, 323-336

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

Sherrington, R, Rogaev, EI, Liang, Y, Rogaeva, EA, Levesque, G, Ikeda, M, Chi, H, Lin, C, Li, G, Holman, K, Tsuda, T, Mar, L, Foncin, JF, Bruni, AC, Montesi, MP, Sorbi, S, Rainero, I, Pinessi, L, Nee, L, Chumakov, I, Pollen, D, Brookes, A, Sanseau, P, Polinsky, RJ, Wasco, W, Da Silva, HAR, Haines, JL, Pericak-Vance, MA, Tanzi, RE, Roses, AD, Fraser, PE, Rommens, JM & St George-Hyslop, PH (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature, 375, 754-760

Siegel, LS (1994) Working memory and reading: A life span perspective. International Journal of Behavioural Development, 17, 109-124

Silva, PA (1980) The prevalence, stability and significance of developmental language delay in preschool children. Developmental Medicine and Child Neurology, 22, 768-777

Silva, PA, McGee, R & Williams, SM (1983) Developmental language delay from three to seven years and it's significance for low intelligence and reading difficulties at age seven. Developmental Medicine and Child Neurology, 25, 783-793

Silva, PA, Williams, S & McGee, R (1987) A longitudinal study of children with developmental language delay at age three: Later intelligence, reading and behaviour problems. Developmental Medicine and Child Neurology, 29, 630-640

Skinner, BF (1957) Verbal Behaviour. New York: Appleton-Century-Crofts

Skuse, DH (1998) Extreme deprivation in early childhood. In D Bishop & K Mogford (Eds) Language development in exceptional circumstances. Edinburgh: Churchill Livingstone

Smalley, SL (1997) Genetic influences in childhood-onset psychiatric disorders: Autism and Attention-Deficit/Hyperactivity Disorder. American Journal of Human Genetics, 60, 1276-1282

Smalley, SL, Bailey, JN, Palmer, CG, Cantwell, DP, McGough, JJ, Del'Homme, MA, Asarnow, JR, Woodward, JA, Ramsey, C & Nelson, SF (1998) Evidence that the dopamine D4 receptor is a susceptibility gene in attention deficit hyperactivity disorder. Molecular Psychiatry, 3, 427-430

Smith, EE & Jonides, J (1998) Neuroimaging analyses of working memory. Proceedings of the National Academy of Science USA, 95, 12061-12068

Smith, SD, Kimberling, WJ, Pennington, BF & Lubs, HA (1983) Specific reading disability: Identification of an inherited form through linkage analysis. Science, 219, 1345-1347

Snowling, M, Bishop, DVM & Stothard, SE (2000) Is preschool language impairment a risk factor for dyslexia in adolescence? Journal of Child Psychology and Psychiatry,41, 587-600

Snowling, MJ Adams, JW, Bishop, DVM & Stothard, SE (2001) Educational attainments of school leavers with a preschool history of speech-language impairments. International Journal of Language and Communication Disorders, 36, 173-183

Spencer, T, Biederman, J, Wilens, T, Harding, M, O'Donnell, D & Griffin, S (1996) Pharmacotherapy of attention-deficit hyperactivity disorder across the life cycle. Journal of the American Academy of Child and Adolescent Psychiatry, 35, 409-432

Spielman, RS, McGinnis, RE & Ewens, WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American Journal of Human Genetics, 52, 506-516

Spielman, RS, McGinnis, RE & Ewens, WJ (1994) The transmission/disequilibrium test detects cosegregation and linkage. American Journal of Human Genetics, 54, 559-560

Spielman, RS & Ewens, WJ (1998) A sibship tests for linkage in the presence of association. The sib transmission/disequilibrium tests. American Journal of Human Genetics, 62, 450-458

Stafford, L (1987) Maternal input to twin and singleton children: implications for language acquisition. Human Communication Research, 13, 429-462

Stark, RE & Tallal, P (1981) Selection of children with specific language deficits. Journal of Speech and Hearing Disorders, 46, 114-122

Stattin, H & Klackenberg-Larsson, I (1993) Early language and intelligence development and their relationship to future criminal behaviour. Journal of Abnormal Psychology, 102, 369-378

Stefantos, GA, Green, GGR & Ratcliff, GG (1989) Neurophysiological evidence of auditory channel anomalies in developmental dysphasia. Archives of Neurology, 46, 871-875

Stevenson, J & Richman, N (1976) The prevalence of language delay in a population of three year old children and its association with general retardation. Developmental Medicine and Child Neurology, 18, 431-441

Stoel-Gammon, C (1998) Role of babbling and phonology in early linguistic development. In AM Wetherby, SF Warren & J Reichle (Eds.), Transitions in prelinguistic communication (pp 87-110). Baltimore, MD: Paul H Brookes

Stoneking, M (2001) From the evolutionary past... Nature, 409, 821-822

Stott, CM, Merricks, MJ, Bolton, PF and Goodyer, IM (2002) Screening for speech and language disorders: the reliability, validity and accuracy of the general language screen. International Journal of Communication Disorders, 37, 133-151

Stromswold, K (1998) Genetics of spoken language disorders. Human Biology, 70, 297-324

Sunohara, GA, Roberts, W, Malone, M, Schachar, RJ, Tannock, R, Basile, VS, Wigal, T, Wigal, SB, Schuck, S, Moriarty, J, Swanson, JM, Kennedy, JL & Barr, CL (2000) Linkage of the dopamine D4 receptor gene and attention-deficit/hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 39, 1537-1542

Sussman, JE (1993) Perception of formant transition cues to place of articulation in children with specific language impairments. Journal of Speech and Hearing Research, 36, 1286-1299

Swanson, JM, Sunohara, GA, Kennedy, JL, Regino, R, Fineberg, E, Wigal, T, Lerner, M, Williams, L, LaHoste, GJ & Wigal, S (1998) Association of the dopamine receptor D4 (DRD4) gene with a refined phenotype of attention deficit hyperactivity disorder (ADHD): a family-based approach. Molecular Psychiatry, 3, 38-41

Swanson, HL (1999) What develops in working memory? A life-span perspective. Developmental Psychology, 35, 986-1000

Tachibana, M, Takeda, K, Nobukuni, Y, Urabe, K, Long, JE, Meyers, KA, Aaronson, SA & Miki, T (1996) Ectopic expression of MITF, a gene for Waardenburg syndrome type 2, converts fibroblasts to cells with melanocytes characteristics. Nature Genetics, 14, 50-54

Tager-Flusberg, H (1999) A psychological approach to understanding the social and language impairments in autism. International Review of Psychiatry 11:325-334

Tager-Flusberg H, Joseph R & Folstein S (2001) Current directions in research on autism Mental Retardation and Developmental Disabilities Research Review 7:21-29

Tahir, E, Yazgan, Y, Cirakoglu, B, Ozbay, F, Waldman, I & Asherson, PJ (2000a) Association and linkage of DRD4 and DRD5 with attention deficit hyperactivity disorder (ADHD) in a sample of Turkish children. Molecular Psychiatry, 5, 396-404

Tahir, E, Curran, S, Yazgan, Y, Ozbay, F, Cirakoglu, B & Asherson, PJ. (2000b) No association between low- and high-activity catecholamine-methyl-transferase (COMT) and attention deficit hyperactivity disorder (ADHD) in a sample of Turkish children. American Journal of Medical Genetics, 96, 285-288

Tallal, P & Piercy, M (1973a) Defects of non-verbal auditory perception in children with developmental aphasia. Nature, 241, 468-469

Tallal, P & Piercy, M (1973b) Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. Neuropsychologia, 11, 389-398

Tallal, P & Piercy, M (1974) Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. Neuropsychologia, 12, 83-93

Tallal, P & Piercy, M (1975) Developmental aphasia: The perception of brief vowels and extended stop consonants. Neuropsychologia, 13, 69-74

Tallal, P (1976) Rapid auditory processing in normal and disordered language development. Journal of Speech and Hearing Research, 19, 561-571

Tallal, P, Stark, R & Curtiss, B (1976) Relation between speech perception and speech production in children with developmental aphasia. Brain and Language, 3, 305-317

Tallal, P & Stark, RE (1981) Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. Journal of the Acoustical Society of America, 69, 568-574

Tallal, P, Stark, RE, Kallman, C & Mellits, D (1981) A re-examination of some non-verbal perceptual abilities of language-impaired and normal children as a function of age and sensory modality. Journal of Speech and Hearing Research, 24, 351-357

Tallal, P, Ross, R & Curtiss, S (1989) Familial aggregation in specific language impairment. Journal of Speech and Hearing Disorders, 54, 167-173

Tallal, P, Townsend, J, Curtiss, S & Wulfeck, B (1991) Phenotypic profiles of language-impaired children based on genetic/family history. Brain and Language, 41, 81-95

Tallal, P, Jernigan, T, & Trauner, D (1994) Developmental bilateral damage to the head of the caudate nuclei: implications for speech-language pathology. Journal of Medical Speech-Language Pathology, 2, 23-28

Tallal, P, Miller, SL, Bedi, G, Byma, G, Wang, X, Najarajan, SS, Schreiner, C, Jenkins, WM & Merzenich, MM (1996) Language comprehension in language-learning impaired children improved with acoustically modified speech. Science, 271, 81-84

Tallal, P, Hirsch, LS, Realpe-Bonilla, Miller, S, Brzustowicz, LM, Bartlett, C & Flax, JF (2001) Familial aggregation of specific language impairment. Journal of Speech, Language and Hearing Research, 44, 1172-1182

The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable in Huntington's disease chromosomes. Cell, 72, 971-983

The IBD International Genetics Consortium (2001) International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn Disease and chromosome 16. American Journal of Human Genetics, 68, 1165-1171

The International SNP Map Working Group (2001) A map of the human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409, 928-933

Tirosh, E & Cohen, A (1998) Language deficit with attention deficit disorder: A prevalent comorbidity. Journal of Child Neurology, 13, 493-497

Todd, RD, Neuman, RJ, Lobos, EA, Jong, Y-JI, Reich, W & Heath, AC (2001) Lack of association of dopamine D4 receptor gene polymorphisms with ADHD subtypes in a population sample of twins. American Journal of Medical Genetics, 105, 432-438

Tomasello, M & Farrar, MJ (1986) Joint attention and early language. Child Development, 57, 1454-1463

Tomasello, M, Mannle, S & Kruger, AC (1986) Linguistic environment of 1-2-year-old twins. Developmental Psychology, 22, 169-176

Tomblin, JB (1989) Familial concentrations of developmental language impairment. Journal of Speech and Hearing Disorders, 54, 287-295

Tomblin, JB, Hardy, JC & Hein, HA (1991) Predicting poor-communication status in pre-school children using risk factors present at birth. Journal of Speech and Hearing Research, 34, 1096-1105

Tomblin, JB & Buckwalter, PR (1994) Studies of the genetics of specific language impairment. In R Watkins, & M Rice (Eds) Specific Language impairments in children. Baltimore, MD: Paul H Brookes

Tomblin, JB, Abbas, PJ, Records, NL & Brenneman, LM (1995) Auditory evoked responses to frequency-modulated tones in children with specific language impairment. Journal of Speech and Hearing Research, 38, 387-392

Tomblin, JB (1996) The big picture of SLI: Results of an epidemiologic study of SLI among kindergarten children. Paper presented at the Symposium on Research in Child Language Disorders, Madison, Wisconsin

Tomblin, JB, (1997) Epidemiology of specific language impairment. In M Gopnik (Ed) The inheritance and innateness of grammars. New York: Oxford University Press

Tomblin, JB, Records, N, Buckwalter, P, Zhang, X, Smith, E & O'Brien, M (1997a) Prevalence of specific language impairment in kindergarten children. Journal of Speech and Hearing Research, 40, 1245-1260

Tomblin, JB, Smith, E & Zhang, X (1997b) Epidemiology of specific language impairment: Prenatal and perinatal risk factors. Journal of Communication Disorders, 30, 325-344

Tomblin, JB & Buckwalter, PR (1998) Heritability of poor language achievement among twins. Journal of Speech and Hearing Research, 41, 188-189

Tomblin, JB, Nishimura, C, Zhang, X & Murray, JC (1998) Association of developmental language impairment with loci at 7q3. American Journal of Human Genetics Supplement, 63, A312

Trauner, D, Wulfeck, B, Tallal, P & Hesselink, J (2000) Neurological and MRI profiles of children with developmental language impairment. Developmental Medicine and Child Neurology, 42, 470-475

Treitel, L (1893) Über aphasie im kindesalter. Sammlung klinischer vorträge, 64, 629-654

Ullman, M & Gopnik, M (1994) The production of inflectional morphology in hereditary specific language impairment. In J Matthews (Ed) Linguistic aspects of familial language impairment. Special issue of the McGill Working Papers in Linguistics, vol 10. Montréal, Canada: McGill University

Väisse, L (1866) Des sourds-muets et de certains cas d'aphasie congénitale. Bulletin de la société d'Anthropologie de Paris, 1, 146-150

Van der Lely, HKJ & Harris, M (1990) Comprehension of reversible sentences in specifically language impaired children. Journal of Speech and Hearing Research, 55, 101-117

Van der Lely, HKJ & Howard, D (1993) Children with specific language impairment: Linguistic impairment or short-term memory deficit? Journal of Speech and Hearing Research, 36, 1193-1207

Van der Lely, HKJ (1994) Canonical linking rules forward versus reverse linking in normally developing and specifically language-impaired children. Cognition, 51, 29-72

Van der Lely, HKJ (1996) Specifically language impaired and normally developing children: Verbal passive versus adjectival passive sentence interpretation. Lingua, 98, 243-272

Van der Lely, H & Stollwerck, K (1996) A grammatical specific language impairment in children: An autsosomal dominant inheritance? Brain and Language, 52, 484-504

Van der Lely, HJK & Stollerwerck, L (1997) Binding theory and grammatical specific language impairment in children. Cognition, 245-290

Van Eerdewegh, P, Little, RD, Dupuis, J, Del Mastro, RG, Falls, K, Simon, J, Torrey, D, Pandit, S, McKenny, J, Braunschweiger, K, Walsh, A, Liu, Z, Hayward, B, Foiz, C, Manning, SP, Bawa, A, Saracino, L, Thackson, M, Benchekroun, Y, Capparell, N, Wang, M, Adair, H, Gibson, R, Aleen, KM, Pedan, A, Danzig, MR, Umland, SP, Egan, RW, Cuss, FM, Rorke, S, Clough, JB, Holloway, JW, Holgate, ST & Keith, TP (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature, 418, 426-430

Vargha-Khadem, F, Watkins, K, Alcock, K, Fletcher, P & Passingham, R (1995) Praxic and non-verbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proceedings of the National Academy of Sciences, 92, 930-933

Vargha-Khadem, F, Watkins, KE, Price, CJ, Ashburner, J, Alcock, KJ, Connelly, A, Frackowiak, RS, Friston, KJ, Pembrey, ME, Mishkihn, M, Gadian, DG & Passingham, RE (1998) Neural basis of an inherited speech and language disorder. Proceedings of the National Academy of Science. USA, 13, 12695-12700

Veenstra-VanderWeele, J, Gonen, D, Leventhal, BL & Cook, EH Jr (1999) Mutation screening of the UBE3A/E6-AP gene in autistic disorder. Molecular Psychiatry, 4, 64-67

Veenstra-VanderWeele, J, Kim, SJ, Lord, C, Courchesne, R, Akshoomoff, N, Leventhal, BL, Courchesne, E & Cook, EH Jr (2002) Transmission disequilibrium studies of the serotonin 5-HT (2A) receptor gene (HTR2A) in autism. American Journal of Medical Genetics, 114, 277-283

Vincent, JB, Herbrick, J, Gurling, HMD, Bolton, PF, Roberts, W & Scherer, SW (2000) Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. American Journal of Human Genetics, 67, 510-514

Viskochil, D, Buchberg, AM, Xu, G, Cawthorn, RM, Stevens, J, Wolff, RK, Culver, M, Carey, JC, Copeland, NG, Jenkins, NA, White, R & O'Connell, P (1990) Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell, 62, 187-192

Waldman, ID, Rowe, DC, Abramowitz, A, Kozel, ST, Mohr, JH, Sherman, SL, Cleveland, HH, Sanders, ML, Gard, JMC & Stever, C (1998) Association and Linkage of the Dopamine Transporter Gene and Attention-Deficit Hyperactivity Disorder in Children: Heterogeneity owing to Diagnostic Subtype and Severity. American Journal of Human Genetics, 63, 1767-1776

Wallace, MR, Marchuk, DA, Andersen, LB, Letcher, R, Odeh, HM, Saulino, AM, Fountain, JW, Brereton, A, Nicholson, J, Mitchell, AL, Brownstein, BH & Collins, FS (1990) Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in three NF1 patients. Science, 249, 181-186

Warburton, P, Baird, G, Chen, W, Morris, K, Jacobs, BW, Hodgson, S & Docherty, Z (2000) Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. American Journal of Medical Genetics, 96, 228-234

Warren, JD, Smith, HB, Denson, LA & Waddy, HM (2000) Expressive language disorder after infarction of left lentiform nucleus. Journal of Clinical Neuroscience, 7, 456-458

Wassink, TH, Piven, J, Vieland, VJ, Huang, J, Swiderski, RE, Pietila, J, Braun, T, Beck, G, Folstein, SE, Haines, JL & Sheffield, VC (2001) Evidence supporting *WNT2* as an autism susceptibility gene. American Journal of Medical Genetics, 105, 406-413

Wassink, TH, Piven, J, Vieland, VJ, Pietila, J, Goedken, RJ, Folstein, SE, Sheffield, VC (2002) Evaluation of FOXP2 as an autism susceptibility gene. American Journal of Medical Genetics, 114, 566-569

Watanabe, A, Takeda, K, Plopis, B & Tachibana, M (1998) Epistatic relationship between Waardenburg syndrome genes MITF and PAX3. Nature Genetics, 18, 283-286

Watkins, KE, Gadian, DG & Vargha-Khadem, F (1999) Functional and structural brain abnormalities associated with a genetic disorder of speech and language. American Journal of Human Genetics, 65, 1215-1221

Watkins, KE, Vargha-Khadem, F, Ashburner, J, Passingham, RE, Connelly, A, Friston, KJ, Fracowiak, RSJ, Mishkin, M & Gadian, DG (2002a) MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain, 125, 465-478

Watkins, KE, Dronkers, NF & Vargha-Khadem, F (2002b) Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. Brain, 125, 452-464

Watkins, R & Rice, M (1991) Verb particle and preposition acquisition in language impaired preschoolers. Journal of Speech and Hearing Research, 34, 1130-1141

Wechsler, D (1992) Wechsler Intelligence Scale for Children – Third UK Edition. Psychological Corporation, London

Weeks, DE & Lathrop, M (1995) Polygenic disease: methods for mapping complex disease traits. Trends in Genetics, 11, 513-519

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

Weistuch, L & Schiff-Myers, NB (1996) Chromosomal translocation in a child with SLI and apraxia. Journal of Speech and Hearing Research. 39, 668-671

Wenger, SL, Boone, LY, Surti, U & Steele, MW (1997) Terminal 2q deletion – a recognisable syndrome. Clinical Genetics, 51, 290

Wexler, K (1994) Optional infinitives. In D Lightfoot & N Hornstein (Eds) Verb movement (pp305-350) New York: Cambridge Press

Whitehurst, GJ, Arnold, Smith, M, Fischel, JE, Lonigan, CJ & Valdez-Menchaca, MC (1991) Family history in developmental expressive language delay. Journal of Speech and Hearing Research, 34, 1150-1157

Wiktor, A, Feldman, GL, Bawle, EV, Czarnecki, P, Conard, JV & VanDyke, DL (2001) Deletion of 2q37 and duplication of 10q24: Two cases in the same family and review of the literature. Annales de Génétique, 44, 129-134

Wilcox, ER, Rivolta, MN, Plopis, B, Potterf, SB & Fex, J (1992) The *PAX3* gene is mapped to human chromosome 2 together with a highly informative CA dinucleotide repeat. Human Molecular Genetics, 1, 215

Wilson, R (1975) The Louisville twin study: Developmental synchronies in behaviour. Child Development, 54, 298-316

Wolff, DJ, Clifton, K, Karr, C & Charles, J (2002) Pilot assessment of the subtelomeric regions of children with autism: Detection of a 2q deletion. Genetics in Medicine, 4, 10-14

Wolpert, CM, Menold, MM, Bass, MP, Qumsiyeh, MB, Donnelly, SL, Ravan, SA, Vance, JM, Gilbert, JR, Abramson, RK, Wright, HH, Cuccaro, ML & Pericak-Vance, MA (2000) Three probands with autistic disorder and isodicentric chromosome 15. American Journal of Medical Genetics, 96, 365-372

World Health Organisation (1993) The ICD-10 classification for mental and behavioural disorders: Diagnostic criteria for research. Geneva, Switzerland

Xu, G, O'Connell, P, Viskochil, D, Cawthon, R, Robertson, M, Culver, M, Dunn, D, Stevens, J, Gesteland, R, White, R & Weiss, R (1990) The neurofibromatosis type 1 gene encodes a protein related to GAP. Cell, 62, 599-608

Xu, M, Hu, XT, Cooper, DC, Moratalla, R, Graybiel, AM, White, FJ & Tonegawa, S (1994) Elimination of cocaine-induced hyperactivity and dopamine-mediated neurophysiological effects in dopamine D1 receptor mutant mice. Cell, 79, 945-

Young, RS, Shapiro, SD, Hansen, KL, Hine, LK, Rainosek, DE & Guerra, FA (1983) Deletion 2q: Two new cases with karyotypes 46,XY, del(2)(q31;q33) and 46, XX, del(2)(q36). Journal of Medical Genetics, 20, 199-202

Zhang, L, Cui, X, Schmitt, K, Hubert, R, Navidi, W & Arnheim N (1992) Whole genome amplification from a single cell: implications for genetic analysis Proceedings of the National Academy of Sciences USA, 89, 5847-5851

Zhong, N, Ye, L, Ju, W, Brown, WT, Tsiouris, J & Cohen I (1999) 5-HTTLPR variants not associated with autistic spectrum disorders. Neurogenetics, 2, 129-131

Zhong, W, Jiang, MM, Weinmaster, G, Jan, LY & Jan, YN (1997) Differential expression of mammalian Numb, Numblike and Notch1 suggests distinct roles during mouse cortical neurogenesis. Development, 124, 1887-1897

Zhong, W, Jiang, MM, Schonemann, MD, Menese, JJ, Pedersen, RA, Jan, LY & Jan, YN (2000) Mouse numb is an essential gene involved in cortical neurogenesis. Proceedings of the National Academy of Sciences, USA, 97, 6844-6849

Zilian, O, Saner, C, Hagedorn, L, Lee, HY, Säuberli, E, Suter, U, Sommer, LN & Auget, M (2001) Multiple roles of mouse Numb in tuning developmental cell fates. Current Biology, 11, 494-501

Zoghbi, HY & Orr, HT (2000) Glutamine repeats and neurodegeneration. Annual Review of Neuroscience, 23, 217-247

Group	Members	Address	Role
The Wellcome Trust Centre for Human Genetics	Dianne Newbury James Cleak Yumiko Ishikawa-Brush Angela Marlow Simon Fisher Anthony Monaco	The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Headington, Oxford. OX3 7BN. http://www.well.ox.ac.uk	Genome screen and subsequent genetic analyses
Newcomen Centre, Guys Hospital	Leila Jannoun Vicky Slonims Gillian Baird	Newcomen Centre, Guys Hospital, St Thomas' Street, London. SE1 9RT.	Collection of genome screen sample and psychometric testing
The Cambridge Language and Speech Project (CLASP)	Carol Stott Melanie Merricks Ian Goodyear Patrick Bolton	Developmental Psychiatry Section, Douglas House, 18B Trumpington Rd, Cambridge, CB2 2AH.	Collection of genome screen sample and psychometric testing
Human Communication & Deafness, University of Manchester	Zoe Simpkins Gina Conti-Ramsden	Human Communication and Deafness, School of Education, University of Manchester, Manchester. M13 9PL. http://www.hcd.man.ac.uk/	Collection of wave II sample and psychometric testing
University of Aberdeen	Elizabeth Hennessy Andrea Everitt Peter Helms Margaret Main Alan Kindley Duncan Shaw	Department of Child Health, University of Aberdeen, Foresterhill, Aberdeen. AB25 2ZD.	Collection of wave II sample and psychometric testing

Group	Members	Address	Role
University of Edinburgh	Ann Hodson Jocelyne Watson Anne O'Hare Jonathan Seckl	Child Life and Health, Department of Reproductive and Developmental Sciences University of Edinburgh, 20 Sylvan Place, Edinburgh. EH9 IUW.	Collection of wave II sample and psychometric testing
School of Epidemiology & Health Science, Manchester University	Andrew Pickles	School of Epidemiology and Health Science and Centre for Census and Survey Research, University of Manchester, Stopford Building, Oxford Road, Manchester. M13 9PT.	Statistical advice and multivariate analyses
Dept Experimental Psychology, University of Oxford	Dorothy Bishop	Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford. OX1 3UD. http://epwww.psych.ox.ac.uk/oscci/	Linguistic advice

For each sub-test a brief description of the task involved is given, followed by the trial example that the examiner goes through with the child. Finally two examples of the actual test items are given. Since the test items get harder throughout each sub-test I have typically listed the first (easiest) and last (hardest) item for each sub-test. Also given the aspect of language which each sub-test measures and the age range for which the test is applicable.

Sub-test	Aspect of language measured	Age range
Linguistic concepts	**Receptive language**	**Ages 5-7**
The child is presented with a series of different colour lines and must point to the lines when asked by the examiner. The examiner may ask them to complete an impossible task (e.g. point to the green line when there is no green line), in which case they must point to a stop sign.		
"I am going to show you some pictures with coloured lines on them. Sometimes I will point to the lines by myself. Sometimes I will ask you to point by yourself. And sometimes, we will point together. Listen carefully to my directions. First I will point to some lines. I will point to a yellow line *(point)*, a red one *(point)*, and a blue one *(point)*. Now you point to a blue line *(pause)*. Point to a red line *(pause)*. Now lets do it together. Let's point to a yellow line at the same time. Let's point to a red line at the same time *(point)*. If at any time you can't do what I ask you to do, point to the stop sign *(point)*. Now you are going to point to some lines. Listen carefully because I can't say it again"		
'point to the line that is not yellow', 'point to all the red lines and all but one of the yellow lines'		
Word structure	**Expressive language**	**Ages 5-7**
The children are shown pictures and the examiner describes the situation shown on each picture. The child must finish off the examiner's descriptions with the correct word or words. The examiner may repeat the sentence once.		
"I'm going to show you some pictures and say some things about them. I want you to help me by finishing some of the things I say. Let's try one. Here is a boy *(point to the picture of a boy)* and here is *(point to the picture of a girl and pause)*"		
'Here is a dog. Here are two...', 'Here is Butch painting a picture. This is the picture Butch...'		

Sub-test	Aspect of language measured	Age range
Sentence structure	**Receptive language**	**Ages 5-7**

The child is given an array of four pictures and must point to the one which depicts the situation the examiner is describing. The examiner is not allowed to repeat their descriptions
"Let's look at these pictures. I am going to point to the picture that shows the boy has a big dog *(point)*. Now you show me, The boy has the ball *(pause)*"
'The man who is carrying his umbrella is walking out the door', 'The girl asked "Where did you hide the present?"'

Orai directions	**Receptive language**	**Ages 5-17**

The child is presented with a series of different shapes and the examiner asks them to complete a series of actions involving these shapes. The child must not respond until each set of instructions is completed.
"I am going to show you some pictures. Each of the pictures has shapes on it. I will ask you to point to some of the shapes. Point to the square *(pause)*, point to the triangle *(pause)*. Now I will say 'Go' when I want you to point. Point to the small square and point to the white circle. Go *(pause)*"
'Point to the white circle; point to the black square' 'Point to the third square, the second triangle and the last circle'

Formulating sentences	**Expressive language**	**Ages 5-17**

The child is verbally presented with a list of words which they must incorporate into a sentence. They are also shown pictures which they may use to help them construct each sentence. The examiner may repeat the list of words once only.
"Here is a picture of a room. I will use the word books to talk about it in one sentence *(pause)*. 'There are lots of books in this room'. Or I could say 'They like to read books'. Now you make a sentence with the word shoes. You may use the picture to make your sentence. The picture may help you to think of what to say. If you would like to talk about something else, you don't have to use the picture. But you do have to say the word shoes"
'car', 'after and unless'

Sub-test	Aspect of language measured	Age range
Recailing sentences	**Expressive language**	**Ages 5-17**

The child is verbally presented with sentences which they must remember and recall. No visual aid is given and the examiner may repeat the sentences once only.
"Now I am going to say some things to you. I want you to listen and repeat what I say. If I say 'The boat sailed across the lake', you repeat it exactly as I say it. Let's try 'The boat sailed across the lake'"
'The dog chased the cat', 'The man in the house next door promised to water our flowers during our vacation'

Word classes	**Receptive language**	**Age 8+**

The child is verbally presented with a list of four words. From these four words they must chose the two which go together the best and repeat them to the examiner. No visual aid is given and no repetition is allowed.
"I am going to read some words to you. Two of the words go together. Listen to the words and tell me the two that go together the best *(pause)* girl, boy, car, table"
'tiger, lion, tree, baby', 'below, away, mile, distant'

Sentence assembly	**Expressive language**	**Age 8+**

The child is presented with a list of written words which the examiner reads. The child is then required to construct a sentence involving these words. The examiner may repeat the list of words once only.
"Here are some words that can be made into a sentence. I'll make a sentence with these words: tall, the boy, is *(pause)*. 'The boy is tall'. This sentence tells something. That's one way of doing it. Here's another sentence with the same words *(pause)* 'Is the boy tall?' This sentence asks something"
'saw, the dog, the woman', 'the lamp, the woman, the table, put, didn't, on'

Sub-test	Aspect of language measured	Age range
Semantic relationships	**Receptive language**	**Age 8+**

The child is given an incomplete sentence and a list of four words which may complete the sentence, only two of which make sense. The child must chose the two answers which can be used to correctly complete the sentence. The words are presented in written and verbal form and the examiner may repeat the list of words once only.

"I'm going to read you some problems to figure out. Each problem has two correct answers. Let's do one and see if you can tell me the two correct answers *(pause)*. A man is bigger than: a house, a penny, a spoon or a plane?"

'Water is wetter than...' 'sand, milk, ice, juice', 'Spring comes between...' 'autumn and summer, autumn and winter, summer and autumn, winter and summer'

Word associations — **Expressive language** — **Ages 5-17**

The child is given a semantic class and must recall as many members of that class as they can within 60 seconds

"I am going to tell you some things to wear: shoes, pants, hats, shirts and jackets. Can you think of other things to wear? *(pause)*. If the child gives a correct response repeat it, if the child gives an incorrect response say We don't wear that! But this is something else to wear...' Now tell me the names of some foods you eat"

'Tell me the names of as many animals as you can think of', 'Tell me as many names as you can think of for kinds of work people do'

Listening to paragraphs — **Receptive language** — **Ages 5-17**

The child is given a short story and then asked questions about the story afterwards. The examiner may repeat the questions once but may not repeat any part of the story

"Listen carefully to what I am going to read to you. I will ask you questions about what I read. 'For her birthday, Lisa's grandmother gave her a pretty gold ring. The ring had Lisa's name on it. The ring was old and a little scratched, but it still sparkled beautifully."

1. What did Lisa get for her birthday?
2. Who gave the ring to Lisa?
3. What was on the ring?

1) Dopelate
2) Glistering
3) Pennel
4) Defermication
5) Contramponist
6) Hampnent
7) Reutterpation
8) Perplisterork
9) Blonterstaping
10) Sepretennial
11) Detratapillic
12) Gistow
13) Frescovent
14) Bannifer
15) Stopograttic
16) Woolgalamic
17) Ballop
18) Confrantually
19) Fenneriser
20) Altupatory
21) Pristoractions
22) Underbrantuand
23) Trumpetine
24) Sladding
25) Commeecitate
26) Tafflest
27) Loddernapish
28) Barrazon
29) Commerine
30) Empliforvent
31) Thickery
32) Voltularity
33) Versatrationist
34) Rubid
35) Brasterer
36) Diller
37) Penneriful
38) Bannow
39) Prindle
40) Skiticul

1) Tirroge
2) Merhayba
3) Boodallower
4) Yarsteth
5) Malpirony
6) Griodarippal
7) Trogamusp
8) Axoniobic
9) Brufid
10) Pascotantic
11) Stryamect
12) Penneriful
13) Cheenlope
14) Dexiptecastic
15) Shimitet
16) Nembid
17) Empliforverit
18) Zubinken
19) Doduloppity
20) Strunfabe
21) Perplisteronk
22) Instadrontally
23) Usnat
24) Frescovent
25) Pranstitutiary
26) Tridercory
27) Donderificam
28) Brasterer

Please complete a separate form as fully as possible for each sample sent

PATIENT DETAILS *Delete as appropriate

Surname:.....

Forenames:.....

Identification No. : __/___/__

Sender's Name:..... Site: .....

Nature of sample: *BLOOD / MOUTH-SWAB

Phenotypic Sex: *MALE / FEMALE / AMBIGUOUS

Age (at sampling): ..... Date of Birth:.....

Race:
*CAUCASIAN / BLACK / ASIATIC INDIAN / ORIENTAL / OTHER (please specify)

.....

Clinical phenotype:
*ABNORMAL OR AFFECTED / NORMAL OR UNAFFECTED / NOT KNOWN

Diagnosis:.....

Is this the proband? *YES / NO

Other relevant clinical details:.....

Karyotype: *UNKNOWN / KNOWN.....

*BALANCED / UNBALANCED

Have any other samples from this family been sent to the WTCHG: *YES / NO

Please draw a detailed pedigree below:-

Signature:.....**Date:**.....

Marker	Panel	Reason	Comments
Chromosome 1			
D1S234	1	PCR Failed	
D1S425	2	PCR Failed	
Chromosome 2			
D2S367	3	PCR Failed	
D2S305	4	PCR Failed	
D2S347	4	Hard to genotype/bin	
D2S396	4	Shifting between alleles	
Chromosome 3			
D3S1271	5	Weak products	Typed for boxes 5b, 6 and 7
D3S1614	5	Weak products	Typed for boxes 5b, 6 and 7
D3S1267	7	Hard to genotype/bin	
Chromosome 4			
D4S1534	5	Hard to genotype/bin	
D4S405	5	PCR Failed	
D4S414	5	PCR Failed	
D4S391	6	Hard to genotype/bin	
Chromosome 5			
D5S410	10	Hard to genotype/bin	
D5S630	10	Hard to genotype/bin	
Chromosome 6			
D6S262	8	Shifting between alleles	
D6S257	9	Hard to genotype/bin	
D6S434	9	PCR Failed	
Chromosome 7			
D7S517	11	Re-designed	
D7S640	11	PCR Failed	
Chromosome 8			
D8S1771	11	PCR Failed	
D8S270	12	Not run	Typed for boxes 5b, 6 and 7
D8S284	12	Not run	
Chromosome 9			
D9S158	16	Weak products	
D9S1826	16	Shifting between alleles	
Chromosome 10			
D10S1868	14	Hard to genotype/bin	
D10S196	16	Shifting between alleles	
Chromosome 13			
D13S285	17	Hard to genotype/bin	
D13S159	19	Shifting between alleles	
Chromosome 14			
D14S468	20	Shifting between alleles	
Chromosome 15			
D15S127	21	Shifting between alleles	
Chromosome 16			
D16S3136	21	Hard to genotype/bin	
Chromosome 17			
D17S921	24	Shifting between alleles	
Chromosome 18			
D18S1102	23	Weak products	
D18S70	23	Shifting between alleles	

Marker	Panel	Reason	Comments
Chromosome 19			
D19S884	26	PCR Failed	Typed for boxes 6 and 7
D19S571	27	Hard to genotype/bin	
Chromosome 20			
D20S196	25	Shifting between alleles	
D20S178	27	Hard to genotype/bin	
Chromosome 21			
D21S1252	26	PCR Failed	
D21S1256	27	PCR Failed	
D21S263	27	Shifting between alleles	

Locus	Product Size	Dye	Origin[a]	Excluded Markers
Panel B1				
DXS8080	81-109	FAM	ABI HD-5	
D10S580	138-148	FAM	ABI HD-5	
D18S1127	184-210	FAM	ABI HD-5	
D12S1583	227-255	FAM	ABI HD-5	PCR Failed
D16S3100	274-288	FAM	ABI HD-5	
D12S304	314-328	FAM	ABI HD-5	
D16S3040	92-111	HEX	ABI HD-5	
D20S871	133-165	HEX	ABI HD-5	
D8S1779	197-211	HEX	ABI HD-5	
D11S4190	232-252	HEX	ABI HD-5	
D20S912	286-304	HEX	ABI HD-5	
D4S1586	107-125	NED	ABI HD-5	
D12S313	145-161	NED	ABI HD-5	
D6S1549	196-212	NED	ABI HD-5	Shifting between alleles
D21S1922	248-260	NED	ABI HD-5	
Panel B2				
D8S1720	141-155	FAM	ABI HD-5	
D20S102	169-177	FAM	Généthon	
D1S507	191-211	FAM	ABI HD-5	
D14S1044	230-244	FAM	ABI HD-5	
D2S2354	263-283	FAM	ABI HD-5	
D3S3706	106-122	HEX	ABI HD-5	
D21S1911	137-159	HEX	ABI HD-5	
D2S2150	172-204	HEX	ABI HD-5	
D12S1725	224-247	HEX	ABI HD-5	
D16S3034	263-271	HEX	ABI HD-5	
D3S3725	80-108	NED	ABI HD-5	
D4S1615	123-133	NED	ABI HD-5	
D8S275	147-165	NED	ABI HD-5	
D19S865	199-235	NED	ABI HD-5	PCR Failed
D16S3041	252-282	NED	ABI HD-5	Typed for boxes 6 and 7
D10S1655	312-328	NED	ABI HD-5	Typed for boxes 6 and 7
Panel B3				
D4S2994	99-131	FAM	ABI HD-5	
D10S1656	144-162	FAM	ABI HD-5	
D22S1170	201-215	FAM	ABI HD-5	Shifting between alleles
D10S1730	235-269	FAM	ABI HD-5	
D8S256	108-136	HEX	ABI HD-5	
D8S277	154-188	HEX	ABI HD-5	
D4S1587	226-238	HEX	ABI HD-5	
D14S1040	90-120	NED	ABI HD-5	
D19S886	148-164	NED	ABI HD-5	PCR Failed
D1S2766	190-202	NED	ABI HD-5	
D7S2496	215-241	NED	ABI HD-5	
D21S1255	312-322	NED	ABI HD-5	

a: ABI = Applied Biosystems HD-5 PRISM markers. http:www.appliedbiosystems.com
Généthon. http://www.genethon.fr/

Locus	Product Size	Dye	Origin[a]	Excluded Markers
Panel B4				
D11S912	109-131	FAM	ABI HD-5	
DXS1055	145-161	FAM	ABI HD-5	
D16S3102	193-211	FAM	ABI HD-5	
D9S1674	221-241	FAM	ABI HD-5	
D4S2924	263-279	FAM	ABI HD-5	
D3S3574	95-115	HEX	ABI HD-5	PCR Failed
D17S1795	139-149	HEX	ABI HD-5	
DXS984	162-192	HEX	ABI HD-5	
D4S2983	220-254	HEX	Généthon	
D1S423	94-100	NED	ABI HD-5	
D7S1870	116-140	NED	ABI HD-5	
D4S1579	150-168	NED	ABI HD-5	
D4S428	195-211	NED	ABI HD-5	
D16S3049	237-263	NED	ABI HD-5	
D17S1816	273-304	NED	ABI HD-5	
Panel B5				
D6S1693	85-93	FAM	Généthon	
D8S543	117-141	FAM	ABI HD-5	
D1S2850	153-161	FAM	ABI HD-5	
D16S418	174-196	FAM	ABI HD-5	
D19S904	216-230	FAM	ABI HD-5	
D2S2166	240-258	FAM	ABI HD-5	
D17S927	117-133	HEX	ABI HD-5	
D4S2971	143-163	HEX	ABI HD-5	
D18S1163	196-212	HEX	Généthon	
D16S3061	241-253	HEX	Généthon	
D17S1824	96-116	NED	ABI HD-5	
D20S861	128-140	NED	ABI HD-5	
D17S957	157-189	NED	ABI HD-5	
DXS8064	217-233	NED	ABI HD-5	
D3S3668	243-265	NED	ABI HD-5	
Panel B6				
D20S906	91-109	FAM	ABI HD-5	
D14S274	122-142	FAM	ABI HD-5	
D3S3609	171-193	FAM	ABI HD-5	
D12S1675	218-232	FAM	ABI HD-5	
D5S2073	245-261	FAM	ABI HD-5	
D2S2163	123-133	HEX	ABI HD-5	PCR Failed
D21S262	142-152	HEX	Généthon	Shifting between alleles
D10S1765	170-192	HEX	ABI HD-5	
D8S1762	228-246	HEX	ABI HD-5	
D16S3140	286-324	HEX	ABI HD-5	
D20S882	78-90	NED	ABI HD-5	
D4S3022	129-157	NED	ABI HD-5	
D3S1597	170-188	NED	ABI HD-5	
D2S149	218-236	NED	ABI HD-5	
D6S452	269-288	NED	ABI HD-5	

***a*:** ABI = Applied Biosystems HD-5 PRISM markers. http:www.appliedbiosystems.com
Généthon. http://www.genethon.fr/

Locus	Product Size	Dye	Origin[a]
Panel B7			
D5S1960	125-163	FAM	ABI HD-5
D4S2912	175-201	FAM	ABI HD-5
D19S903	214-236	FAM	ABI HD-5
D10S570	295-313	FAM	ABI HD-5
D8S549	78-88	HEX	ABI HD-5
D8S1836	129-163	HEX	ABI HD-5
D4S2936	174-192	HEX	ABI HD-5
D21S1896	207-219	HEX	Généthon
D10S575	255-277	HEX	ABI HD-5
DXS8069	139-153	NED	ABI HD-5
D10S1709	164-182	NED	ABI HD-5
D6S305	212-238	NED	ABI HD-5
D2S352	271-299	NED	ABI HD-5

Excluded Markers

***a*:** ABI = Applied Biosystems HD-5 PRISM markers. http:www.appliedbiosystems.com
Généthon. http://www.genethon.fr/

Locus	Product Size	Dye	Origin[a]	Excluded Markers/ Comments
Fine Mapping Panel 19.1				
D19S572	124-142	FAM	ABI	
D19S606	172-190	FAM	Généthon	
D19S208	167-175	HEX	Généthon	
D19S921	189-215	HEX	ABI	
D19S888	178-198	NED	ABI	
Fine Mapping Panel 19.2				
D19S566	147-169	FAM	ABI	
APOC2	297-353	FAM	Généthon	
D19S867	102-116	HEX	Généthon	
D19S900	157-177	HEX	Généthon	PCR Failed
D19S178	143-189	NED	GDB	
D19S882	269-283	NED	Généthon	
Fine Mapping Panel 19.3				
DM	128-194	FAM	GDB	Bad Inheritance
D19S879	251-265	FAM	Généthon	
BLC3	127-137	HEX	GDB	PCR Failed
D19S246	185-233	HEX	GDB	
D19S881	133-143	NED	Généthon	
KLK	176-202	NED	GDB	
***NUMBL* Panel**				
D19S412	89-113	FAM	Généthon	
NUMBL	141-150	FAM	Margolis et al (1997)	Trinucleotide repeat
D19S908	200-232	FAM	Généthon	
D19S425	252-280	FAM	Généthon	
D19S225	170-180	HEX	Généthon	
D19S596	213-221	HEX	Généthon	
D19S224	240-262	HEX	Généthon	
D19S402	301-359	HEX	Marshfield	PCR Failed
D19S553	408-479	HEX	Marshfield	Tetranucleotide repeat

***a*:** ABI = Applied Biosystems HD-5 PRISM markers. http:www.appliedbiosystems.com
Généthon. http://www.genethon.fr/
GDB = Genome Database. http://www.gdb.org
Marshfield = Marshfield Center for Medical Genetics
http:/research.marshfieldclinic.org/genetics

REAGENTS

Agarose Loading Buffer (2x)
For 100ml
5g ficoll
50ml MilliQ dH_2O
Boil gently until the ficoll has dissolved
Make up to 100ml with MilliQ dH_2O
Add bromophenol blue and xylene cyanol FF
Keep at room temperature

Blocking Solution / Antibody Dilution Solution
To make 1ml
0.5g non fat dried milk
10ml 1xPBS/Tween
Dissolve milk in PBS by shaking
Pipette 1.5ml into an eppendorf and centrifuge at 13000rpm for 20 minutes
Transfer top 1ml supernatant to a fresh eppendorf
Store at -20°c

Buffer A (0.1MTEAA)
For 1 litre
50ml TEAA
250µl acetonitrile
Make up to 1 litre with MilliQ dH_2O
Keep at room temperature for up to 1 week

Buffer B (0.1M TEAA/25% acetonitrile)
For 1 litre
50ml TEAA
250ml acetonitrile
Make up to 950ml with MilliQ dH_2O
Leave to warm to room temperature
Make up to 1 litre with MilliQ dH_2O
Keep at room temperature for up to 1 week

2.5x Dilution Buffer
For 1ml
200µl 1mM TrisHCl (pH7.0)
5µl 25mM $MgCl_2$
Make up to 1ml with MilliQ dH_2O
Keep at room temperature

0.5M EDTA (pH8.0)
For 250ml
46.52g $Na_2EDTA.2H_2O$
175ml MilliQ dH_2O
Add NaOH tablets until EDTA dissolves (stirring)
Adjust pH to 8.0 using NaOH and HCl
Make up to 250mls with MilliQ dH_2O
Autoclave
Keep at room temperature

Fixative
1 part acetic acid to 3 parts methanol
Make up fresh each time, store at –20c before use

Formamide Loading Buffer
1 part loading dye (ABI)
5 parts deionised formamide (Sigma)
Keep at 4°c

Freeze mix
10 parts FCS
1 part DMSO
Keep at -20°c

GTE
For 50ml
2.5ml Glucose (1M)
1.25ml Tris (1M)
1.00ml EDTA (0.5M)
Make up to 50ml with milliQ dH2O
Keep at room temperature

Hybridisation Buffer
50% formamide/10% dextran sulphate/2xSSC
Keep at 4°c

Luria Broth (LB)
For 400ml
8g LB broth base
400ml MilliQ dH_2O
Autoclave
Keep at room temperature

REAGENTS (cont.)

LB Agar
For 400ml (~ 9 plates)
8g LB broth base
6g bacto-agar
400ml MilliQ dH_2O
Autoclave
Keep at 60°c

Precipitation Mix
Equal volumes of:
10mg/ml tRNA
10mg/ml sonicated salmon sperm DNA
Store at -20°c

Reagent A
For 1 litre
109.5g sucrose
10ml Tris (1M pH8.0)
5ml $MgCl_2$ (1M)
Make up to 900ml with MilliQ dH_2O
Autoclave
Add 10ml Titron
Make up to 1 litre with MilliQ dH_2O
Keep at room temperature

1% SDS + 0.2M NaOH
For 50ml
2ml NaOH (5M)
43ml MilliQ dH_2O
5ml 10% SDS
To prevent precipitation of NaOH add reagents in order given
Do not keep – make up fresh each time

1xTE (pH8.0)
For 200ml
2.0ml TrisHCl (1M)
0.4ml EDTA (0.5M)
Make up to 200ml with MilliQ dH_2O
Keep at room temperature

Transport Buffer
For 1 litre
584g NaCl (solid)
10ml Tris (pH8.0)
50ml EDTA (pH8.0)
Make up to 900ml with MilliQ dH_2O
Autoclave
Add 125ml 20% SDS
Make up to 1 litre with MilliQ dH_2O
Keep at room temperature

Mr Tris = 157.6
Mr $MgCl_2$ = 203.3
Mr Ammonium acetate = 77.08
Mr NaOH = 40.00
Mr Sodium Acetate = 82.03
Mr Guanidine HCl = 95.53

SUPPLIERS

AB Gene
http://www.abgene/com
0.2ml, 0.5ml and 0.7ml microcentrifuge tubes

Applied Biosystems (ABI)
http://www.appliedbiosystems/com
Taq Gold
ROX size standard
PRISM LMS-MD10 primers
PRISM LMS-HD5 primers
BigDye terminator

Beckman
http://www.beckman.com
96 deep well plates

Children's Hospital Oakland Research Institute (CHORI)
http://www.chori.org BACs

Corning
http://www.corning.com
15ml and 50ml tubes
Vented tissue culture flasks
96 well Costar plates

European Collection of Cell Cultures (ECACC)
http://www.ecacc.org.uk
EBV transformation of cell lines

Gibco BRL http://www.lifetech.com
Cot^{-1} DNA

SUPPLiERS (cont.)

Millipore
http://www.millipore.com
dH_2O purification systems
96 well purification plates

MWG Biotech
http://www.mwg-biotech.com
All primers were supplied by MWG, unless otherwise stated

New England Biolabs (NEB)
http://www.neb.com
All restriction enzymes, buffers and BSA
All agarose markers

Nucleon Biosciences
http://www.tepnel.com
Blood extraction kits.
Nucleon resin also used for buccal swab extraction protocol

PE Biosystems
http://www.appliedbiosystems.com
10x buffer (KCl)
dNTPs (8mM)
$MgCl_2$ (25mM)

Sigma
http://www.sigmaaldrich.com
All reagents and chemicals were supplied by Sigma, unless otherwise noted

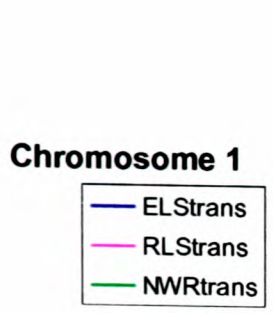
Stratagene
http://www.stratagene.com
Pfu polymerase

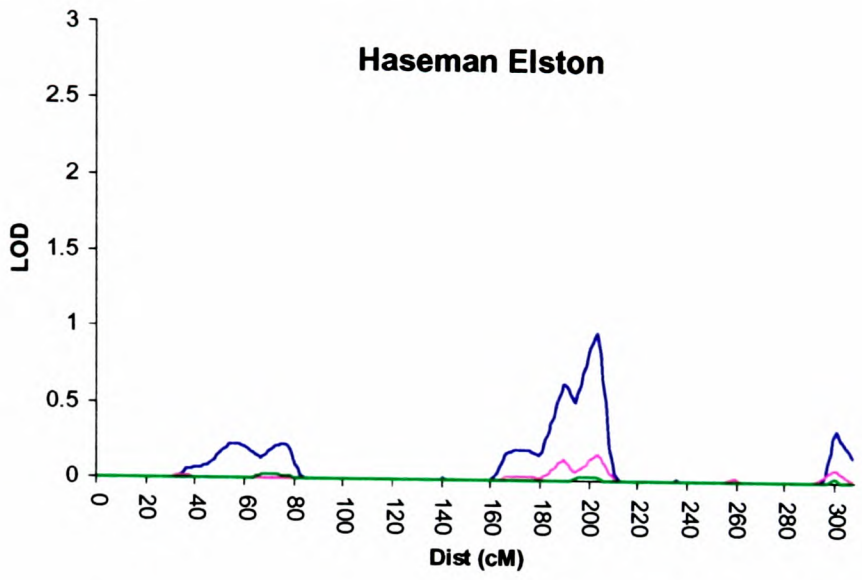
Transgenomic
http://www.transgenomic.com
WAVE reagents (TEAA and acetonitrile)

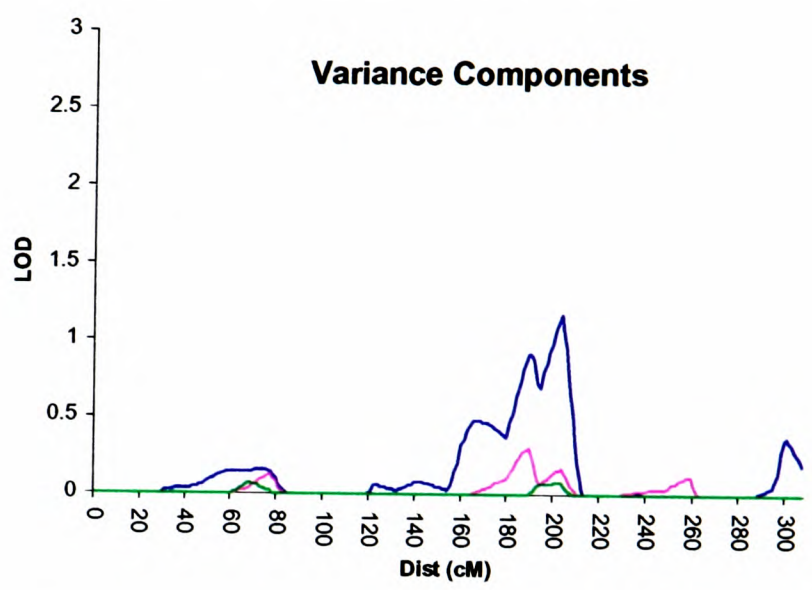
Vector Labs
http://www.vectorlabs.com
Antibodies
Propidium iodide
DAPI

Vysis
http://www.vysis.com
Nick translation kits (Biotin-dUTP, DIG-dUTP, dNTP mix, nick translation buffer and nick translation enzyme)

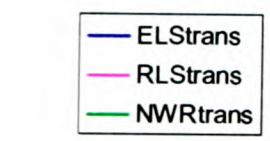
Chromosome 1

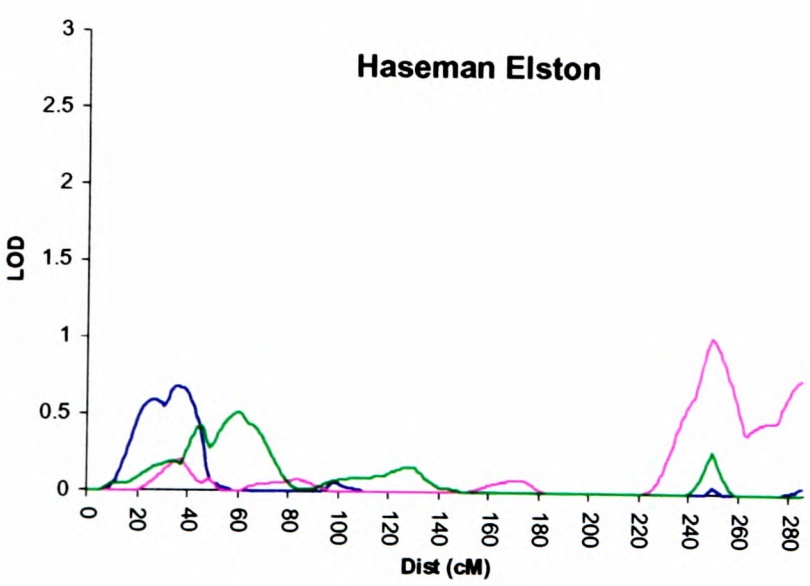


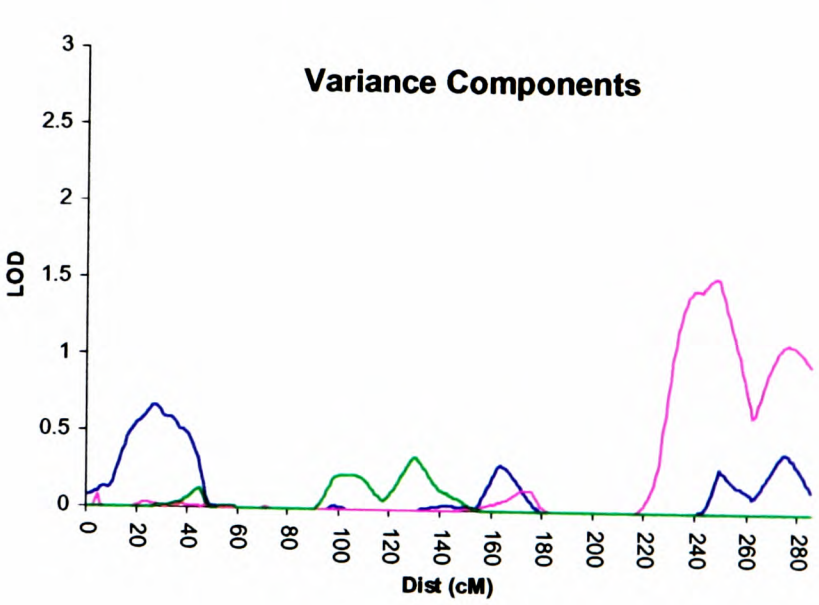


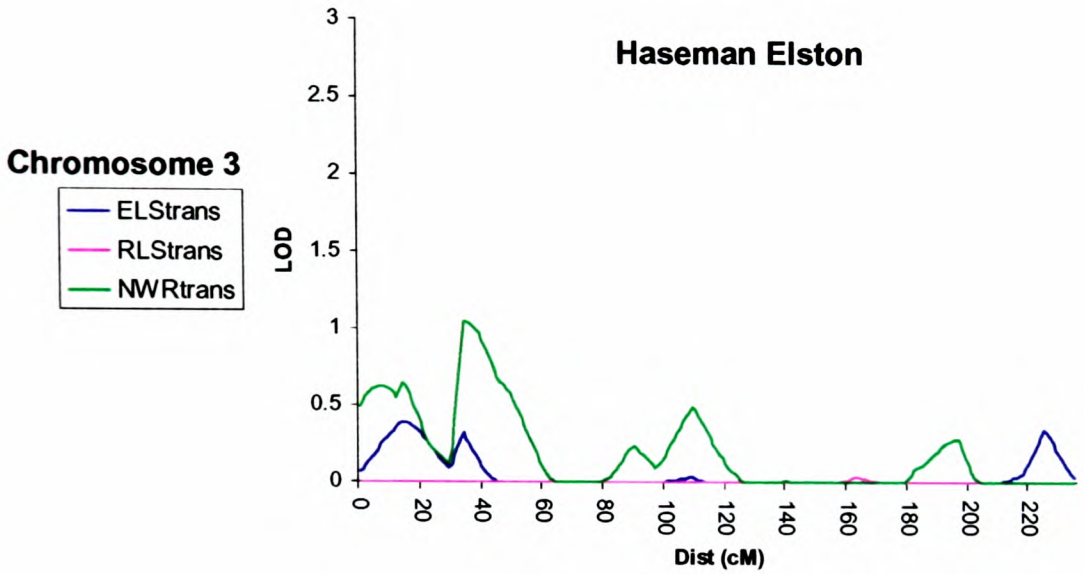


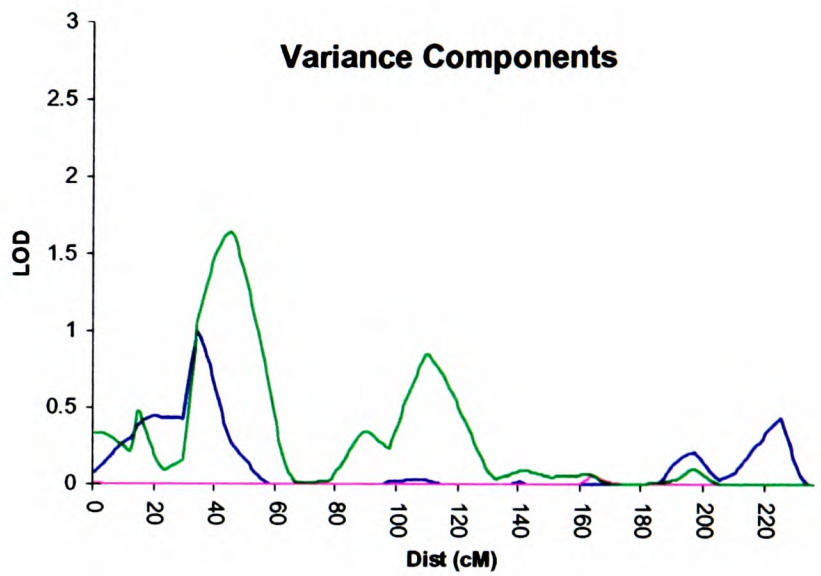
Chromosome 2

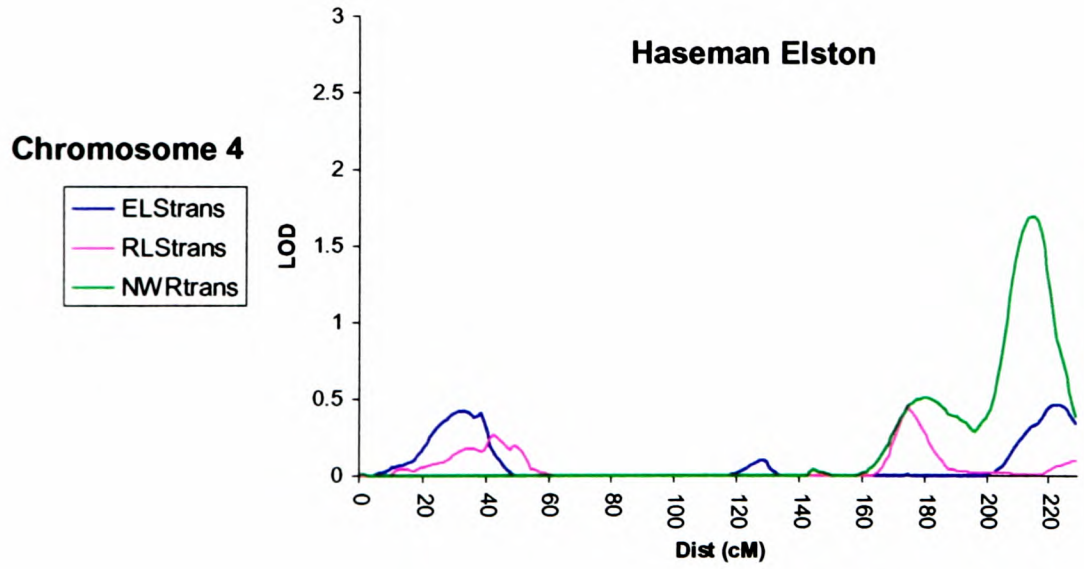


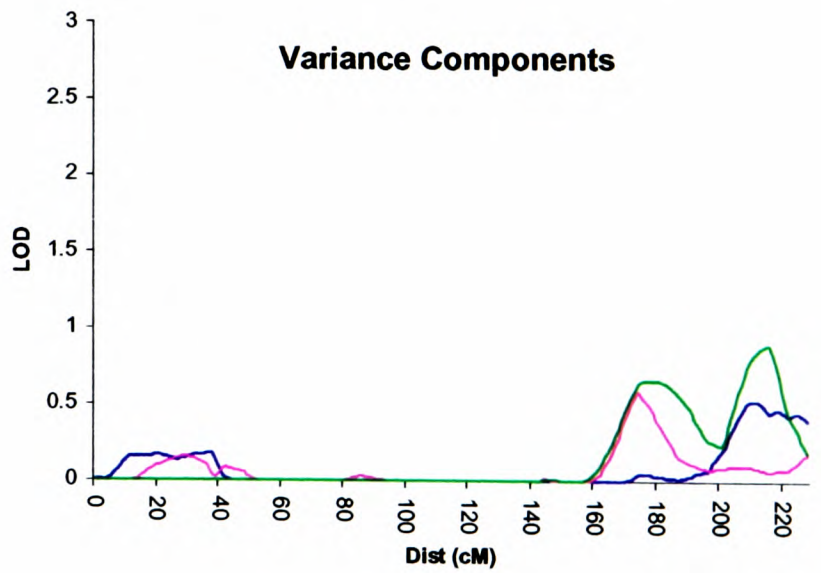


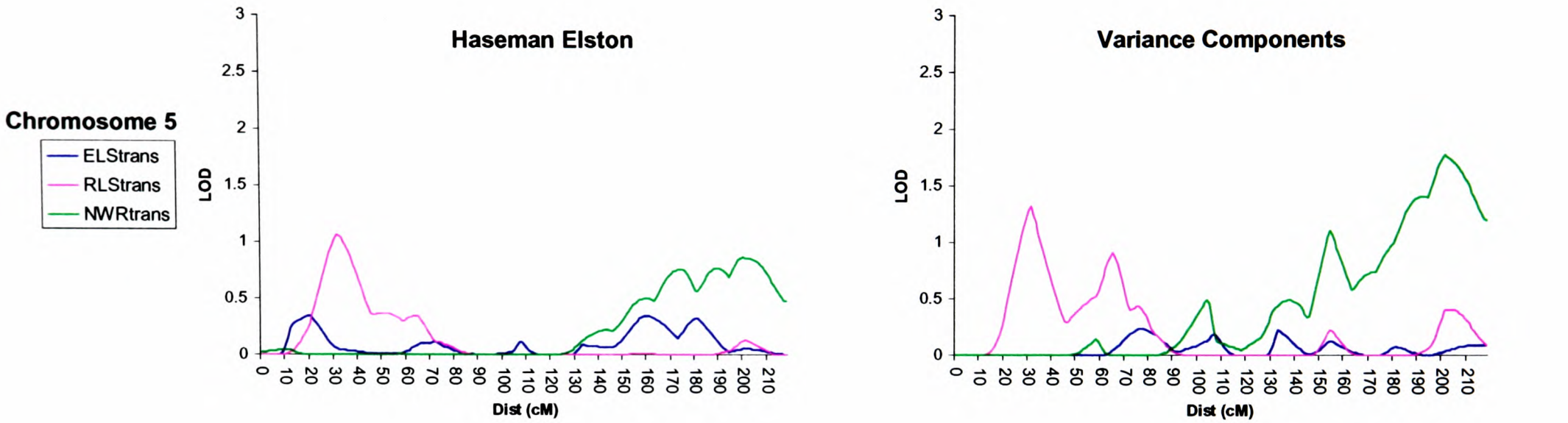


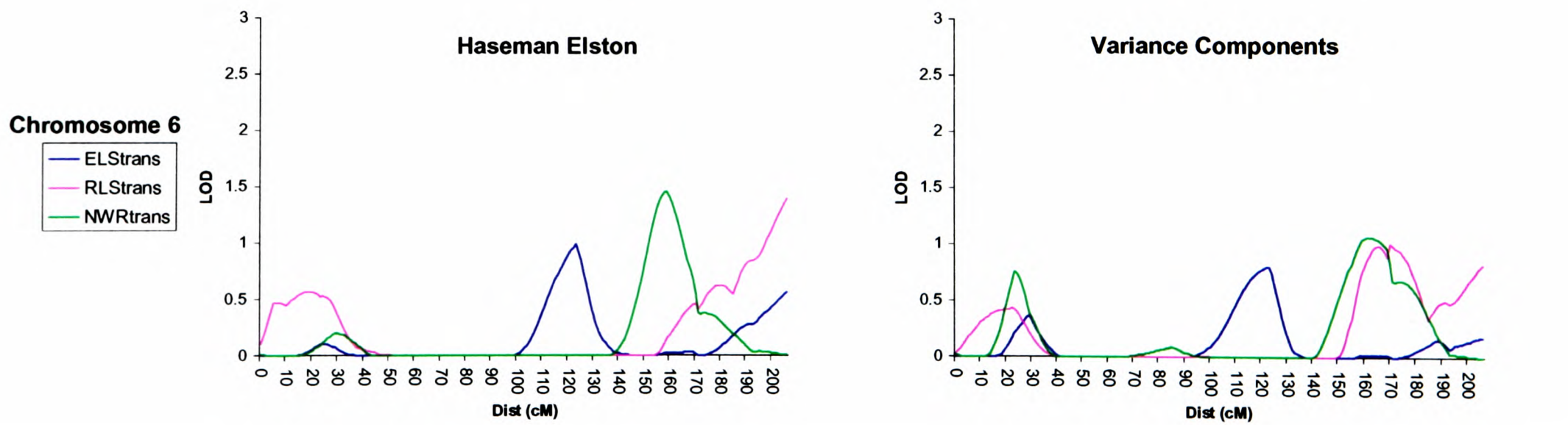

Chromosome 3
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
20
40
60
80
100
120
140
160
180
200
220

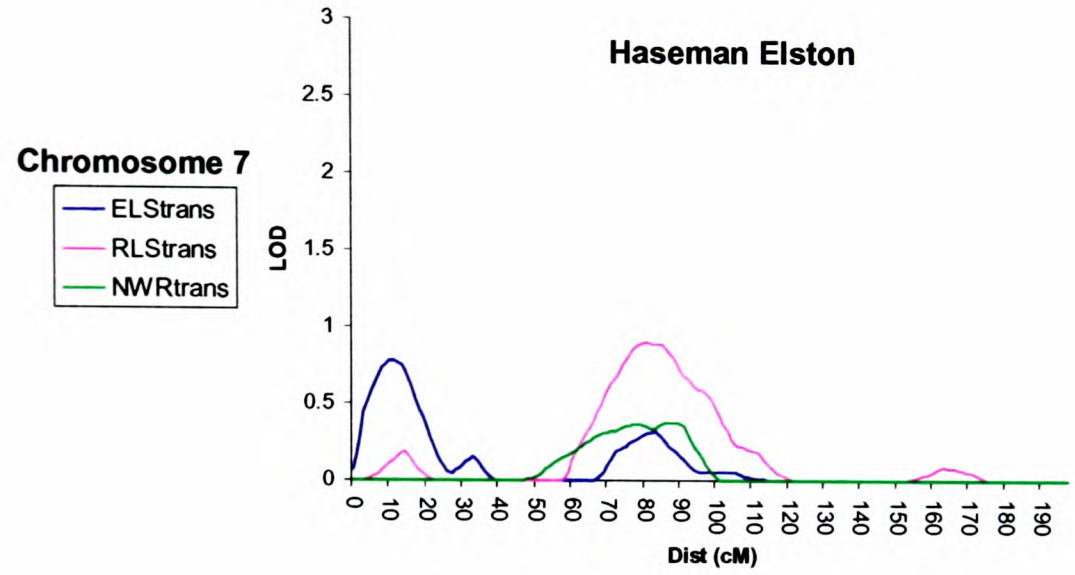

Variance Components
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
20
40
60
80
100
120
140
160
180
200
220

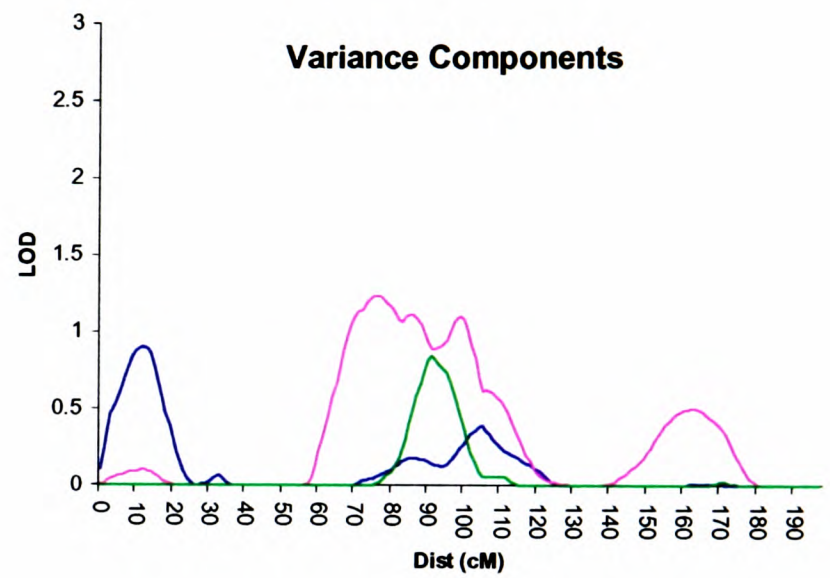

Chromosome 4
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
20
40
60
80
100
120
140
160
180
200
220

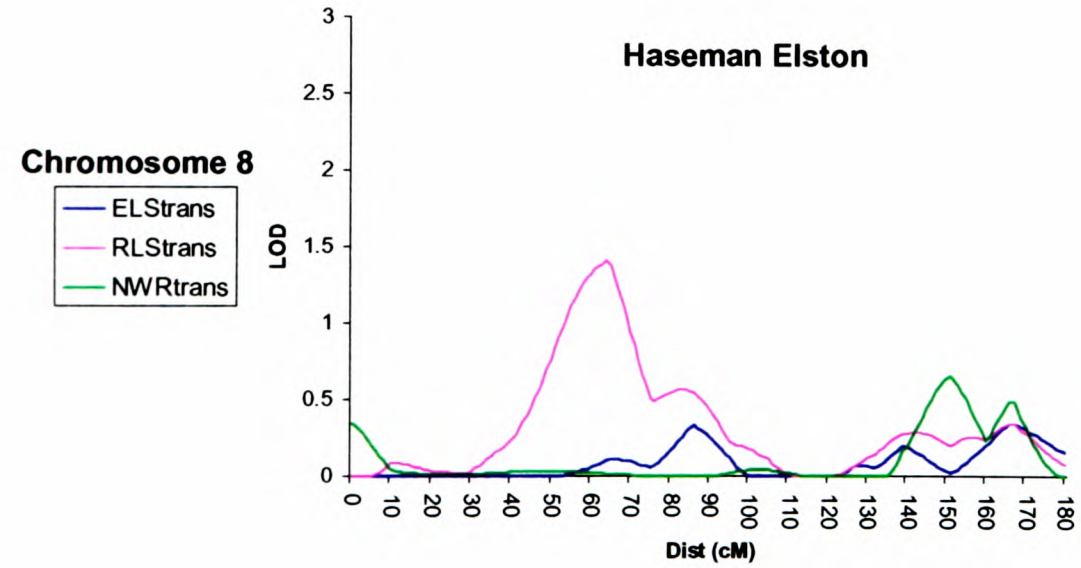

Variance Components
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
20
40
60
80
100
120
140
160
180
200
220

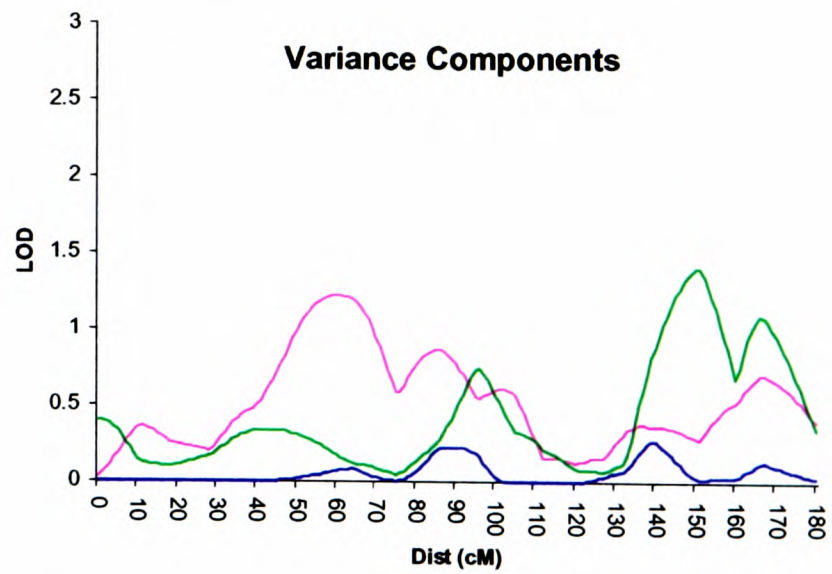

Chromosome 5
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
Variance Components
LOD
Dist (cM)

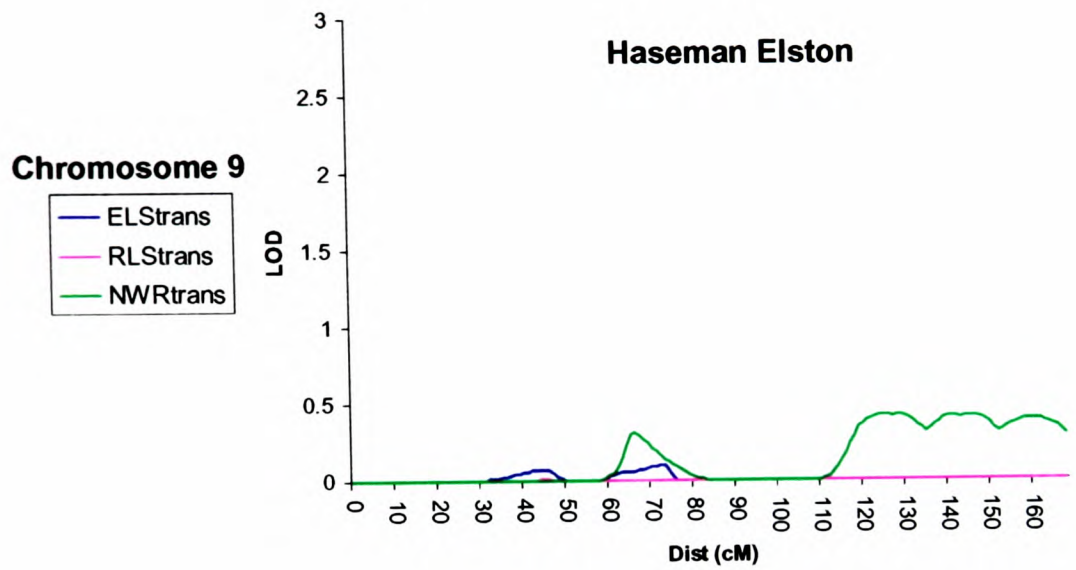

Chromosome 6
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
Variance Components
LOD
Dist (cM)

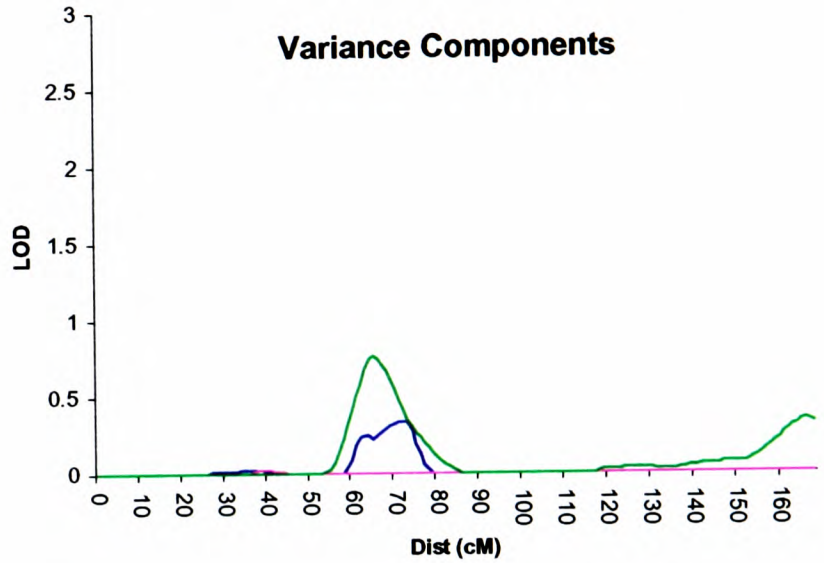

Chromosome 7
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

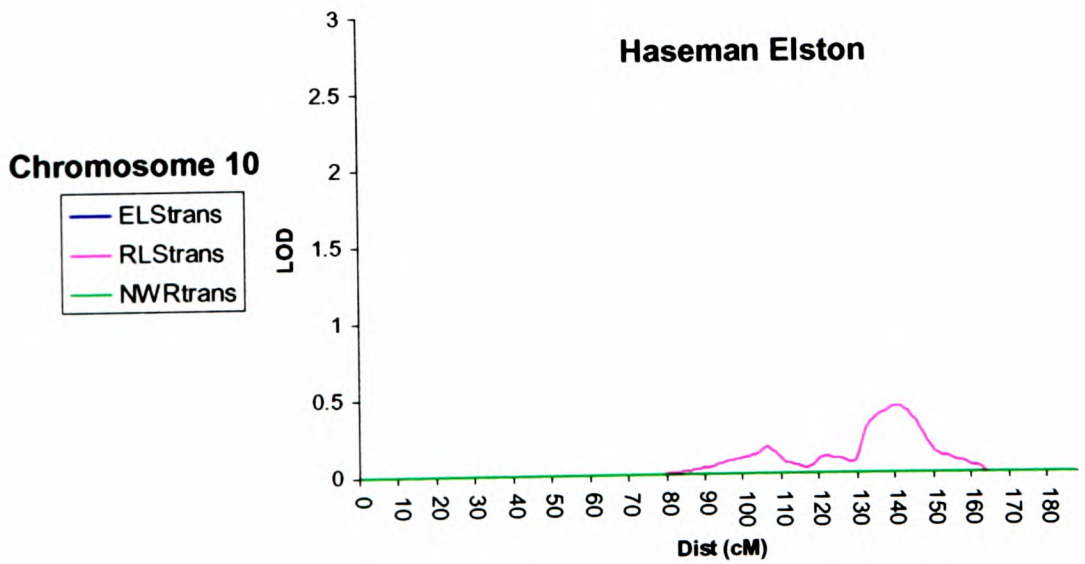

Variance Components
LOD
Dist (cM)

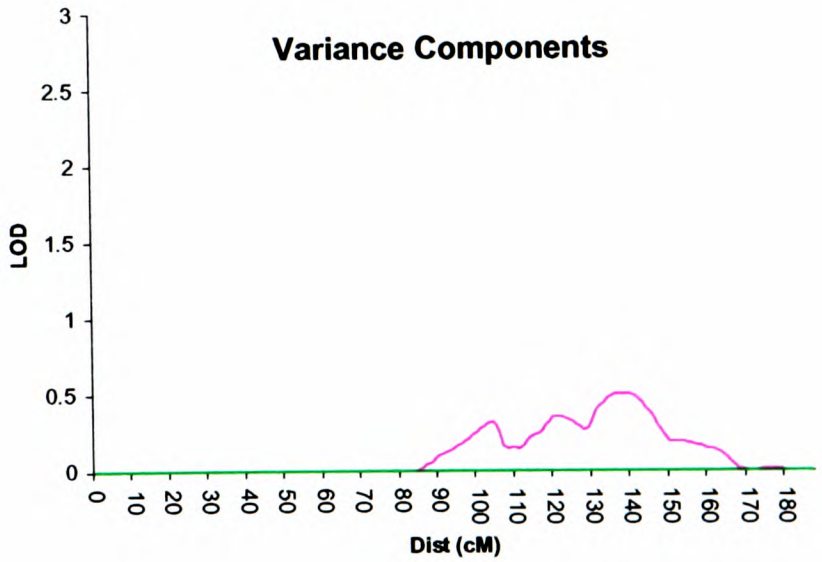

Chromosome 8
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

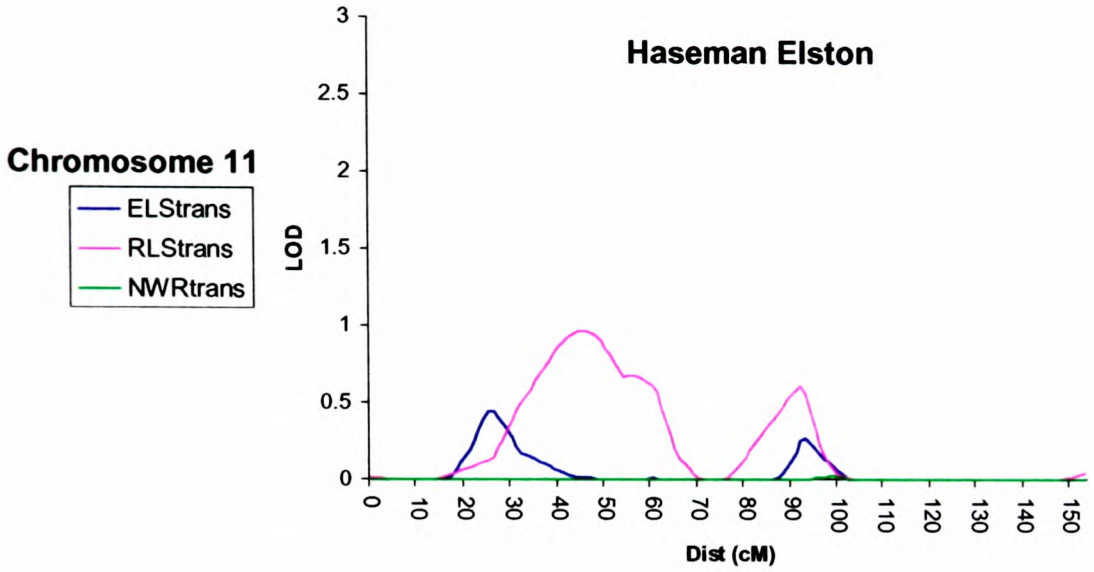

Variance Components
LOD
Dist (cM)

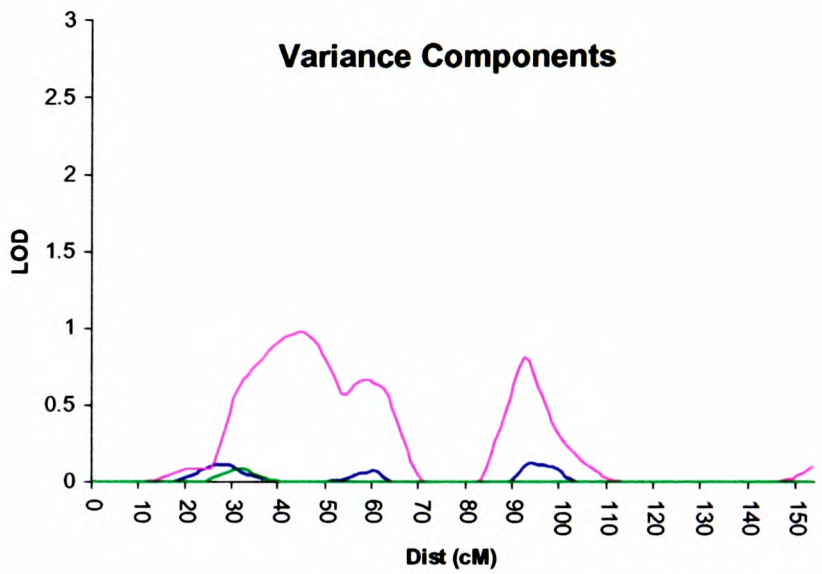

Chromosome 9
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

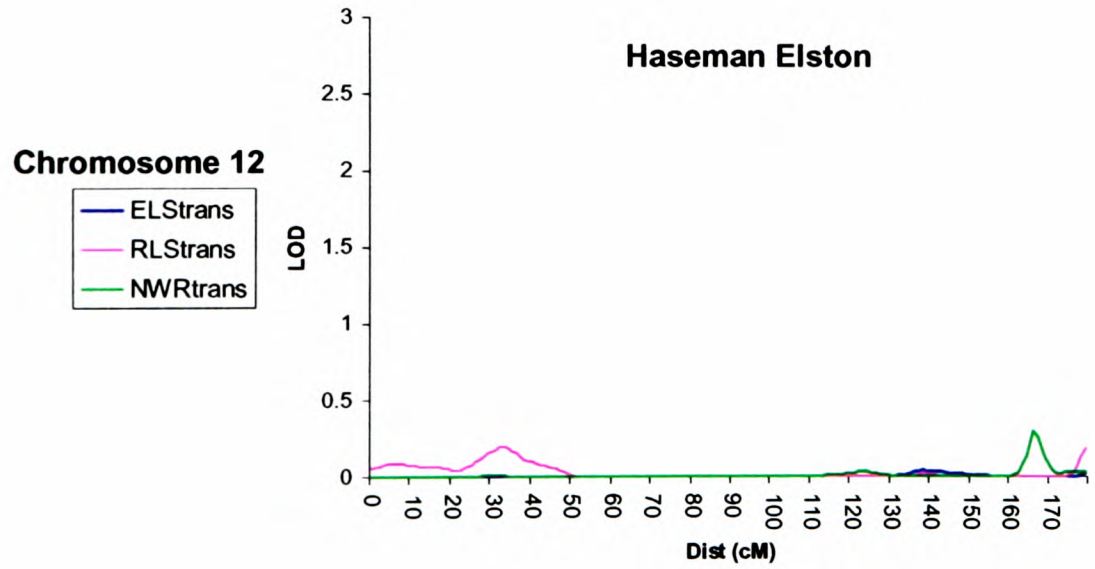

Variance Components
LOD
Dist (cM)

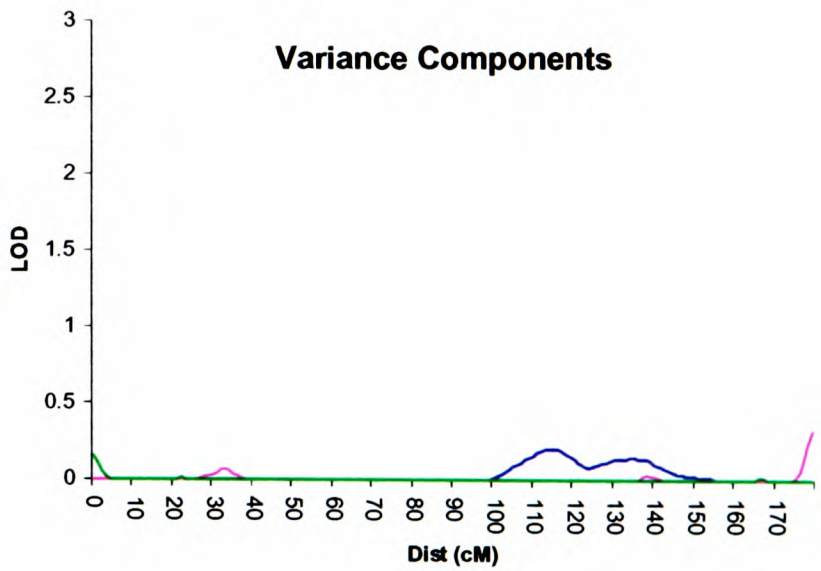

Chromosome 10
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

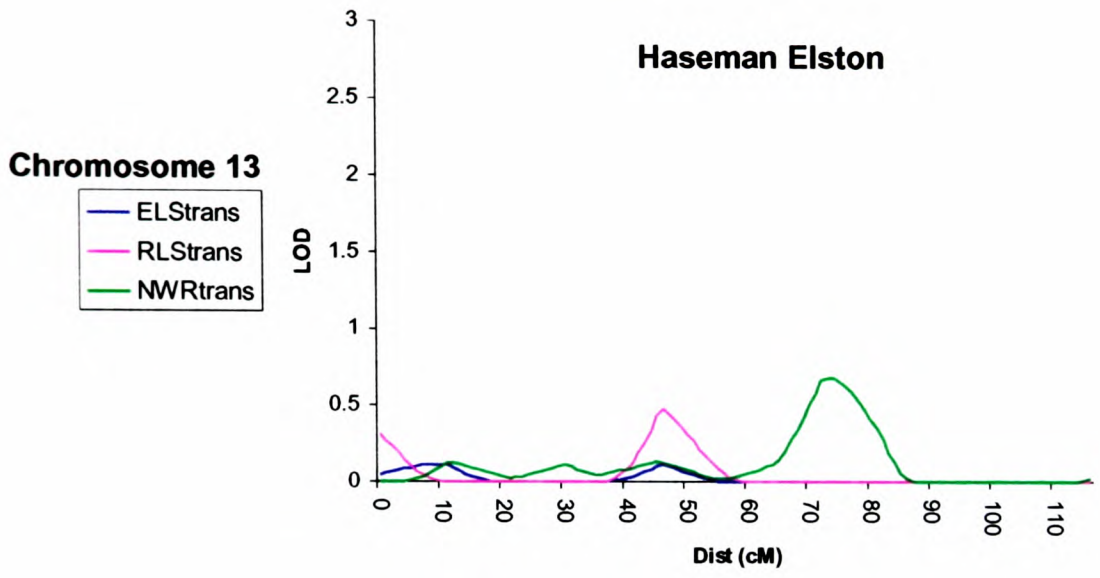

Variance Components
LOD
Dist (cM)

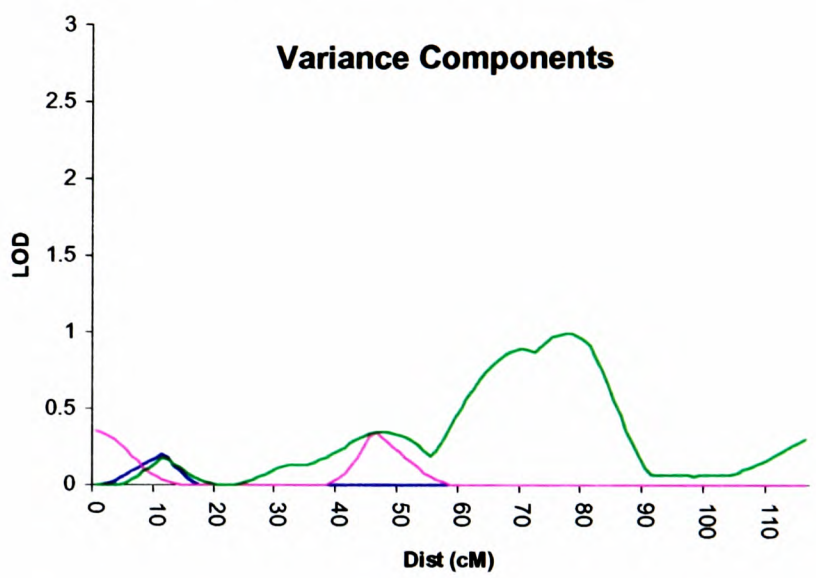

Chromosome 11
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
0 0.5 1 1.5 2 2.5 3

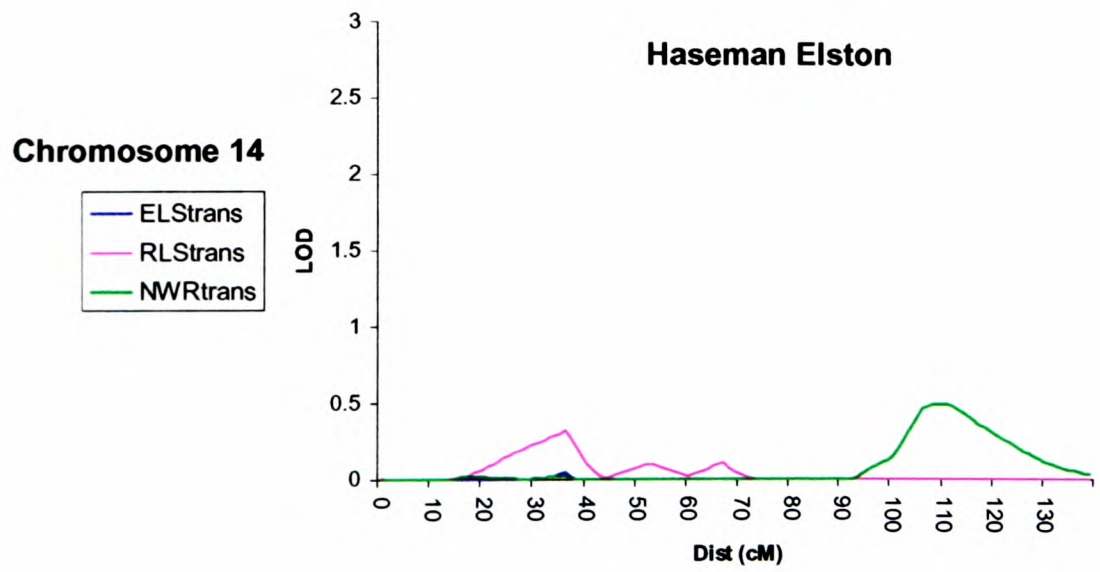

Variance Components
LOD
Dist (cM)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
0 0.5 1 1.5 2 2.5 3

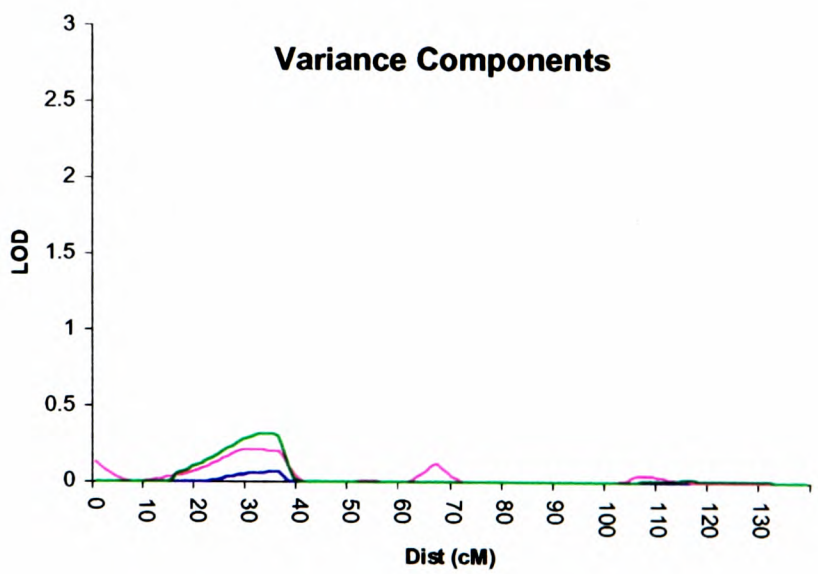

Chromosome 12
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
0 0.5 1 1.5 2 2.5 3

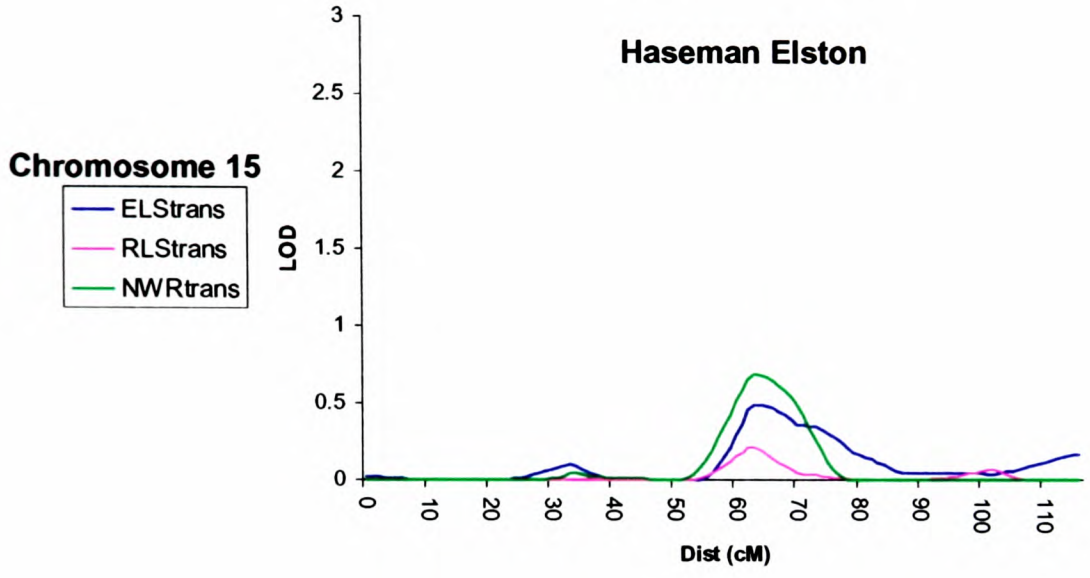

Variance Components
LOD
Dist (cM)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
0 0.5 1 1.5 2 2.5 3

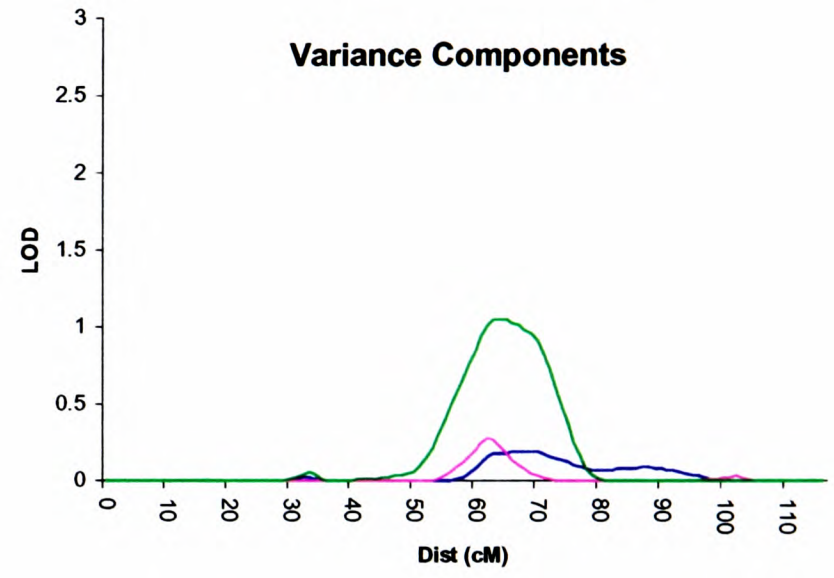

Chromosome 13
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90
100
110

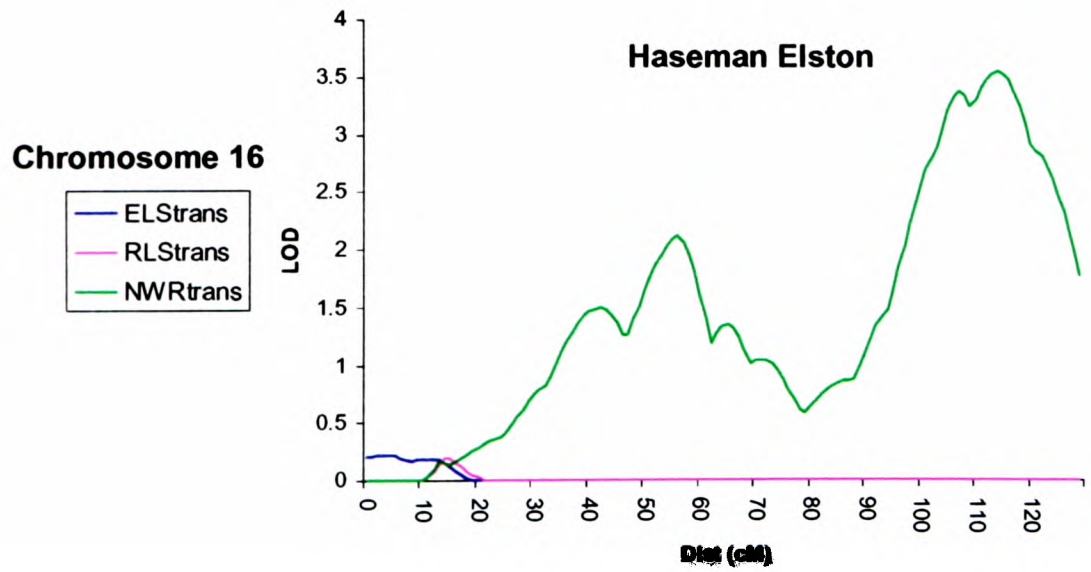

Variance Components
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90
100
110

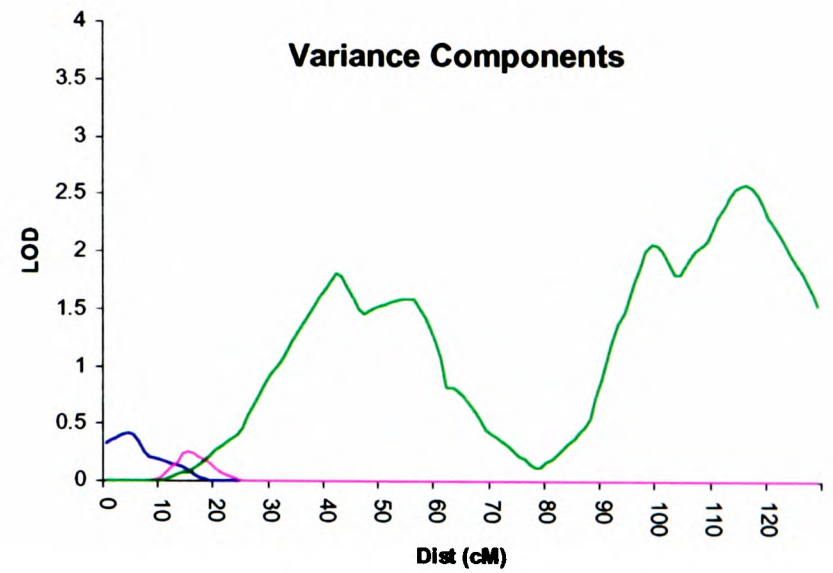

Chromosome 14
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90
100
110
120
130

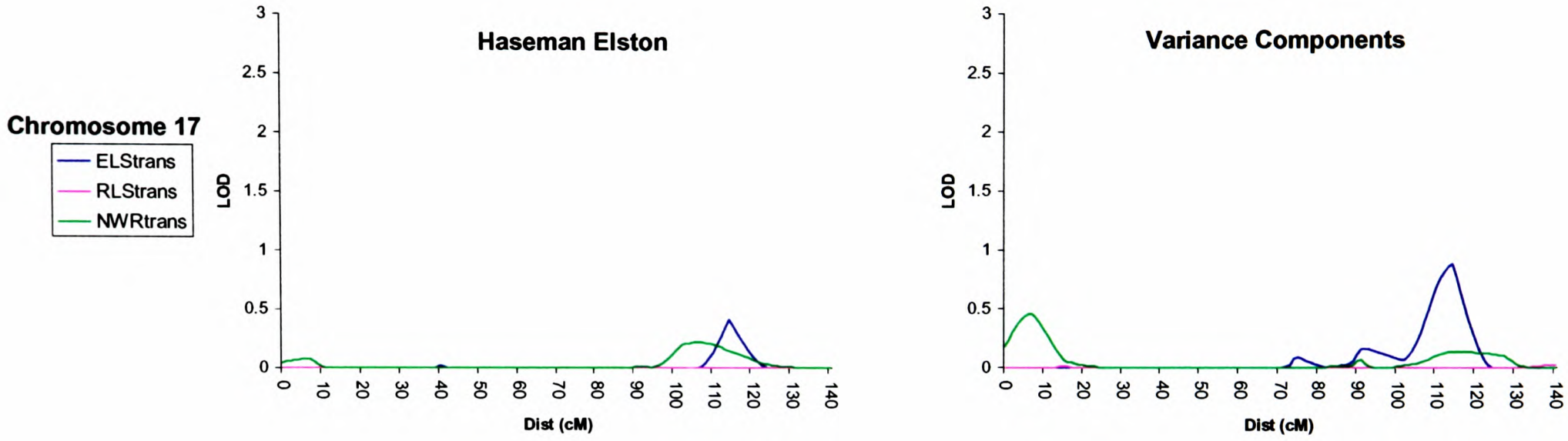

Variance Components
LOD
Dist (cM)
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60
70
80
90
100
110
120
130

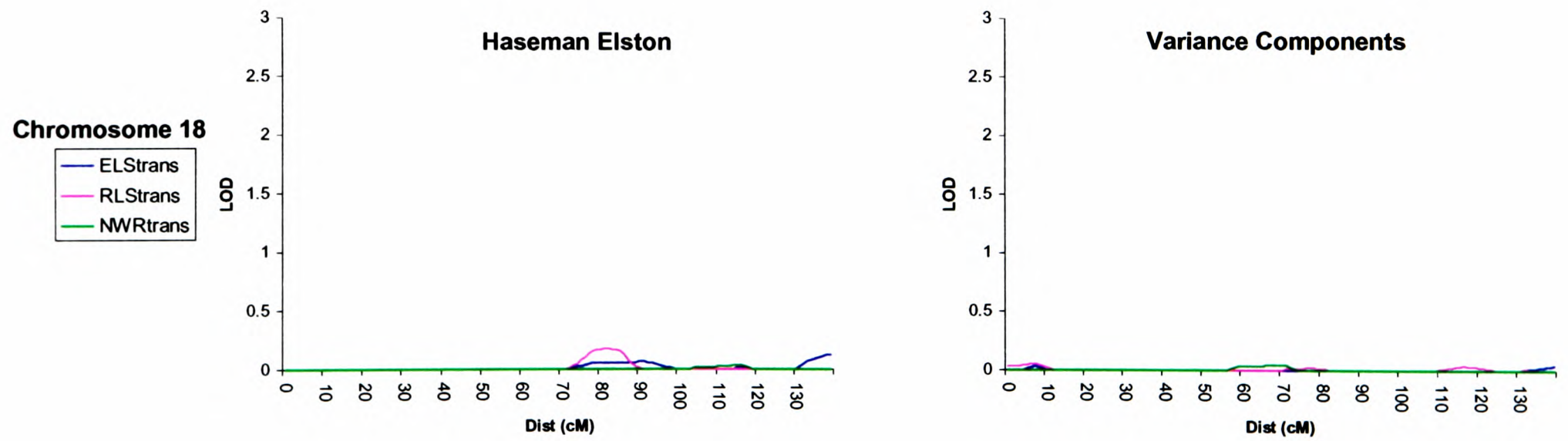

Chromosome 15
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

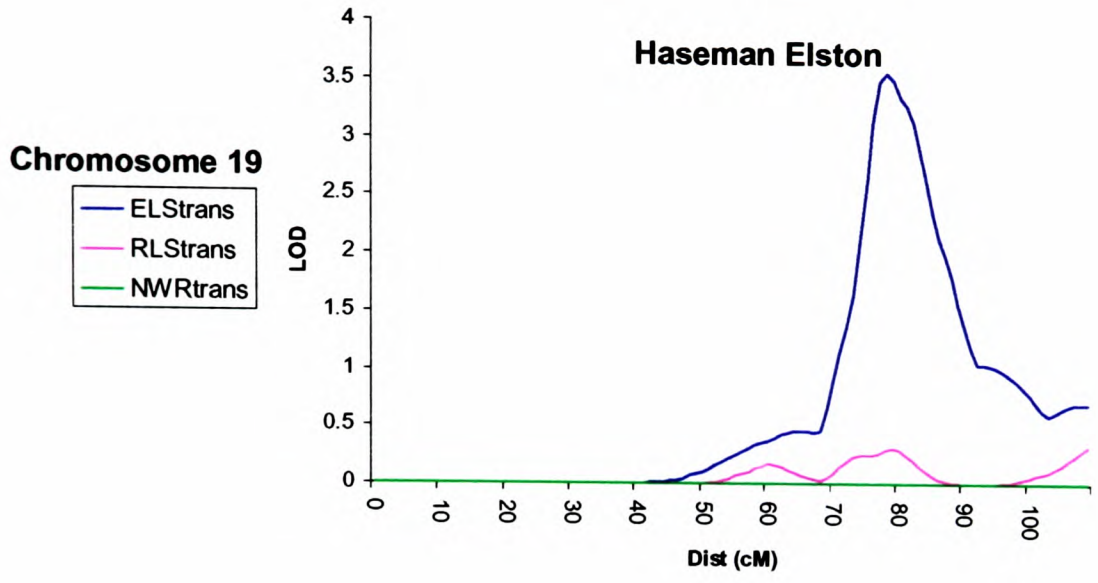

Variance Components
LOD
Dist (cM)

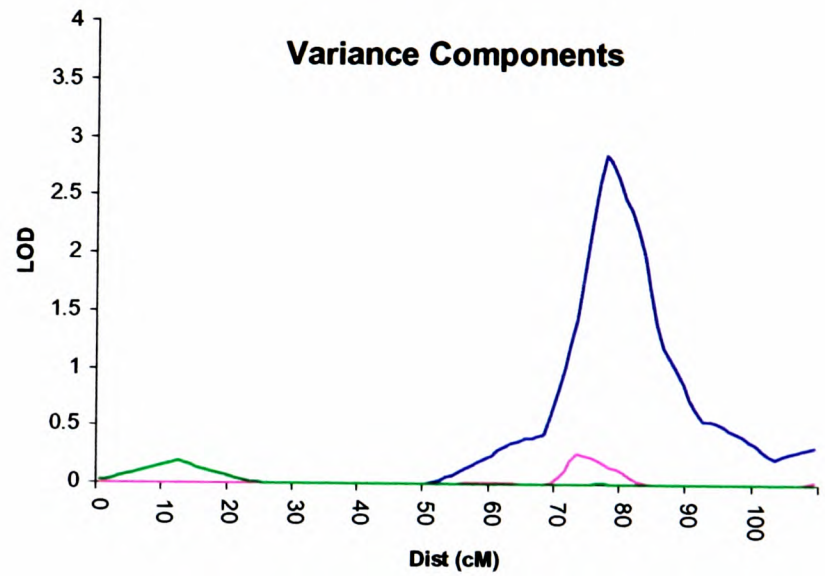

Chromosome 16
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

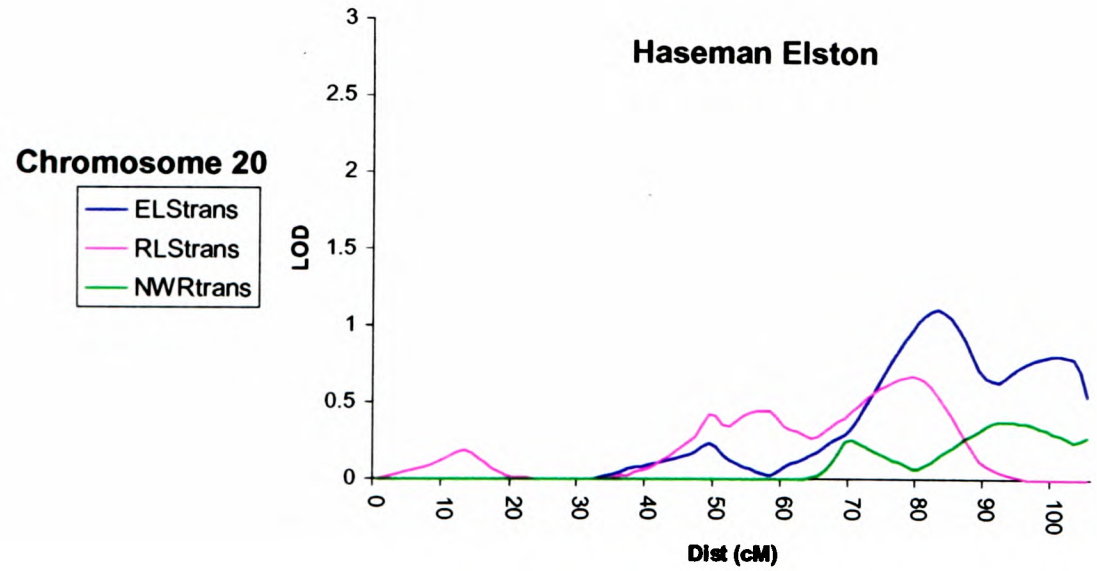

Variance Components
LOD
Dist (cM)

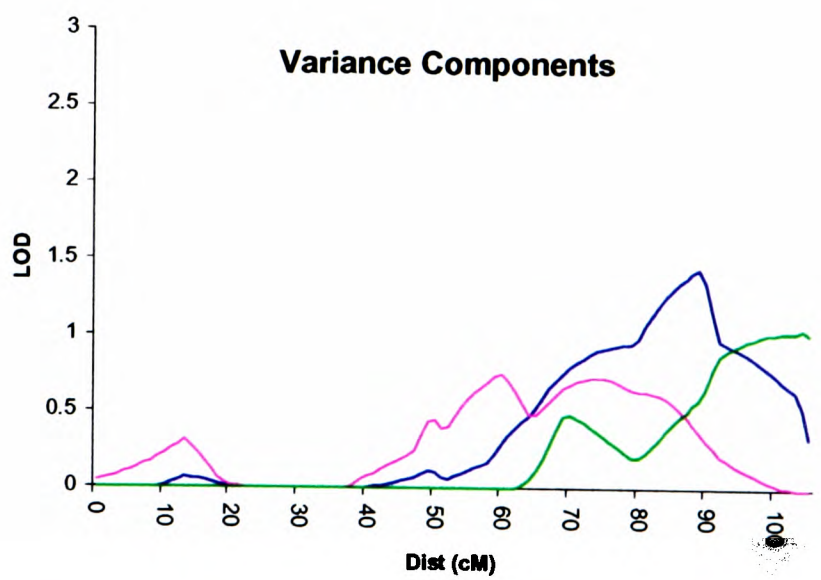

Chromosome 17
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
Variance Components
LOD
Dist (cM)

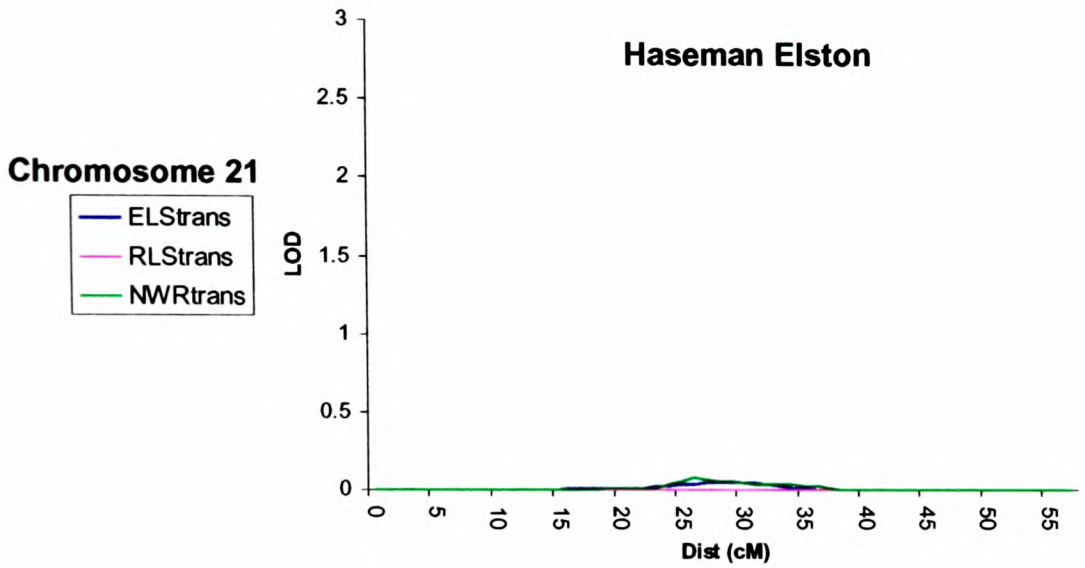

Chromosome 18
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)
Variance Components
LOD
Dist (cM)

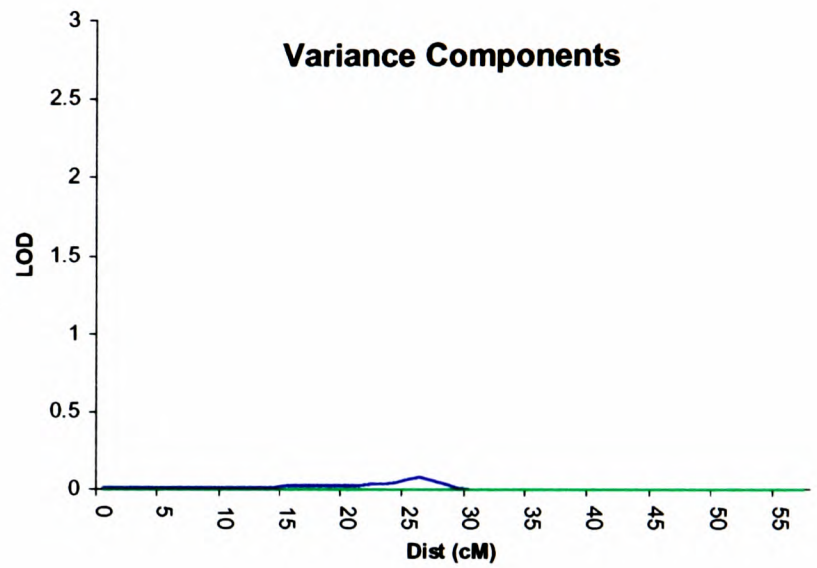

Chromosome 19
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

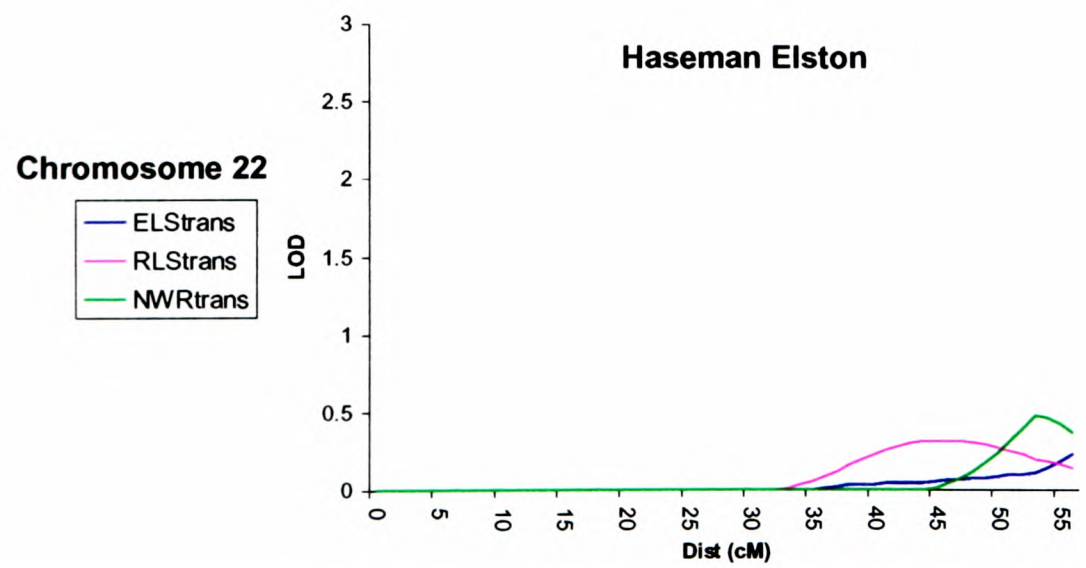

Variance Components
LOD
Dist (cM)

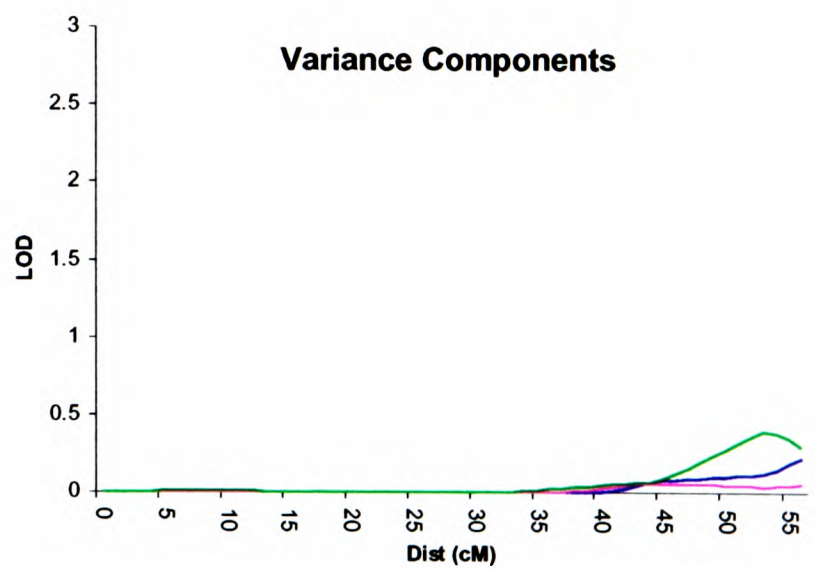

Chromosome 20
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)

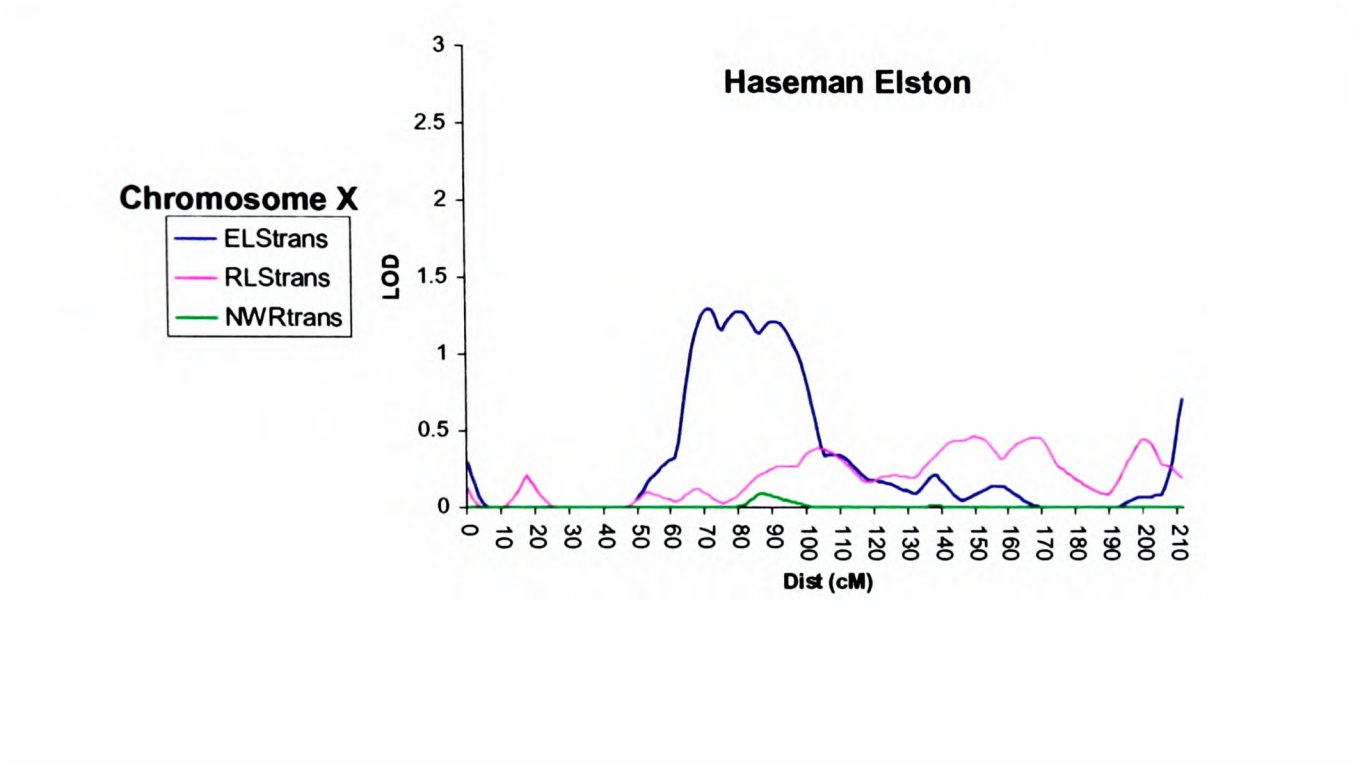

Variance Components
LOD
Dist (cM)


Chromosome 21
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)


Variance Components
LOD
Dist (cM)


Chromosome 22
ELStrans
RLStrans
NWRtrans
Haseman Elston
LOD
Dist (cM)


Variance Components
LOD
Dist (cM)


Haseman Elston
Chromosome X
ELStrans
RLStrans
NWRtrans
LOD
3
2.5
2
1.5
1
0.5
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
Dist (cM)

All *NUMB-R* coding sequence was contained within a single BAC (AC010412) which was 155086 bases in length. This appendix shows the coding sequence for all *NUMB-R* exons, the SNPs used for association analyses, the SNPs detected by DHPLC and some upstream sequence containing a putative CpG island.
The numbers given refer to base positions within the BAC AC010412

```
31381 ccagaaaagg cttcggaaga ccccggggaa ccctgtggtg ctatgaagga cgagggaaat gtgtagacag atgatacgat
31461 gcaggtcaga ggcctggagc aggagctgca aggggcagYg ggggagacag gagagggaca gatgcaggtc agaggcagga
31541 gcaggagctg cgtgggagac aggagaggga cagatgtggg tcacaggcct agagtaggag ctgtgaggga cagcgcggga
31621 gacaggagag agagggttgg caggagaagg catccaccag tgggcgctcg gggcccttcc ctggctgaca cacctccccc

43141 ttactaatgg gctctaatgt ccctgccagc tgctttgcca aacccacagg attggcgtag ggggcggcgg gggcgccgta
43221 gcggccagaa ccgcgccctg gtcccacccc ggcgcccccg ttcccaaagc cgaaccgatc gcgcatcagg acgccttggt
43301 ctgttcacac cccctcccct tccttgcaaa aagtgcctcg tcaccgctcc tgcccgcccc cgccgggctg gggaaccatg
43381 gggttagcgg gaagatttgg gggtgtgggg cggcaaggcg cctctagaag caactctttc ggacgcggag ctggagggga
43461 tctgaggtca ggtgtcacgg agggaaaggg agcccaggcc aggacggacc gggggaaagg actcggcgat gttcataggc
43541 atgctcagct cgggactggc caatgggcgt ggcggggcgg tggccgcgcc ggggaaggcg ggatggtgtc tgcgtgtggg
43621 cggggtcttc aggccaaggg gcggggcggg gcgggacctt tctgcatata cggcaggcgg cgggccaatg ggcggggtcg
43701 gggcggggag ccgaggccga gggacgctcc cccgcccccg gaacggaatc ctcccgggag ccgaggaggc tcgcgcgcgt
43781 gcatcccgca ccatcccccg gccccgggcc ctggccggcg tcagaccgag ctgccgccgc caccaccgca gcagcagcag
43861 tcggggagcc aggcccagcc agggccgcgg gaggcggggg cgcccgggcc ctggatgtcc cgcagcgcgg cggccagcgt
43941 gagaagggcg ggagaggggg gctcgggtcg ggggcggcgc cgctgctgat gtggcggagg gtgggaggcg gccgcggcgc

46431 tgtgaccaga tttgagcacc tatgactgag tatgtgtgac cacgtgacag tgtgtggcca aaaacagtgt gttggtgtga
46511 ctggtctggt tgaggctgca tgtggtctcc gtgggaccaa atgtgtgtct gtccgtgtgc atctgtgacc agatgtcagt
46591 gtgcagcagt ggagatacgt gtaacaagtg tatatagctg cgattgtgtc gtgtgggact gggtcagcaa atgggtgaac
46671 tctgggtgtg attctatctg ggtgggcctg tgacgtcagt atcttgggcg ggYgcctgag gaaacatagc tgtcactggg
46751 tgtggccacg catgtgcggc tctagtgggt tgtctgggcc ccgttctttg tatgtggctg agaatgtgga agtgaccaaa
46831 tttcagtggg tggtgtgacc aggtgtctgc ggtgggaact gagtgggcca ggacatgctg catgtatgtg gacaaatgtc

47341 aagtatcaga ctgagcccca acccagacaa acttctcacc ttctcctccc ctactcgggc aggcggaccc ccggaggcct
47421 gagcggcacc tgccccccagc cccctgtggg gccccggggc ccccagaaac ctgcaggacg gagccaggtg agcgaccagc
47501 ctgggacagg gcagggctga atgggtatgt ggtcttgcct aacacgacca gaggagctgt ccctgcggtc acacaggctt
```

SNP 2604887

Putative CpG

Exon 1
228bp

SNP 2561537

Exon 2
81bp

```
49741 tgttatctga ccccatgctc cctgagcact ggtccctccc cgttgagctc cataccttct gacccccacc cttgacccac
49821 atcactgtcc tgtgtcacca gacggggcgg gcaccatgaa caagttacgg cagagcctgc ggcggaggaa gccagcctac
49901 gtgcccgagg cgtcgcgccc gcaccagtgg caggcagacg aggacgcggt gcggaagggc acgtgcagct tcccggtcag
49981 ggtgagtggg cgcggcaggg tgcggcgggg cgaggggggac gggggccgct gtgctgtgct gtgcgggtgc gcggcctggc

51361 cagtggacac tgagaatgag tgattactgt gctgtttcat gtcccccccc ccattccctc caacccaagt acctgggtca
51441 cgtggaggta gaggaStccc ggggaatgca cgtgtgtgaa gatgcggtga agaagctgaa ggcggtgagt gagggggctg
51521 ccatgaggga gggggcaagg tcagacaggg agaggcaccc tgacccggct cctttgtccc cttgcctccc tccagatggg
51601 ccgaaagtcc gtgaagtctg tcctgtgggt gtcagccgat gggctccgag tggtggacga caaaaccaag gtaggaggcc
51681 tgtgggggtg gacgccaggg ggcctgtctg ctcaccctgc tggaatcagg gcggaggaca ccttcccgcc tcctctgcac

53221 cagtgactgc tgggtcccca ccctgttcag gactggattt cagcagacac agagagaatg ttgagtgaac acttgaaaga
53301 gagttcacac tcttctcctg gcctccccgg cccctcctag gatcttctgg tcgaccagac catcgaaaag gtctcctttt
53381 gtgctcctga ccgcaacctg gacaaggctt tctcctatat ctgccgtgac gggactaccc gccgctggat ctgccactgt
53461 tttctggcac tgaaggactc cgtgagtatt gccacagctg gccaactctt ccctcccatg ccctggcatc ccagccccac

56881 gggcaaccca gtacctagcg gctataaccc agtgcccctg gcccctgtcc gcactgccca gcgcctctaa ccccccccgcc
56961 catcctctcc ccatagggcg agaggctgag ccacgctgtg ggctgtgctt ttgccgcctg cctggagcga aaacagcgac
57041 gggagaagga atgtggggtc acggccgcct tcgatgccag ccgcaccagc ttcgcccgcg agggctcctt ccgcctgtct
57121 gggggtgggc ggcctgctga gcgagaggcc ccggacaaga agaaaggtgg gtgttgcccc aagcgccagc agagcacaga
57201 gtgggcacag ggtacccagt atcaggggggc agggagctgc agaggtgcag agactcaaca gtttacaaag taggggatgc

60661 ccttggagaa tgatcaagag aaccaggtgg gagaggcctt gcatgccatg gtgagatatc taaccatcca tctcccccat
60741 gtctccagca gaggcagcag ctgcccccac tgtggctcct ggccctgccc agcctgggga cgtgtccccg acaccagcca
60821 ccacatcccc tggtgagaag ggtgaggcag gcacccctgt ggctgcaggc accactgcgg ccgccatccc ccggcgccat
60901 gcacccctgg agcagctggt tcgccagggc tccttccgtg ggttcccagc actcagccag aagaactcgc ctttcaaacg
60981 gcagctgagc ctacggctga atgagctgcc atccacgctg cagcgccgca ctgacttcca ggtgaagggc acaggtgagc
61061 cctgggccca gtgggaaacg gcgttggaat caattggcac tatctgatgt gggcaaggga tggaggcgag atggttcaca
```

Exon 3
151bp

Exon 4
75bp
Exon 4 SNP
Exon 5
75bp

Exon 6
141bp

Exon 7
189bp

Exon 8
306bp

```
62541 aagatagagg tggcacgcta ctgagttttt gcctgaggga ggcaggcaga agggctgtct tccccaccag cagccagtga
62621 acccatgtgg tctggcccac tctcttgctc ctggatgcca aatctgcatc gcctacccgt tatgtcctcg gggattgatg
62701 tgggagggtg ttgagctgta tccaggaggc tgSgctgctc caaaaggtaa ccggtgagca gatgccaacc cttgttgaga
62781 gtgacgcacg ctccctccat gcctgcagca ttcctgggcc atcacacacc ctcacatttg ctgtccttaa cgcagcctga
62861 catcaggcac actccccagg ctgtcacacg tgctcctgag aggtggcagg agctccaggc tttatttgga agcctcacac

64321 ggggctgcgg cagggtctgg gcctgtttgc tgacctctgt cctgccctgc ttcccagtgc ctgagatgga gcctcctggt
64401 gccggcgaca gtgacagcat caacgctctg tgcacacaga tcagttcatc ttttgccagt gctggagcgc cagcaccagg
64481 gccaccacct gccacaacag gtaaagccag gctttctgct gggcagacac cctgcatgtc cctcctgtat cccctcctct

66121 gtgcagagct gggattcaga ggcaggcggt ctggctccag agtccgtgct ctgatcccca caccatgcca gttcccagcc
66201 tggggttctt ctgctgaagc cgttgccccc tccctcacct gtccctctcc tctctccagg gacttctgcc tggggtgagc
66281 cctccgtgcc ccctgcagct gccttccagc ctgggcacaa gcggacacct tcagaggctg agcgatggct ggaggaggtg
66361 tcacaggtgg ccaaggccca gcagcagcag cagcagcaac agcaacagca gcagcagcag cagcagcaac agcagcaagc
66441 agcctcagtg gccccagtgc ccaccatgcc tcctgccctg cagcctttcc ccgcccccgt ggggcccttt gacgctgcac
66521 ctgcccaagt ggccgtgttc ctgccacccc cacacatgca gccccctttt gtgcccgcct acccgggctt gggctaccca
66601 ccgatgcccc gggtgcccgt ggtgggcatc acaccctcac agatggtggc aaacgccttc tgctcagccg cccagctcca
66681 gcctcagcct gccactctgc ttgggaaagc tggggccttc ccgccccctg ccatacccag tgcccctggg agccaggccc
66761 gccctcgccc caatggggcc ccctggcccc ctgagccagc gcctgcccca gctccagagt tggacccctt tgaggcccag
66841 tgggcggcat tagaaggcaa agccactgta gagaaaccct ccaacccctt tctggcgac ctgcaaaaga cattcgagat
66921 tgaactgtag Yccgagccgc cccacccact ccatcatctc caggtgcccc acgcctgggg gtggaggcac aacctctccc
67001 ctaaccctgc tccctggggc tgcgcccctc aacaccctct caacaccccc ctccctcaac caccccgaca accactacag
67081 aaccaacatt gtgacgccca ggttgcaaca ggatggaatt cagggacgga cccagcctgg ctaagggaac catttcactg
67161 ccggacttag gctggcaatg ccccccttccc caaccccaga cacaggggtt ggccacaatc ccactgaatg cccttggttc
67241 acactccatt tcccagtttc tgttgacccc caccttccag tgttggacag gatggagggg ggacacttgc ttaggggctc
67321 tcctgggccc cacaccagtg cccaccccaa atctggtcgt ctcctccccc catgcacagc acaagctaag ggctgccctc
67401 tgcccacacg ctgcgttcac tgccaatgct gtactcacct ccatcaccct ccaactttgg ggcccatgtc ttccttgggc
67481 caaggtctca tgggggctag ggccaagttg ggggcccagg aggcggggag ggaggaggag gagaagatgc gcagttacct
67561 catgtcggtg cccgctgggg aggggtccgg gaagaagggg aaggggtgcc tggcRggtac ttttctatct tttatttcca
67641 gatttttttt gtatctaaac ttgaagattt gtattataca aggacagcca ataaaggaag aatataatgg tgcccctcgg
67721 gaaacgaggg tgtgtgtagg gcgctggtga ggccctagaa cctttcctcc tgagtcatgg actgaaacag agtcagaagg
```

SNP 2604892

Exon 9
126bp

Exon 10
1444bp
5 fragments
(10a, 10b, 10c,
10d + 10e)

Exon 10c SNP

Exon 10e SNP

```
67981 ctgtaatccc agcactttgg gaggtcgagg caggcagatc acttaaggtc aggagttcga gaccagcctg gccgccatgg
68061 agaaacccag tctctaccaa aaattcaaaa attagccggg cgtggtggca ggcgcctgta attccaggag gctgaggcac
68141 aagaatcgct tgaacccRgg cagcagaggt tgcagtgacc tgaggtcaca ccattgcact cctgcctggg caacagaatg
68221 agactctgtc tcaaaaaaaa aaaacaaaaa aaaaaaaact acttctaatt agctcaatat taatatttta acaagttggg
68301 ttggtaacag tatatctttg cccatgctgg caaattcttg ttttgtcagc attttccata actctggcca aagtgtcacc
68381 tgatgtggca acgttttaca gtcttgctat tgtttcttga gtcctttaat ctataagatg tatttttaaa aatatataac
```

SNP 2561551

KEY:

Coding sequence is shown in red
5' and 3' untranslated regions are shown in pink
Non-coding sequence is shown in black
SNPs are shown in blue
Putative CpG islands are highlighted in grey

Conserved residues within the *NUMB-R* gene are boxed in red
The polyglutamine tract is boxed in black
Primers for the amplification of all regions are underlined (NB all primer sequences can also be found in appendix M)

atg|tcc|cgc|agc|gcg|gcg|gcc|agc|ggc|gga|ccc|cgg|agg|cct|gag|cgg|cac|ctg|ccc|cca|gcc|ccc|tgt|ggg|gcc|ccg|ggg|ccc|
M S R S A A A S G G P R R P Q R H L P P A P C G A P G P
cca|gaa|acc|tgc|agg|acg|gag|cca|gac|ggg|gcg|ggc|acc|atg|aac|aag|tta|cgg|cag|agc|ctg|cgg|cgg|agg|aag|cca|gcc|tac|
P Q T C R T E P D G A G T M N K L R Q S L R R R K P A Y
gtg|ccc|gag|gcg|tcg|cgc|ccg|cac|cag|tgg|cag|gca|gac|gag|gac|gcg|gtg|cgg|aag|ggc|acg|tgc|agc|ttc|ccg|gtc|agg|tac|
V P E A S R P H Q W Q A D E A A V R K G T C S F P V R Y
ctg|ggt|cac|gtg|gag|gta|gag|gag|tcc|cgg|gga|atg|cac|gtg|tgt|gaa|gat|gcg|gtg|aag|aag|ctg|aag|gcg|atg|ggc|cga|aag|
L G H V E V E E S R G M H V C E D A V K K L K A M G R K
tcc|gtg|aag|tct|gtc|ctg|tgg|gtg|tca|gcc|gat|ggg|ctc|cga|gtg|gtg|gac|gac|aaa|acc|aag|gat|ctt|ctg|gtc|gac|cag|acc
S V K S V L W V S A D G L R V V D D K T K D L L V D Q T
atc|gaa|aag|gtc|tcc|**ttt**|**tgt**|gct|cct|gac|cgc|aac|ctg|gac|aag|gct|ttc|tcc|tat|atc|tgt|cgt|gac|ggg|act|acc|cgc|cgc|
I E K V S **F** **C** V P D R N L D K A F S Y I C R D G T T R R
tgg|atc|tgc|cac|tgt|ttt|ctg|gca|ctg|aag|gac|tcc|ggc|gag|agg|ctg|agc|cac|gct|gtg|ggc|tgt|gct|**ttt**|gcc|gcc|tgc|ctg|
W I C H C F L A L K D S G E R L S H A V G C A **F** A A C L
gag|cga|aaa|cag|cga|cgg|gag|aag|gaa|tgt|ggg|gtc|acg|gcc|gcc|ttc|gat|gcc|agc|cgc|acc|agc|ttc|gcc|cgc|gag|ggc|tcc|
E R K Q R R E K E C G V T A A F D A S R T S F A R E G S
ttc|cgc|ctg|tct|ggg|ggt|ggg|cgg|cct|gct|gag|cga|gag|gcc|ccg|gac|aag|aag|aaa|gca|gag|gca|gca|gct|gcc|ccc|act|gtg|
F R L S G G G R P A E R E A P D K K K A E A A A A P T V
gct|cct|ggc|cct|gcc|cag|cct|ggg|cac|gtg|tcc|ccg|aca|cca|gcc|acc|aca|tcc|cct|ggt|gag|aag|ggt|gag|gca|ggc|acc|cct|
A P G P A Q P G H V S P T P A T T S P G E K G E A G T P

gtg|gct|gca|ggc|acc|act|gcg|gcc|gcc|atc|ccc|cgg|cgc|cat|gca|ccc|ctg|gag|cag|ctg|gtt|cgc|cag|ggc|tcc|ttc|cgt|ggg|
V A A G T T A A A I P R R H A P L E Q L V R Q G S F R G
ttc|cca|gca|ctc|agc|cag|aag|aac|tcg|cct|ttc|aaa|cgg|cag|ctg|agc|cta|cgg|ctg|aat|gag|ctg|cca|tcc|acg|ctg|cag|cgc|
F P A L S Q K N S P F K R Q L S L R L N E L P F T L Q R
cgc|act|gac|ttc|cag|gtg|aag|ggc|aca|gtg|cct|gag|atg|gag|cct|cct|ggt|gcc|ggc|gac|agt|gac|agc|atc|aac|gct|ctg|tgc|
R T D F Q V K G T V P E M E P P G A G D S D S I N A L C
aca|cag|atc|agt|tca|tct|ttt|gcc|agt|gct|gga|gcg|cca|gca|cca|ggg|cca|cca|cct|gcc|aca|aca|ggt|act|tct|gcc|tgg|ggt|
T Q I S S S F A S A G A P A P G P P P A T T G T S A W G
gag|ccc|tcc|gtg|ccc|cct|gca|gct|gcc|ttc|cag|cct|ggg|cac|aag|cgg|aca|cct|tca|gag|gct|gag|cga|tgg|ctg|gag|gag|gtg|
E P S V P P A A A F Q P G H K R T P S E A E R W L E E V
tca|cag|gtg|gcc|aag|gcc|cag|cag|cag|cag|cag|cag|caa|cag|caa|cag|cag|cag|cag|cag|cag|cag|caa|cag|cag|caa|gca|gcc|
S Q V A K A Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q A A
tca|gtg|gcc|cca|gtg|ccc|acc|atg|cct|cct|gcc|ctg|cag|cct|ttc|ccc|gcc|ccc|gtg|ggg|ccc|ttt|gac|gct|gca|cct|gcc|caa|
S V A P V P T M P P A L Q P F P A P V G P F D A A P A Q
gtg|gcc|gtg|ttc|ctg|cca|ccc|cca|cac|atg|cag|ccc|cct|ttt|gtg|ccc|gcc|tac|ccg|ggc|ttg|ggc|tac|cca|ccg|atg|ccc|cgg|
V A V F L P P P H M Q P P F V P A Y P G L G Y P P M P R
gtg|ccc|gtg|gtg|ggc|atc|aca|ccc|tca|cag|atg|gtg|gca|aac|gcc|ttc|tgc|tca|gcc|gcc|cag|ctc|cag|cct|cag|cct|gcc|act|
V P V V G I T P S Q M V A N A F C S A A Q L Q P Q P A T
ctg|ctt|ggg|aaa|gct|ggg|gcc|ttc|ccg|ccc|cct|gcc|ata|ccc|agt|gcc|cct|ggg|agc|cag|gcc|cgc|cct|cgc|ccc|aat|ggg|gcc|
L L G K A G A F P P P A I P S A P G S Q A A P A P N G A

ccc|tgg|ccc|cct|gag|cca|gcg|cct|gcc|cca|gct|cca|gag|ttg|gac|ccc|ttt|gag|gcc|cag|tgg|gcg|gca|tta|gaa|ggc|aaa|gcc|

P W P P E P A P A P A P E L D P F E A Q W A A L E G K A

*act|gta|gag|aaa|ccc|tcc|**aac|ccc|ttt**|tct|ggc|gac|ctg|caa|aag|aca|ttc|gag|att|gaa|ctg*

T V E K P S ***N P F*** S G D L Q K T F E I E L

KEY: Alternate exons are shown in normal text or *italics* accordingly
The PTB binding domain is shown in red, the polyglutamine stretch is shown in blue and the EH binding domain is shown in green
The critical residues of each binding domain are shown in **bold**
The position of the base change found in family 27 is shown in pink

Region	Forward Primer Sequence	Reverse Primer Sequence	PCR Conditions[a]
Exons			
Exon 1	GGGCGGGACCTTTCTGCATAT	GCCGCCTCCCACCCTCCGCCACA	2.0mM WAVE 63-55.5°cx15 59°cx25
Exon 2	TGAGCCCCAACCCAGACAAACT	CACATACCCATTCAGCCCTGCC	2.0mM WAVE 64-56.5°cx15 60°c x25
Exon 3	TCCCCGTTGAGCTCCATACCTTCTGA	GCACCCGCACAGCACAGCACAG	1.5mM WAVE 68-60.5°c x15 64°c x30
Exons 4+5	TACTGTGCTGTTTCATGTCCC	ATTCCAGCAGGGTGAGCAGA	1.5Mm WAVE 54-46.5°c x15 52°c x25
Exon 6	TGGATTTCAGCAGACACAGAGAGA	AGGGCATGGGAGGGAAGAGTT	1.5Mm WAVE 54-46.5°c x15 52°c x25
Exon 7	CCCTGGCCCCTGTCCGCACTG	AGCTCCCTGCCCCCTGATACTGG	2.0mM WAVE 68-60.5°c x15 65°c x25
Exon 8	TCAAGAGAACCAGGTGGGAGAG	TCAGATAGTGCCAATTGATTCCA	1.5mM WAVE 54-46.5°c x15 54°c x25
Exon 9	CTGGGCCTGTTTGCTGACCTCTG	GGATACAGGAGGGACATGCAGGG	C38 2mM Ta=61°c
Exon 10a	CCCACACCATGCCAGTTCCCAGC	CGGGTAGGCGGGCACAAAAGGG	1.0mM WAVE 68-60.5°c x15 65°c x25
Exon 10b	CCAGTGCCCACCATGCCTCCT	TTGCCTTCTAATGCCGCCCAC	1.5mM WAVE 65-57.5°c x15 61°c x25
Exon 10c	ACTCTGCTTGGGAAAGCTGG	GCGTCACAATGTTGGTTCTG	1.5mM WAVE 64-56.5°c x15 60°c x25
Exon 10d	CGCCCCTCAACACCCTCTCAACA	CAGCATTGGCAGTGAACGCAGC	1.5mM WAVE 65-57.5°c x15 63°c x25
Exon 10e	ACACCAGTGCCCACCCCAAA	CGCCCTACACACACCCTCGTT	1.5mM WAVE 62-54.5°c x15 59°c x25
SNPs			
Exon 10c SNP	See exon 10c		
SNP 2604887	AAGGACGAGGGAAATGTGTAGA	GAGGTGTGTCAGCCAGGGAAGG	C38 2mM Ta=59°c
SNP 2561537	CAGATTTGAGCACCTATGAC	AGTTCCCACCGCAGACACCT	C38 2mM Ta=53°c
SNP 2604892	GCTACTGAGTTTTTGCCTGAG	CTGGGGAGTGTGCCTGATGTC	C38 2mM Ta=55°c
SNP 2561551	GGAGGTCGAGGCAGGCAGAT	CTCAAGAAACAATAGCAAGA	2.0mM WAVE 56-48.5°c x15 51°cx25

a: For PCR protocols please see section 2.4 'The polymerase chain reaction (PCR)'

The *FOXP2* coding sequence was contained within 3 BACs (AC073626 (exon 1), AC003992 (exons2, 3a and 3b) and AC020606 (exons 4 to 17)). The 5' SNP and CpG islands are contained within BAC AC092606 and lie approximately 340Kb upstream of exon 1.
This appendix shows the coding sequence for all *FOXP2* exons, the SNPs and microsatellites used for association analyses, the SNPs detected by DHPLC and some upstream sequence containing two putative CpG islands. The numbers given refer to base positions within the appropriate BAC sequence

```
138151 cctgtccccc accccgcccc tccgaggaga ggcagccccg cgagcagtgt gtgtacagtt cgcattgcca acaacacgcg
138231 cgcgcggccg ggtgcgcgcg cctgccagcg cacattcaca cacacagaca cacagacaca cacacacaca cacacacaca
138311 cacagatccg ggtgggggga gcggctgcgc gagacagcgc gagcctccga gaaagcgcga gacacgccgg cgcgtgcagc
138391 tccgcggccg ccgcttcgcc ctagctctag ccccgcgcca cccgcagccc gcccgcgaac gcccgccccg gttatttatg
138471 cggcggccgc gtccgctggc tgcggcttcc tcggcccccc ccctccccgg gcgcgccccc cactcgcggc agcagctgcc
138551 cggactcgcg cgtgggtgtg ttgtttgggg gcttctgcct cgccgccgcg ggtgccacct cccgggacgc tgcccacggc
138631 gtccccggtc gcggtaagtt tcttggccct cactctggcg cgctacacct ccgcacccca ccctgtccca gccacctcca
138711 cgctgggccg agctgcgact ttactctgct ccgcgctcct cccgcggtgg cgacaaagtt tcgccccaaa ggcagcgccc
138791 tgcttgccgg ggcgagtgtg acatgtgtca aatttgggct cggcgttggg gtcgcattcc ggacccagca tcaccacttg
138871 ttgttctctt tcattttgct ttgtgatggg ggaaaaaggg gtggagaggg gagatttgct gttggctcac ggatgatttt

139511 cgtgggagga gcgcagacac ctttcggtga tagggagggc ttctcacttg gctgttacct ggaagtccac agtggccccg
139591 gcgggcaggc gggcgggcag agcgcgggtc cgaacgccct tcgcgctcgc ggcggcgcac gtgcggcggc ggcggcggcg
139671 ccggcggcgc gggcgggacc cactgggtgg ccgcgcgcgc cacccgcgcg ttttctgcgc gcctctgccc acccgcgagg
139751 ccgcaacgcg ccctgacacc cgacctcagt gtggaccctc acttgctgtg ggtgttgggg cctctctaga gcagggagga

146881 gaggtctcgc tatgttgtcc aggttgggaa aaaaaaccca caaaactatt aactcacacc tactcccatt cccttgcccc
146961 ttaaatgccc agtgtcctcc atgttgtgac ctggctattc caaatctcag tgMactctcc tttccttgtt gcgagtagga
147041 aattttacat taaaggacat gtatgatgat aataatgcat ataagactga gaatgaagat gaatagttgt gaaaagtatc

131101 aaatgccatc agtctgggac gtgatcgggc agaggtgtac tcacagtagt gtaaatactg ctgtaaatag ttgtctgatg
131181 gtggctttga cagtgagcta gcttctgagt tttcccttct ttttatactg ttttctgtgc tggcttttt gaatcttcct
131261 aatttttcat ctctttaaca aactcctatg aagttgaaac cgggaagttt gctctaacat ttccagagaa ggtacgttac
131341 tttttgctaa gagaatatct gtaagagttt agagaggggt gggattttac gattgcttta ctggtagatt tggcaattgc
131421 tttacttttt tttttttctg agactgtagc tcattgaaat gaactagagc attgtatctg tttgcgaggg aaaagccggg
```

Putative CpG
%GC=0.68

Putative CpG
%GC=0.67

SNP 923875

Exon 1

```
 5641 tgactattca aggctttttt cttccccctc ttctaaagat gctgtctctg taattggcag agagggacat cttgataatg
 5721 gaagtgtgaa ggtgtaacgt gtgttaattg atacttctta atcactttta ggtattaagt catgatgcag gaatctgcga
 5801 cagagacaat aagcaacagt tcaatgaatc aaaatggaat gagcactcta agcagccaat tagatgctgg cagcagagat
 5881 ggaagatcaa gtggtgacac cagctctgaa gtaagcacag tagaactgct gcatctgcaa caacagcagg taagttttgt
 5961 tttcctcagt gcttccttaa aacaaataag cttgttttat tgaatatgaa aaatgtattc atagcaaact gagcatgttg
 6041 tgttaagcta aaagatgctg agaatttatt attatttgga acatgattga aggatgagta aagagataag attttcatat

60921 ttattgaagg taggatcaag tcacctctag gaaattcccc cagagtattt aattccctgg ctcaagtaaa tgttactatt
61001 actgttcgtt gcagatgatc tcatatggaa actctatcta tctatctatc tatctatcta tctatctatc tatctatcta
61081 tatatctatc tatctaatct ataatggtta gaagttgtaa cccggtaaat agaaaggcct gatttatagc actatcagcc
61161 cccttatctg agacattttc tgttgacatt gatgtataat caggaaaatg ttccagacct gctttaaaaa gaggggggca
61241 ctatggcttg tgttctgatt ttaacttaat caccattgat tactcttctt tcttagttac tattcctcct agttggcagg
61321 gtcccagaac tacatgcagt actctagaaa ttaaatccta ttagtactca atataacagg aggatcgaga gagagggatt

93221 atttatgaca ttgtggccat accccagtac aaattttatc actaacattt tttaaattga taagttaaca ttgtacatat
93301 gtgtatatct atatatcact aacattttta aaattgatct taacattgta tcttatacac acacacacac acacacacac
93381 acacacacgc cataaatata tacaatacta tgctgtatgt atacaaagtt atattgtata tatacatgct gttttcatat

113751 gataaacaac aatgattctc tctagaaagg aatatgggag ttcttgtaca ttgaagcctt ttactattaa ccaagagata
113831 attgataact taactttcaa caaaatattt agataaggtt tcatttttac ttctaggctc tccaggcagc aagacaactt
113911 cttttacagc agcaaacaag tggattgaaa tctcctaaga gcagtgataa acagagacca ctgcaggtta gtaaagcact
113991 cctgtccttg gggtcttatt ttaaaagatg ttcataatga taacagatgt gcagacccac ttgtatatat acaggaaatt

117401 aaagagtaag ttaacacatc accccacaca cacataaatg ccaacatgca catacacaca tgcacacttc cataatctgt
117481 aaatttaaaa aaaWttttgt aaaatacatt ttcaattttc tagcctatga caaatctact taatttagtt tagatatgaa
117561 aattggtcgc attattgttc ttgttacatt tacagccatg agacagactt gtggttcttc aacagctgtg cagagcagag
117641 acgtaaattt aaggtaaatg ccattgcttt gctatcttag atgaatatta gtttttagca ttgtaatcac ttccttcatt
117721 gtaggttaat atataatata aagttctcct ttctaatagg taattgaatt ataaaatttt tataattctg tcttaagaag
```

Exon 2

Microsatellite 1
(TAGA)n

Microsatellite 2
(CA)n

Exon 3

Intron 3 SNP

Exon 3a

```
149901 tatttttaaa atctatactg gttaactgtg agacaatgat ggaatattct tccctatttt gacctaccaa gataataatt
149981 ccaaaaaatg atttttaaaa atgataatgt acgttattag cacaagccYt aagattgaaa aaaaRgccaa ctcctgattc
150061 tctttgttta acctaggaat tgcttccaga aacaaaatta tgtatctgtg gccactcttc tggtgatggg catcctcaca
150141 acacatttgc agtaagttac actgggcaat tagaaattca gctactaggc acaaggccat ggacccagtc ccactaaata
150221 gattatatgt gattttctat agccaatccc ttccattggt gcagataatc aagatacaaa taatgttaag atacagatgt

160281 tagcagtaga gaaccaggag ccagcagtca aaggccaata gaagatacac tcacataaac acatgcacac acacacacac
160361 acacacacac acacacaagc acacggacag attggatttg ttgcttctat cttcttccgt ttatttattt atttatttat
160441 ttatttattt atttatttat ttatttattt tgagatgcag tctggctctg tcacccaggc tggagtgcag tggcacagtc
160521 tctgctcact gcaacctctg cctcctaggc ttaagcgatc ctccctcctc agcctcccca gtagttggca ccacaggcac
160601 aagccaccat atctgactaa ttattgtatt ttttgtagag atggggtttc accattgccc aggctggtct tgaactcctg
160681 gccttacaga cgtgagccac tgtgcctggc catcttcttt ataattaaca ctttttgtg tgtcagatta gttttaggtt

 31801 gctaaagaat ttataaaaga taacatacta tttgtgaagt tgattaacag atatttggtt atgaccacga attttctttc
 31881 tctttctctt tctgtgcaag gtgcctgtgt cagtggccat gatgactccc caggtgatca cccctcagca aatgcagcag
 31961 atccttcagc aacaagtcct gtctcctcag cagctacaag cccttctcca acaacagcag gctgtcatgc tgcagcaggt
 32041 aatgtgggtt acctgctttg gtgttctagc atgacttaga aggtgtgcac ttattttgaa agtgcagtgg gcatatttga
 32121 caattcagtc tccatttgaa aggacaacct attagcgtgc tagaacataa atgtaacatt ttaattatag aactcgtaag

 33001 aatgtatgta gagctgtctc tttgaatcca atgtatattt tttgttgttt ttatgggatg aatcttaatg gatactctgc
 33081 catatgccag tctagaagag tttaggagat ttataatacg tgaaactttt gcctttattt attaaagtca aaatggttta
 33161 ttcagcaaca actacaagag ttttacaaga aacagcaaga gcagttacat cttcagcttt tgcagcagca gcagcaacag
 33241 cagcagcagc aacaacagca gcaacaacag cagcagcaac aacaacaaca acagcagcaa caacagcagc agcagcagca
 33321 acagcagcag cagcagcaac agcatcctgg aaagcaagcg aaagaggtag gatccggtta tcKcattgat acataacagt
 33401 ttgcagttaa gaaggggctt tgagcggcaa gaatagtctt agatcttcct ataacaaggc tcttgctgtc aaagttgcag

 34801 atgaaaggag tgtgcatttc cctgtaagag agctgtttgt acagaccatg ttctctgctg tttactggtt tgggttttct
 34881 gataccagca gcagcagcag cagcagcagc aacagcaatt ggcagcccag cagcttgtct tccagcagca gcttctccag
 34961 atgcaacaac tccagcagca gcagcatctg ctcagccttc agcgtcaggg actcatctcc attccacctg gccaggcagc
 35041 acttcctgtc caatcgctgc ctcaaggtac atacaaaatg ttgtgcactc ttcatttcaa atcttgtact ttctaccatt
 35121 tcatggcctt tctgccttat accttgagaa attttgtttt tatgactttc aaatttatta ttaaaacgtg attgttaaca
```

Intron 3a SNPs
Exon 3b

Microsatellite 3
(CA)n

Exon 4

Exon 5

Intron 5 SNP

Exon 6

```
45601 aagacaaaag ctttggtgct gatagtgtaa aagtgatttt ttgttgtata ttaatttatg caggtaacat cactgttgtt
45681 gttggtatta atctattaat aacacaaagc tcattatata aactttacct ttaccttgtt atgctagtga agctttcttt
45761 cattttatag ctggcttaag tcctgctgag attcagcagt tatggaaaga agtgactgga gttcacagta tggaagacaa
45841 tggcattaaa catggagggc tagacctcac tactaacaat tcctcctcga ctacctcctc caacacttcc aaagcatcac
45921 caccaataac tcatcattcc atagtgaatg gacagtcttc agttctaagt gcaagacgag acaggtaaat ctcatgagct
46001 ttattctata tttatctatt ttcagatttt attttcacca ttgagacaat gaaaaagaac aatttgtact tggagcaagg

47941 cctgttttga cctgtttgtc actgatcgta acctgacagg cgctagcagc ctgcaacgat tatagctttt tgagatgaat
48021 ctgacgtcgt gttcttttgc tacagctcgt cacatgagga gactggggcc tctcacactc tctatggcca tggagtttgc
48101 aaatggccag gctgtgaaag catttgtgaa gattttggac agtttttaaa gtaggttttt tacttttttt tggtgggggg
48181 cgggggctgg attatatggg cattttaggc acattgtaaa taaacaaaag atgaaggagg aatgtaaaag tccattataa

55441 aaggctttct gttttcccag tgcacagaaa catctgactt cagtgtagtg ctttttaagt gtagcctatg ccactaagat
55521 cgacatcact ttacattctg ttttgtgtct tctgtttgtt taggcacctt aacaatgaac acgcattgga tgaccgaagc
55601 actgctcagt gtcgagtgca aatgcaggtg gtgcaacagt tagaaataca ggtttgttaa atgctcactt aatggactct
55681 tcttagattt ctggtgtgtt acattgagta ctgagcagat tctgggtctt tcagcctgca ttttactgtc ttagccattc

57181 aggaatcatt attctgaggc aagctcaatg ataagatgta tcactgcaat aaaatagctg tatcagtcat ttctaaaacg
57261 cttctgatct cactctttct taacagcttt ctaaagaacg cgaacgtctt caagcaatga tgacccactt gcacatgcga
57341 ccctcagagc ccaaaccatc tcccaaacct gtaagtgcat attgctttat aaacagtaaa tagctctacc aatgtaacag
57421 actaagaaaa tgaacaattt agtgacagtt agaaaacaat gagtgtgatg aaaaatacgg caataaaatg aaagtaaaat
57501 gtaatcgctt gtcaaattgt atgtttcttt taaagtaatg ctatctttta caaaatctat ccaatctaca ccatgggtga

61261 tatgatacaa tttcatcctg taattagttg agattggctg cttccttttg cagcttatta gtggcaacac agctgtagaa
61341 gtgaatccac tctcatttgt caaacctttt taagtctttt tctcttgtct ctacctttct cctgcccttc tcttgggcct
61421 ttgcagctaa atctggtgtc tagtgtcacc atgtcgaaga atatgttgga gacatcccca cagagcttac ctcaaacccc
61501 taccacacca acggccccag tcaccccgat tacccaggga ccctcagtaa tcaccccagc cagtgtgccc aatgtgggag
61581 ccatacgaag gcgacattca gacaaataca acattcccat gtcatcaggt aggataYgaa tgctcagtag agcactttta
61661 ctttgggaga ggaaaactgt agctgaatca actgtacatg agccattcca gagcacatgg aaactcatgt catgatttat
```

Exon 7

Exon 8

Exon 9

Exon 10

Exon 11

Intron 11 SNP (1)

```
62581 ctgtcatgct ttgtgctttc actagtcggt ggcttcctca ctgaatcact ttaccaRtac tagaggaaat atgaaaattc
62661 aattatatat aatgtgaatt attagcagaa ttaacaccta gtttttattt ttatagaaat tgccccaaac tatgaatttt
62741 ataaaaatgc agatgtcaga cctccattta cttatgcaac tctcataagg caggtaagta gaaggaaaat taactttgcc
62821 tgattaaatt aaaattcatt aatgcttaaa gtaaggcttg gatgagaaag tgtcatctag tctagttagt aaaccattat
62901 tttatgtcac tatgtatctt gtctcatttc aggctatcat ggagtcatct gacaggcagt taacacttaa tgaaatttac
62981 agctggttta cacggacatt tgcttacttc aggcgtaatg cagcaacttg gaaggtaact acttttccag cagttttaag
63061 atgSctacca cagttcctta cagatagcac aaggaacatt tatttacatg acataagcag attagtatct gcttcccctc

65341 agtaagatca ataatgtagt atgttgggct gccttattag acaatattat ttttgccatt ttttcttctc ttctgtctgc
65421 tttagaatgc agtacRtcat aatcttagcc tgcacaagtg ttttgttcga gtagaaaatg ttaaaggagc agtatggact
65501 gtggatgaag tagaatacca gaagcgaagg tcacaaaaga taacagggta tgtttgtgat agttttgtaa Yctgtatcc
65581 tgcatccacc aggaaaagta aatactttaa gttggggctg gggtggggag acagttagct taatggcatg ctaacattct
65661 cgtggcaatg aactgcattt tgaatctagg tgtaggtggc ataaatcaac agtagcctat caatttttc cgtaaaaatc

66661 actctattac cttatttttg tggccttata tacatccagt atggactatt agtgtgagac aagccagaac atacctttat
66741 agctgagact atcgatatcc acgttttata ttttgacgta taaatgatct ttatatattt ttttttcag aagtccaacc
66821 ttagtaaaaa atatacctac cagtttaggc tatggagcag ctcttaatgc cagtttgcag gtaatgtact ttcccagttt
66901 tgttgtattt gaatgtttag ggctttttt ttttttttgg catgtctttg tatttaggtg agattgtgat tgttcttaat

67441 tttcacatcg cacttaattt ccttgcatct ctgccacaag tagccagtta ggaatttttt ttcaatacat tttcttttaa
67521 tccttctaaa atatctattt ccttattgga acccattaaa agaagataca tgttttaaaa attattttaa atgccatttt
67601 gaaagttgtt ttacacaatc ttcatttcac taggctgcct tggcagagag cagtttacct ttgctaagta atcctggact
67681 gataaataat gcatccagtg gcctactgca ggccgtccac gaagacctca atggttctct ggatcacatt gacagcaatg
67761 gaaacagtag tccgggctgc tcacctcagc cgcacatgta agtgtggtta acagactctc taaagggaag aatctatatt
67841 aatatgcttc ttaaatttag aatttttccg aagtttttac tgttgtataa gttgaaaata caaatttaag tggagttaaa

91401 agtatggatg aaactgcaaa agccatttgt gaagtaaaaa atgagataaa acttttttcat aaacttcagg gttttttttt
91481 ttttaatgtg cttgtccttt tcctttcttt taatttttcc ttttattctt ctctcctaca aacacacaca aacacatgca
91561 tacatgcata cacacacaca cacacacaca cacacacaca cacacacatc ttttttttt tttggcctag ttttgcattt
91641 caacatgcct ggttcaaact agctaaaaaa cgtataattt actaacacca aacaacaaaa acaaaggcta aagcaaaaca
```

Intron 11 SNP (2)
Exon 12

Exon 13
Intron 13 SNP

Exon 14
Intron 14 SNP

Exon 15

Exon 16

Microsatellite 4
(CA)n

```
93061 gttgacctct tcactgcaaa gttggccaaa ctctgtttgt tgcttactta gtaaaatttt ggtgtatcta catgtttttc
93141 agacattcaa tccacgtcaa ggaagagcca gtgattgcag aggatgaaga ctgcccaatg tccttagtga caacagctaa
93221 tcacagtcca gaattagaag acgacagaga gattgaagaa gagcctttat ctgaagatct ggaatgagaa ctgacttgtg
93301 aaacctcagc gtgaagggac atatcactga ccttcataac cactccacaa ccatgaatat ttgacaaatt tttactgtga
```

Exon 17

KEY:

Coding sequence is shown in red
5' and 3' untranslated regions are shown in pink
Non-coding sequence is shown in black
SNPs and microsatellites shown in blue
Putative CpG islands are highlighted in grey

The KE mutation is boxed in red
The polyglutamine tracts are boxed in black
Primers for the amplification of all regions are underlined
(NB all primer sequences can also be found in appendix P)

atg|atg|cag|gaa|tct|gcg|aca|gag|aca|ata|agc|aac|agt|tca|atg|aat|caa|aat|gga|atg|agc|act|cta|agc|agc|caa|tta|gat|gct|

M M Q E S A T E T I S N S S M N Q N G M S T L S S Q L D A

ggc|agc|aga|gat|gga|aga|tca|agt|ggt|gac|acc|agc|tct|gaa|gta|agc|aca|gta|gaa|ctg|ctg|cat|ctg|caa|caa|cag|cag|*gct|ctc|*

G S R D G R S S G D T S S E V S T V E L L H L Q Q Q Q A *L*

*cag|gca|gca|aga|caa|ctt|ctt|tta|cag|cag|caa|aca|agt|gga|ttg|aaa|tct|cct|aag|agc|agt|gat|aaa|cag|aga|cca|ctg|cag|*gtg|

Q A A R Q L L L Q Q Q T S G L K S P K S S D K Q R P L Q V

cct|gtg|tca|gtg|gcc|atg|atg|act|ccc|cag|gtg|atc|acc|cct|cag|caa|atg|cag|cag|atc|ctt|cag|caa|caa|gtc|ctg|tct|cct|cag|

P V S V A M M T P Q V I T P Q Q M Q Q I L Q Q Q V L S P Q

cag|cta|caa|gcc|ctt|ctc|caa|caa|cag|cag|gct|gtc|atg|ctg|cag|cag|*caa|caa|cta|caa|gag|ttt|tac|aag|aaa|cag|caa|gag|cag|*

Q L Q A L L Q Q Q Q A V M L Q Q *Q Q L Q E F Y K K Q Q E Q*

*tta|cat|ctt|cag|ctt|ttg|*cag|cag|cag|cag|caa|cag|cag|cag|cag|caa|caa|cag|cag|caa|caa|cag|cag|cag|caa|caa|caa|caa|caa|

L H L Q L L Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q

cag|cag|caa|caa|cag|cag|cag|cag|cag|caa|cag|cag|cag|cag|cag|caa|cag|*cat|cct|gga|aag|caa|gcg|aaa|gag|*cag|cag|cag|cag|

Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q *H P G K Q A K E* Q Q Q Q

cag|cag|cag|caa|cag|caa|ttg|gca|gcc|cag|cag|ctt|gtc|ttc|cag|cag|cag|ctt|ctc|cag|atg|caa|caa|ctc|cag|cag|cag|cag|cat

Q Q Q Q Q Q L A A Q Q L V F Q Q Q L L Q M Q Q L Q Q Q Q H

ctg|ctc|agc|ctt|cag|cgt|cag|gga|ctc|atc|tcc|att|cca|cct|ggc|cag|gca|gca|ctt|cct|gtc|caa|tcg|ctg|cct|caa|*gct|ggc|tta*

L L S L Q R Q G L I S I P P G Q A A L P V Q S L P Q *A G L*

agt|cct|gct|gag|att|cag|cag|tta|tgg|aaa|gaa|gtg|act|gga|gtt|cac|agt|atg|gaa|gac|aat|ggc|att|aaa|cat|gga|ggg|cta|gac

S P A E I Q Q L W K E V T G V H S M E D N G I K H G G L D

ctc|act|act|aac|aat|tcc|tcc|tcg|act|acc|tcc|tcc|aac|act|tcc|aaa|gca|tca|cca|cca|ata|act|cat|cat|tcc|ata|gtg|aat|gga

L T T N N S S S T T S S N T S K A S P P I T H H S I V N G

cag|tct|tca|gtt|cta|agt|gca|aga|cga|gac|agc|tcg|tca|cat|gag|gag|act|ggg|gcc|tct|cac|act|ctc|tat|ggc|cat|gga|gtt|tgc

Q S S V L S A R R D S S S H E E T G A S H T L Y G H G V C

aaa|tgg|cca|ggc|tgt|gaa|agc|att|tgt|gaa|gat|ttt|gga|cag|ttt|tta|aag|cac|ctt|aac|aat|gaa|cac|gca|ttg|gat|gac|cga|agc

K W P G C E S I C E D F G Q F L K H L N N E H A L D D R S

act|gct|cag|tgt|cga|gtg|caa|atg|cag|gtg|gtg|caa|cag|tta|gaa|ata|cag|ctt|tct|aaa|gaa|cgc|gaa|cgt|ctt|caa|gca|atg|atg

T A Q C R V Q M Q V V Q Q L E I Q L S K E R E R L Q A M M

acc|cac|ttg|cac|atg|cga|ccc|tca|gag|ccc|aaa|cca|tct|ccc|aaa|cct|cta|aat|ctg|gtg|tct|agt|gtc|acc|atg|tcg|aag|aat|atg

T H L H M R P S E P K P S P K P L N L V S S V T M S K N M

ttg|gag|aca|tcc|cca|cag|agc|tta|cct|caa|acc|cct|acc|aca|cca|acg|gcc|cca|gtc|acc|ccg|att|acc|cag|gga|ccc|tca|gta|atc

L E T S P Q S L P Q T P T T P T A P V T P I T Q G P S V I

acc|cca|gcc|agt|gtg|ccc|aat|gtg|gga|gcc|ata|cga|agg|cga|cat|tca|gac|aaa|tac|aac|att|ccc|atg|tca|tca|gaa|att|gcc|cca

T P A S V P N V G A I R R R H S D K Y N I P M S S E I A P

aac|tat|gaa|ttt|tat|aaa|aat|gca|gat|gtc|aga|**cct**|**cca**|ttt|act|**tat**|gca|act|**ctc**|ata|agg|cag|**gct**|atc|atg|gag|tca|tct|gac

N Y E F Y K N A D V R **P** **P** F T **Y** A T **L** I R Q **A** I M E S S D

agg|cag|tta|aca|**ctt**|aat|gaa|**att**|tac|agc|tgg|ttt|aca|cgg|aca|ttt|gct|tac|ttc|agg|cgt|aat|gca|gca|act|**tgg**|aag|**aat**|gca

R Q L T **L** N E **I** Y S W F T R T F A Y F R R N A A T **W** K **N** A

gta|**cgt**|**cat**|aat|**ctt**|agc|ctg|cac|aag|tgt|**ttt**|gtt|cga|gta|gaa|aat|gtt|**aaa**|**gga**|gca|gta|**tgg**|act|gtg|gat|gaa|gta|gaa|tac

V **R** **H** N **L** S L H K C **F** V R V E N V **K** **G** A V **W** T V D E V E Y

cag|aag|cga|agg|tca|caa|aag|ata|aca|gga|agt|cca|acc|tta|gta|aaa|aat|ata|cct|acc|agt|tta|ggc|tat|gga|gca|gct|ctt|aat

Q K R R S Q K I T G S P T L V K N I P T S L G Y G A A L N

gcc|agt|ttg|cag|gct|gcc|ttg|gca|gag|agc|agt|tta|cct|ttg|cta|agt|aat|cct|gga|ctg|ata|aat|aat|gca|tcc|agt|ggc|cta|ctg

A S L Q A A L A E S S L P L L S N P G L I N N A S S G L L

cag|gcc|gtc|cac|gaa|gac|ctc|aat|ggt|tct|ctg|gat|cac|att|gac|agc|aat|gga|aac|agt|agt|ccg|ggc|tgc|tca|cct|cag|ccg|cac

Q A V H E D L N G S L D H I D S N G N S S P G C S P Q P H

ata|*cat*|*tca*|*atc*|*cac*|*gtc*|*aag*|*gaa*|*gag*|*cca*|*gtg*|*att*|*gca*|*gag*|*gat*|*gaa*|*gac*|*tgc*|*cca*|*atg*|*tcc*|*tta*|*gtg*|*aca*|*aca*|*gct*|*aat*|*cac*|*agt*

I *H S I H V K E E P V I A E D E D C P M S L V T T A N H S*

cca|*gaa*|*tta*|*gaa*|*gac*|*gac*|*aga*|*gag*|*att*|*gaa*|*gaa*|*gag*|*cct*|*tta*|*tct*|*gaa*|*gat*|*ctg*|*gaa*|

P E L E D D R E I E E E P L S E D L E

KEY: Alternate exons are shown in normal text or *italics* accordingly.
The polyglutamine stretches are shown in blue and the FOX domain is shown in red
Conserved residues of the FOX domain are shown in **bold**
The mutation found in the KE family is boxed in red.

Region	Forward Primer Sequence	Reverse Primer Sequence	PCR Conditions[a]
Exons			
Exon 2	GAGAGGGACATCTTGATAATG	TAGCTTAACACAACATGCTCAG	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 3	GAAAGGAATATGGGAGTTCTTG	GTGGGTCTGCACATCTGTTATC	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 3a	GCACATACACACATGCACACTTC	ACAATGAAGGAAGTGATTACAATGC	2.0 mM WAVE 58-50.5°cx15 50.5°c x25
Exon 3b	CTTCCCTATTTTGACCTACCAAG	CATATAATCTATTTAGTGGGACTGG	2.5 mM WAVE 52.5-45°cx15 45°c x25
Exon 4	GATAACATACTATTTGTGAAGTTG	TCTAGCACGCTAATAGGTTGTCC	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 5	TGAATCTTAATGGATACTCTGCC	TCTAAGACTATTCTTGCCGCTC	2.0 mM WAVE 54.5-47°cx15 47°c x25
Exon 6	AAGGAGTGTGCATTTCCCTG	CAGAAAGGCCATGAAATGGTAG	2.0 mM WAVE 62.5-55°cx15 55°c x25
Exon 7	AATTTATGCAGGTAACATCACTG	CTTTTTCATTGTCTCAATGGTG	C38 2mM Ta=55°c
Exon 8	TGTTTGTCACTGATCGTAACCTG	GTGCCTAAAATGCCCATATAATCC	2.0 mM WAVE 54.5-47°cx15 47°c x25
Exon 9	GCTTTTTAAGTGTAGCCTATGCC	CAGAATCTGCTCAGTACTCAATG	2.0 mM WAVE 54.5-47°cx15 47°c x25
Exon 10	GAGGCAAGCTCAATGATAAGATG	CACCCATGGTGTAGATTGGATAG	2.5 mM WAVE 62.5-55°cx15 55°c x25
Exon 11	AATTAGTTGAGATTGGCTGCTTC	TTGATTCAGCTACAGTTTTCCTC	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 12	CTTCCTCACTGAATCACTTTACC	CACTTTCTCATCCAAGCCTTACT	3.0 mM WAVE 54.5-47°cx15 47°c x25
Exon 13	AGGCTTGGATGAGAAAGTGTCAT	CCTTGTGCTATCTGTAAGGAAC	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 14	TCAATAATGTAGTATGTTGGGCTG	CATTGCCACGAGAATGTTAGC	3.0 mM WAVE 59.5-52°cx15 52°c x25
Exon 15	GTGTGAGACAAGGCCAGAACATAC	CCTAAACATTCAAATACAAC	2.5 mM WAVE 55-47.5°cx15 47.5°c x25
Exon 16	CTGCCACAAGTAGCCAGTTAGG	ATACAACAGTAAAAACTTCG	2.5 mM WAVE 54.5-47°cx15 47°c x25
Exon 17	TGACCTCTTCACTGCAAAGTTGG	GTCAAATATTCATGGTTGTGGAG	2.5 mM WAVE 52.5-45°cx15 45°c x25
Microsatellites			
Microsatellite 1 (TAGA)n	AAGTCACCTCTAGGAAATTCCC	CCCTCTCTCTCGATGGTCCT	2.5 mM WAVE 60-52.5°cx15 55°c x25
Microsatellite 2 (CA)n	TGTGGCCATACCCCAGTAC	TGTATACATACAGCATAGTATTG	2.5 mM WAVE 55-47.5°cx15 50°c x25
Microsatellite 3 (CA)n	AGAACCAGGAGCCAGCAGT	TAAAGAAGATGGCCAGGCAC	2.5 mM WAVE 65-57.5°cx15 45°c x25
Microsatellite 4 (CA)n	TGCAAAAGCCATTTGTGAAG	CAGGCATGTTGAAATGCAAA	2.5 mM WAVE 60-52.5°cx15 55°c x25
SNPs			
intron 3a SNP	See exon 3b		
SNP 923875	CTCACCTACTCCCATTCC	CTATTCATCTTCATTCTCAG	C38 2mM Ta=48°c

a : For PCR protocols please see section 2.4 'The polymerase chain reaction (PCR)'

SNP	Enzyme	Recognition Sequence	Digestion Buffer	+ BSA?	Digestion Temp	Polymorphism	Size of Undigested Fragment	Size of Digested Fragments
NUMB-R								
Exon 10c SNP	*Ava*I	C|YCGRG	NEB buffer 4	No	37°c	Y	404bp (T)	238bp + 166bp (C)
SNP 2604887	*Bts*I	GCAGTGNN|N	NEB buffer 4	Yes	37°c	Y	271bp (C)	77bp + 194bp (T)
SNP 2561537	*Hae*II	RGCGC|Y	NEB buffer 4	Yes	37°c	Y	434bp (T)	289bp + 145bp (C)
SNP 2604892	*Bst*XI	CCANNNNN|NTGG	NEB buffer 3	No	55°c	S	323bp (C)	173bp + 150bp (G)
SNP 2561551	*Ava*I	C|YCGRG	NEB buffer 4	No	37°c	R	422bp (A)	156bp + 266bp (G)
FOXP2								
SNP 923875	*Apa*LI	G|TGCAC	NEB buffer 4	Yes	37°c	M	174bp (A)	78bp + 82bp (C)
Intron 3a SNP	*Afl*II	C|TTAAG	NEB buffer 2	Yes	37°c	Y	212bp (C)	57bp + 155bp (T)

Gene Symbol	Gene Name	Function	Disorder	Expressed in Brain?
ZNF1A2	Zinc finger protein, subfamily 1A, 2 (Helios)	DNA binding protein	-	No
BARD1	BRCA1 associated RING domain 1	May be involved in mediating tumour suppression by the binding of BRCA1.	Defects in BARD1 are found in primary breast, ovarian and uterine cancers	No
ABCA12	ATP-binding cassette, sub-familyA (ABC1), member 12	?	-	?
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase / IMP cyclohydrolase	Catalyses last two steps of the synthesis of inosine 5'-monophosphate	-	Yes
FN1	Fibronectin1	Glycoprotein involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defence and metastasais	fibronectin-deficient mice show increased neuronal apoptosis following cerebral ischemia	Yes
HSA250303	Peroxisomal trans 2-enoyl CoA reductase; putative short chain alcohol dehydrogenase	Enzyme involved in peroxisomal chain elongation pathway	-	?
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5	ATP-dependent helicase II – involved in repair of double stranded DNA breaks	May be involved in Systemic Lupus Erythematosus (SLE)	Yes
SMARCAL1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1	Helicase - Regulates transcription by the alteration of chromatin structure	Schimke Immunosseous dysplasia	Yes
RPL37A	Ribosomal protein L37a	Component of 60S subunit of ribosomes	-	Yes
IGFBP2	Insulin-like growth factor binding protein 2	Binds to and modulates insulin-like growth factor activity	-	Yes
IGFBP5	Insulin-like growth factor binding protein 5	Binds to and modulates insulin-like growth factor activity	-	Yes

Gene Symbol	Gene Name	Function	Disorder	Expressed in Brain?
TNP1	Transisition protein 1	Involved in chromatin condensation during spermiogenesis	-	Yes
TNS	Tensin	Transmembrane protein localises to regions of the plasma membrane where the cell attaches to the extracellular matrix	-	Yes
IL8RA	Interleukin 8 receptor, alpha	G-coupled protein receptor. Mediates neutrophil chemotaxis and binds interleukin 8 (IL8)	-	Yes
ARPC2	Actin related protein 2/3 complex subunit 2	Involved in assembly of actin cytoskeleton	-	Yes
AAMP	Angio-associated, migratory cell protein	Mediates heparin sensitive cell adhesion, may act in cell migration	-	Yes
SLC11A1	Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1	May control antimicrobial activity of macrophages.	May be involved in resistance to TB	Yes
NLI-IF	Nuclear LIM interactor-interacting factor	?	-	?
VIL1	Villin-1	Calcium regulated actin-binding protein	-	Yes
RQCD1	RCD1 required for cell differentiation1 homologue	?	-	Yes
ZNF142	Zinc-finger protein 142	?	-	?
BSC1L	BCS1-like	Mitochondrial protein. May function in the assembly of complex III of the respiratory chain;	-	Yes
RNF25	Ring finger protein 25	?	-	Yes
CYP27A1	Cytochrome P450, subfamily XXVIIA (steroid 27-hydroxylase, cerebrotendinous xanthomatosis), polypeptide 1	Monooxygenase which oxidises cholesterol intermediates as part of the bile synthesis pathway	Cerebrotendinous xanthomatosis	Yes
PRKAG3	Protein kinase, AMP-activated, gamma 3 non-catalytic subunit	Gamma 3 subunit of AMP-activated protein kinase; may directly affect interaction of AMP with kinase	-	No

Gene Symbol	Gene Name	Function	Disorder	Expressed in Brain?
WNT6	Wingless-type MMTV integration site family, member 6	Secreted signalling protein implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis.	Over expressed in cervical cancer lines	Yes
WNT10a	Wingless-type MMTV integration site family, member 10	Secreted signalling protein implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis.	Over expressed in cervical cancer lines	No
CDK5R2	Cyclin-dependent kinase 5, regulatory subunit 2 (p39)	Neuron-specific activator of CDK5 kinase	-	Yes
HSRNAFEV	FEV protein	Transcription factor; possibly a repressor	-	?
CYBA2	Crystallin betaA2	Major protein of vertebrate eyes	-	Yes
ABCB6	ATP-binding cassette sub family B member 6	Protein transporter involved in multidrug resistance and mitochondrial function	candidate gene for lethal neonatal metabolic syndrome, a disorder of mitochondrial function.	Yes
STK16	Serine/threonine kinase 16	Kinase	-	Yes
TUBA1	Tubulin alpha1	Component of microtubules	-	Yes
HSJ1	DnaJ (Hsp40) homologue, subfamily B, member 2	Neuronal member of the DnaJ family of proteins	-	Yes
PTPRN	Protein tyrosine phosphatase, receptor type, N	An autoantigen that is reactive with insulin-dependent diabetes mellitus (IDDM) patient sera	Diabetes mellitus	Yes
DNPEP	Aspartyl aminopeptidase	Member of the M18 family of metalloproteases - Probably functions in intracellular protein and peptide metabolism	-	Yes

Gene Symbol	Gene Name	Function	Disorder	Expressed in Brain?
DES	Desmin	Muscle-specific class III intermediate filament involved in the connection of myofibrils to each other and to the plasma membrane.	Desmin-related myopathy, a familial cardiac and skeletal myopathy (CSM), and with distal myopathies	Yes
APEG1	Nuclear protein, marker for differentiated aortic smooth muscle and down-regulated with vascular injury	?	-	?
GMPPA	GDP-mannose pyrophosphorylase A	Catalyses the conversion of mannose-1-phosphate and GTP to GDP-mannose which is involved in the production of N-linked oligosaccharides.	-	?
ASIC4	Putative acid-sensing ion channel	Acid-sensing ion channel	-	?
INHA	Inhibin, alpha	Joins either the beta A or beta B subunit to form a pituitary FSH secretion inhibitor	-	Yes
SLC4A3	Solute carrier family 4, anion exchanger, member 3	Anion exchanger; may act to regulate intracellular pH and chloride concentration		Yes
EPHA4	Eph-related receptor tyrosine kinase A4	?	-	Yes

Three SNPs were chosen from across the RP11-105N14 clone for association analyses. This appendix shows the fragments amplified for the typing of those SNPs. The numbers given refer to base positions within RP11-105N14.

```
   4801 tgccgtcatt atttttatac aatagctggc agcaagttgc acagaaaccc ttcatcttag aagagttacc aaatagctta
   4881 aacaaatgct gggaccttac agtagtactg accaccttcc taaagccaaa ataagtgagc tctcatRcct tattgatact
   4961 tttctggcca gaatgttagc tgttttacta aaaaaaaaaa aaaaatttac ttattttatt tttttaactg tttaaaaata
   5041 actttttttcc tccttacaca tatattccca tcctctcctt cctgctacta gaaactgtct taaaggtttt cacatttact
   5121 tttagcttta acactttcaa tcaaaattgc aattctgttt ttggaagcag gcaccctcca ggaagtatat ttgaagacat

  28501 aggcagatgt tgactgggtt tacttgcatt cagaatggca gcaattggat atgtgatctt ggacctgagt atgaagccaa
  28581 tatcggtgat agaggtaaca aacacacgga attgtgagtc tcRagaaagg tcagaaatga gaggatcagc atgagaaaca
  28661 aggacaaacc atccaaatca agggaagaaa gcccaggtga tctatctaag gagaggtggg gcaggtgttg tgcaatggtc
  28741 aaaatggcag gcaaatgaaa gagatctaag cagtggggtg ttcagtgttg gaaagtccag ttgcaggaat cagggcccca
  28821 gtcaaggaaa attgagctca gaagtagagt atccattcac cgtggtagca agggacaaga caggaatcaa ttgacaacag

 152761 aagccagtct ctaaaataaa atgacaccct gatatggttt ggctctgtgt ccccacccta atctcacctt gaattgtaat
 152841 aatccgtatg tgtcaagggc aggaccttgt agaggtaatt ggatcatagg Kccagtttcc cccatgctgc tcttgtgata
 152921 atgagtgagt atcgcaagat ctcatggttt tgtaagtgtc tggcatttcc actgcttgca cccattctct ctttcatggt
 153001 gccctgtgaa gaagtgcctt ccgccatgat tgtaagtttc ctgaggtatc ctcagccatg cagaactatg agtcaattaa
 153081 acctcttttc cttataaatt acccagtctt gggtatttct tcatagcagc atgagaatga actaatacgc atccagtgag
```

KEY:

SNPs are shown in blue
Primers are underlined
All SNPs fall within non-coding sequence

SNP 1441165

SNP 1441169

SNP 1488984

SNP	Forward Primer Sequence	Reverse Primer Sequence	PCR Conditions[a]
SNP 1441165	TGTTGACTGGGTTTACTTGC	GGTGAATGGATACTCTACTT	2.0mM WAVE 55.5-48°cx15 50.5°cx25
SNP 1441169	TAGAAGAGTTACCAAATAGC	AAAGCTAAAAGTAAATGTGA	2.0mM WAVE 49-41.5°cx15 46°cx25
SNP 1488984	AATAAAATGACACCCTGATA	TGAAGAAATACCCAAGACTG	2.0mM WAVE 54-46.5°cx15 50°cx25

a : For PCR protocols please see section 2.4 'The polymerase chain reaction (PCR)

SNP	Enzyme	Recognition Sequence	Digestion Buffer	+ BSA?	Digestion Temp	Polymorphism	Size of Undigested Fragment	Size of Digested Fragments
SNP 1441165	*Eco*NI	CCTNN|NNNAGG	NEB buffer 4	No	37°c	R	272bp (A)	67bp + 205bp (G)
SNP 1441169	*Ava*I	C|YCGRG	NEB buffer 4	No	37°c	R	353bp (A)	113bp + 240bp (G)
SNP 1488984	*Ava*II	G|GWCC	NEB buffer 4	No	37°c	K	88bp + 261bp (G)	88bp, 27bp + 234bp (T)

The SLI Consortium (2002) A genomewide scan identifies two novel loci involved in Specific Language Impairment. The American Journal of Human Genetics, 70, 384-399

Newbury, DF, Bonora, E, Lamb, JA, Fisher, SE, Lai, CSL, Baird, G, Jannoun, L, Slonims, V, Stott, CM, Merricks, MJ, Bolton, PF, Bailey, AJ, Monaco, AP & the International Molecular Genetic Study of Autism Consortium (2002) *FOXP2* is not a major susceptibility gene for autism or specific language impairment. American Journal of Human Genetics, 70, 1318-1327

Am. J. Hum. Genet. 70:384–398, 2002

A Genomewide Scan Identifies Two Novel Loci Involved in Specific Language Impairment*

The SLI Consortium*

Approximately 4% of English-speaking children are affected by specific language impairment (SLI), a disorder in the development of language skills despite adequate opportunity and normal intelligence. Several studies have indicated the importance of genetic factors in SLI; a positive family history confers an increased risk of development, and concordance in monozygotic twins consistently exceeds that in dizygotic twins. However, like many behavioral traits, SLI is assumed to be genetically complex, with several loci contributing to the overall risk. We have compiled 98 families drawn from epidemiological and clinical populations, all with probands whose standard language scores fall ≥1.5 SD below the mean for their age. Systematic genomewide quantitative-trait–locus analysis of three language-related measures (i.e., the Clinical Evaluation of Language Fundamentals–Revised [CELF-R] receptive and expressive scales and the nonword repetition [NWR] test) yielded two regions, one on chromosome 16 and one on 19, that both had maximum LOD scores of 3.55. Simulations suggest that, of these two multipoint results, the NWR linkage to chromosome 16q is the most significant, with empirical P values reaching 10^{-5}, under both Haseman-Elston (HE) analysis (LOD score 3.55; $P = .00003$) and variance-components (VC) analysis (LOD score 2.57; $P = .00008$). Single-point analyses provided further support for involvement of this locus, with three markers, under the peak of linkage, yielding LOD scores >1.9. The 19q locus was linked to the CELF-R expressive-language score and exceeds the threshold for suggestive linkage under all types of analysis performed—multipoint HE analysis (LOD score 3.55; empirical $P = .00004$) and VC (LOD score 2.84; empirical $P = .00027$) and single-point HE analysis (LOD score 2.49) and VC (LOD score 2.22). Furthermore, both the clinical and epidemiological samples showed independent evidence of linkage on both chromosome 16q and chromosome 19q, indicating that these may represent universally important loci in SLI and, thus, general risk factors for language impairment.

Introduction

Specific language impairment (SLI) is diagnosed in children who exhibit significant language deficits despite adequate educational opportunity and normal nonverbal intelligence. A diagnosis is made after the presence of other conditions—such as mental retardation, autism, hearing loss, cleft palate, and neurological disorders (e.g., cerebral palsy) that may give rise to language impairments—has been ruled out (Tomblin et al. 1996). Children with SLI differ in the degree to which they have problems articulating speech sounds, expressing themselves verbally, and comprehending the speech of others. Accordingly, SLI is broadly classified into three subtypes: phonological disorder, expressive-language disorder, and mixed expressive- and receptive-language disorder. However, the validity of this subtyping has been questioned, and, instead, it has been proposed that the variability in the profile of deficits may reflect variation in the severity of the underlying disorder.

Although there have been many epidemiological studies of SLI, differences in methodological approaches, in diagnostic criteria, and in category thresholds often render direct comparisons between investigations impractical. The majority of mainstream studies estimate the prevalence among English-speaking pre–primary-school children to be 2%–7% (Law et al. 1998). A substantial proportion of these children are reported to experience severe and persistent language difficulties, which are often associated with additional social, educational, behavioral, and psychological problems (Cantwell and Baker 1987; Beitchman et al. 1994; Snowling et al. 2001). Despite the differences in study design, it is worth noting that most investigations agree on the importance of genetic factors in the development of SLI and that many have demonstrated a strong familial aggregation of cases of language impairment (Bishop and Edmundson 1986; Neils and Aram 1986; Tallal et al. 1989). In a recent review, Stromswold (1998) reported that, across seven family studies, the prevalence of SLI in family members of probands was 24%–78% (mean

Received September 12, 2001; accepted for publication November 12, 2001; electronically published January 4, 2002.

Address for correspondence and reprints: Dr. A. P. Monaco, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, England. E-mail: anthony.monaco@well.ox.ac.uk; or Dr. P. F. Bolton, University of Cambridge, Developmental Psychiatry Section, Douglas House, 18B Trumpington Road, Cambridge, CB2 2AH, England. E-mail: pfb1000@hermes.cam.ac.uk

* Members of the consortium are listed in the Appendix.

0002-9297/2002/7002-0012$15.00

46%), compared with 3%–46% (mean 18%) in the control groups.

In addition, twin studies consistently have indicated a significant increase in MZ concordance rates compared with DZ concordance rates (Lewis and Thompson 1992; Bishop et al. 1995; Tomblin and Buckwalter 1998), suggesting that much of the reported familial aggregation can be attributed to genetic influences. Tomblin and Buckwalter (1998) studied 120 twin pairs in which affected individuals were defined as having both a composite language score (computed from four measures of receptive language and four measures of expressive language) 1 SD below that expected for their ages and a nonverbal IQ (i.e., performance IQ [PIQ]) >70. Using this sample, they demonstrated an MZ concordance rate of 96% and a DZ concordance rate of 69%. Bishop et al. (1995) studied a set of 90 same-sex twin pairs, all with at least one twin affected by a developmental-speech or -language disorder. Using a strict definition of language impairment (i.e., a discrepancy of ≥20 points between verbal and nonverbal abilities), they found a male-male MZ concordance of 70% and a male-male DZ concordance of 46%. Relaxation of the diagnostic criteria to include those cotwins who either lacked a large discrepancy between their (low) verbal skills and nonverbal ability or had a history of speech and language problems resulted in heightened MZ:DZ concordance rates of 92%:62% for male-male twins and 100%:56% for female-female twins. In an extension of this twin-pair study, individuals were subclassified according to the type of disorder that they displayed. Of the four subgroups formed—articulation with or without receptive disorder, articulation and expressive disorder with or without receptive disorder, expressive disorder with or without receptive disorder, and only receptive disorder—those which included children with expressive impairments (i.e., the second and third of these subgroups) showed probandwise MZ:DZ concordance rates close to 100%:50%. In contrast, those with only receptive disorders (i.e., the fourth subgroup) showed little evidence of genetic influence, having a probandwise MZ:DZ concordance rate of 71%:75% (Bishop et al. 1995).

Further support for a genetic etiology in language disorders is provided by estimates of heritabilities of quantitative measures of language-related components. In a series of investigations, Bishop et al. (1995, 1996, 1999) used the DeFries-Fulker method to demonstrate significant heritabilities in several psychometric language measures in families affected by SLI. Many of these measures were comparable to those used in the current study and showed levels of heritability close to 1.0. These include tests of receptive syntactic-language abilities (e.g., the Test for the Reception of Grammar) and tests of expressive-language skills (e.g., the Clinical Evaluation of Language Fundamentals–Revised [CELF-R] repeating-sentences subtest), as well as tests that examine specific processes thought to be important in language acquisition (e.g., tests of nonword repetition). Interestingly, although many language-related traits were shown to be highly heritable, when the discrepancy scores between these traits and PIQ were considered, no significant heritability was seen.

Although there is ample evidence to indicate that genes may play a significant role in the determination of absolute language abilities, family studies have failed to detect any clear cosegregation between phenotype and genotype, and most conclude that the genetic basis is likely to be complex (Bishop et al. 1995). This phenotypic and genotypic complexity has essentially precluded the use of traditional parametric approaches in the genetic mapping of SLI, with one exception. Family KE is a unique three-generation pedigree documented to have a severe speech-and-language disorder that follows an autosomal dominant pattern of inheritance. Investigation of this family and their monogenic trait led to the localization of the SPCH1 locus to chromosome 7q (Fisher et al. 1998) and, ultimately, to the identification of the first gene to be implicated in speech and language development—*FOXP2* (Lai et al. 2001) (MIM 606354). The *FOX* genes encode a large family of transcription factors, all of which possess a winged-helix, or forkhead-box (fox), domain. Lai et al. (2001) have demonstrated that the language impairment in family KE cosegregates with a point mutation in the fox domain of *FOXP2*. They have suggested that the phenotype might result from haploinsufficiency of *FOXP2* at a key stage of embryogenesis, which causes abnormalities in the development of neural structures important for speech and language. Clearly, the *FOX* family represent good candidate genes for SLI; however, their role in the etiology of more common forms of language impairment has yet to be evaluated.

Recent methodological advances have enabled the application of model-free nonparametric approaches to complex disorders, by use of large collections of small nuclear families and analysis of quantitative traits. In the current study, we present the results of the first systematic quantitative-trait locus (QTL)–based genomewide screen for SLI. We use three quantitative measures of different aspects of language abilities and implicate two novel locations, on chromosomes 16 and 19, neither of which coincides with any region previously associated with language impairment. The refinement of the regions reported here may allow the identification of causal genetic factors in cases of SLI and thus aid in the clarification of the etiological mechanisms underlying this disorder.

Am. J. Hum. Genet. 70:384–398, 2002

Subjects and Methods

Subjects

Two centers recruited 473 individuals (including a total of 219 sib pairs) from 98 families. The Newcomen Centre at Guy's Hospital, London, diagnosed and referred a clinically based sample, and the Cambridge Language and Speech Project (CLASP) provided families drawn from an ongoing epidemiological study.

The cases selected at Guy's Hospital were identified through three special schools for language disorders and through Afasic, a support organization for people with developmental and language impairments; thus, these individuals can be considered as representing a self-referred sample of children with persistent language problems needing special schooling and are not representative of the total population in the community. Ethical permission was given by the Guy's and St. Thomas' Trust ethics committee.

CLASP is a community-based longitudinal investigation of speech and language difficulties. The children recruited into the study were initially ascertained during their 3d year of life. A three-stage procedure was employed for case identification, and a standard age design was used to control for divergence between developmental stage and chronological age. Accordingly, at age 36 mo, the population was first defined by means of a questionnaire; then, at age 39 mo, this sample was screened, in more detail, for language difficulties; and, finally, at age 45 mo, age screen–positive cases were assessed in depth. When the children reached 8 years of age, they and their siblings were assessed by the CELF-R and Wechsler Scales of Intelligence–Third UK Edition (WISC-III [Wechsler 1992]), and buccal-DNA samples were collected in families of SLI cases. A detailed description of the ascertainment procedure and sample is available from Stott et al. (in press).

In both the Guy's Hospital sample and the Cambridge sample, probands were selected who, either currently or in the past, had language skills $\geq$1.5 SD below the normative mean for their chronological age, on the receptive and/or expressive scales of the CELF-R battery (Semel et al. 1992). Any proband or sibling found to have a PIQ <80 was excluded from the genome screen. Additional exclusion criteria included MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-10/DSM-IV diagnosis of childhood autism, English being a second language, care provision by local authorities, and known neurological disorders. In the Guy's Hospital sample, those families with chromosome abnormalities, including fragile X, were excluded by cytogenetic testing. A summary of the genome-screen sample is shown in table 1.

Whole-blood or buccal-swab samples were collected from probands and all available siblings and parents, regardless of language ability. DNA was extracted by means of standard protocols, and all buccal-swab DNA samples were preamplified by a preamplification extension protocol (PEP). The PEP technique involves the random amplification of genomic DNA, using a pool of random 15-mer primers, and results in a 50–100-fold increase in template DNA for subsequent microsatellite amplification (Zhang et al. 1992). Prior to the genome screen, this approach was verified, across 20 primers, in a series of 27 controls. All controls showed comparable amplification of both genomic DNA and PEP DNA, and no evidence of preferential preamplification of specific alleles was seen (data not shown).

Table 1

Number of Families and Sib Pairs in the Total Genome-Screen Sample, the Guy's Hospital Sample, and the Cambridge Sample

Category	Distribution When No. of Children in the Family Is				Total
	2	3	4	5	
Total genome-screen sample:					
No. of families	49	44	3	2	98
No. of sib pairs:					
Independent[a]	49	88	9	8	154
All[b]	49	132	18	20	219
Cambridge sample:					
No. of families	31	22	1	1	55
No. of sib pairs:					
Independent[a]	31	44	3	4	82
All[b]	31	66	6	10	113
Guy's Hospital sample:					
No. of families	18	22	2	1	43
No. of sib pairs:					
Independent[a]	18	44	6	4	72
All[b]	18	66	12	10	106

[a] This value is $n - 1$.

[b] All possible pairings of sibs in a sibship; for families with more than two sibs, this is $n(n - 1)/2$.

Phenotypic Measures

Three language measures were assessed for the genome screen: expressive and receptive language skills were scored by CELF-R, and a test of nonword repetition was used as a marker of phonological short-term memory. All three measures showed significant levels of familiality in the genome-screen sample (data not shown). No parental phenotype data were used, since the linkage analysis uses only information from sib-pair phenotype data.

CELF-R

CELF-R is a clinical tool widely used for the identification, diagnosis, and follow-up evaluation of language disorders in school-age children. The battery is split into receptive and expressive scales, which can be combined

to provide a composite language score. Each scale consists of three subtests designed to be primarily receptive or expressive in nature. The exact combination of individual tests used depends on the age of the subject. Additive raw scores from each segment are then transformed to derive a standardized receptive language score (RLS) and an expressive language score (ELS), each with mean 100 and SD 15 in the general-population calibration sample (Semel et al. 1992). The CELF-R tests are generally considered to give a broad overview of a child's general language abilities and are valid for children of age 5–17 years.

Nonword Repetition (NWR)

It has been proposed that children with SLI have language-learning difficulties due to a deficit in working memory. This means that the amount of time during which they are able to hold unfamiliar phonological forms in their short-term memory is insufficient to allow in-depth processing and transfer to the long-term memory (Baddeley and Wilson 1993; Baddeley et al. 1998). To test the capacity that phonological working memory has for novel speech sounds, Gathercole et al. (1994) have developed a measure of NWR. In this test, subjects are required to repeat tape-recorded nonsensical words of increasing length and complexity (e.g., "brufid" and "contramponist"). Studies show that individuals with current language impairments, as well as those who, during early childhood, had language difficulties that later resolved, perform poorly on this test (Gathercole et al. 1994; Bishop et al. 1999). All available children of age 7.5–18 years were tested by the NWR test.

All individuals in the Guy's Hospital sample completed the published version of the children's test of NWR (Gathercole et al. 1994); however, all individuals in the Cambridge sample were examined by a prepublication revision of this test. Although both tasks are similar in administration, and although some words are common to both tests, it was evident that the published standardization introduced flooring effects, which resulted in an undesirable skewing of the distribution of scores. For this reason, as well as to allow combination of the NWR scores across the two samples, both versions of the NWR test were administered to 111 subjects (age 4.8–53.6 years) from both samples, and a between-test-regression calibration coefficient was determined. Raw scores correlated 0.89 ($P < .001$) and were linearly related across the entire range, the relationship being the same for both the adults and the children. A linear-regression calibration equation gave raw scores from the prepublished form of the test that were 0.658 (standard error [SE] 0.009) times the raw score from the published test. Raw scores for the Guy's Hospital sample were therefore multiplied by this factor, to make them comparable to the raw scores for the Cambridge sample. Standard scores for a British population were then obtained by use of norms extended, by S. E. Gathercole (personal communication), for older children.

Intelligence

IQ was assessed by WISC-III (Wechsler 1992). This is a battery of tests that yield measures of verbal IQ and PIQ. The verbal scale comprises tests of comprehension, vocabulary, and abstract reasoning, whereas completion of the performance tasks relies primarily on visual and constructional clues (e.g., mazes, symbol arrangement, and abstract visual problem solving). Verbal IQ and PIQ can then be combined to give a full-scale IQ. The WISC-III requires no reading or writing of words. All children found to have a PIQ <80 were excluded from the study.

Cohort Statistics

A total of 252 children (153 males and 99 females), ages 5–19 years (mean 9.4 years; SD 3.04 years) were assessed, as described above, for CELF-R expressive-language score (by ELS), CELF-R receptive-language score (by RLS), nonword repetition (by NWR), and PIQ. In this sample, which includes unaffected siblings, we found that, although the average level of PIQ was consistent with that of the general population (i.e., mean 100), the means of all language-based measurements fell below the expected mean of 100 (table 2). Thus, the sample selected for the genome screen may be considered to represent a collection of children whose developmental problems are largely language specific.

Comparisons between proband and cosib groups indicated that the probands generally demonstrated language ratings lower than those in the complementary cosibs (table 3). However, although the cosib language scores showed some regression toward the mean, they all remained below that expected (table 3). This is attributable to the high number (~34%) of siblings who also displayed signs of language impairment. In the clinical sample (i.e., that from Guy's Hospital; see the "Recruitment" subsection), 52 (37%) of the children were attending either a special language unit or a special school or had been placed, with a statement of special educational needs, in a mainstream school.

Data Transformation

The data in table 2 present evidence that the Guy's Hospital sample and the Cambridge sample, although both drawn from the general population of children with SLI, are significantly different in the magnitude of severity of their disorders. This is attributable to the fact that, although the diagnostic criteria applied to both samples were identical, the Guy's Hospital sample represents a clinical, severely affected sample, whereas the Cambridge

Table 2

Descriptive Statistics for Each Genome-Screen Phenotype—for the Total Genome-Screen Sample, the Guy's Hospital Sample, and the Cambridge Sample

	Total Genome-Screen Sample				Cambridge Sample			Guy's Hospital Sample[a]		
Statistic	ELS	RLS	NWR[b]	PIQ	ELS	RLS	NWR[b]	ELS	RLS	NWR[b]
Mean	81.68	91.11	96.62	100.24[c]	84.13	94.22	102.42	78.55	87.14	89.12
Median	80	91	99	97	82	93	105	78	87	93.5
SE	1.03	1.17	1.18	1.13	1.38	1.61	1.45	1.49	1.61	1.68
SD	16.05	18.23	18.33	17.39	16.13	18.85	16.97	15.46	16.66	17.34
Interquartile range	21	27	25	24	22	25	21	24	25	26
Skewness	.351	.160	.404	.170	.538	.173	.493	.070	.040	.424
Kurtosis	−.162	−.104	−.384	−.418	−.130	−.174	−.162	−.649	−.313	−.749
Count	244	244	243	239	137	137	137	107	107	106
Mann-Whitney U test (*P*)[d]								.023	.005	.000

[a] Only a selection of the phenotypes available were collected from all siblings: eight siblings were typed for ELS and RLS but not for NWR, and another nine siblings were typed for NWR but not for ELS and RLS. The samples therefore form overlapping subsets of the total sample of 252 children available for genotyping.

[b] Prepublished version.

[c] Consistent with the population average.

[d] Used to ascertain the level of difference between the phenotypic means of the Guy's Hospital sample and the Cambridge sample, prior to amalgamation for the genome screen. In all cases, the means for the two groups varied significantly.

sample represents a more mainstream, epidemiologically selected sample. In order to combine the two samples for variance-components (VC) analysis, which creates a model around a single mean, all phenotypes were standardized to a *Z*-score, $Z = (x - \mu)/\sigma$, where x is the attained score, μ is the mean, and σ is the SD; note that the mean and SD are taken from each group separately. Conversion of the language scores in this manner produces a distribution with a single mean while preserving the variances of the original samples and thus allows a single analysis of the two samples, in the VC model. The standardized scores are hereafter denoted as "RLStrans," "ELStrans, and "NWRtrans" and were used for the combined analysis of both samples, for the genome screen. Correlations between RLStrans, ELStrans, and NWRtrans are given in table 4.

Genotyping and Data Handling

All 473 individuals were genotyped for 400 highly polymorphic dinucleotide-repeat microsatellite markers, taken from the ABI PRISM LMS2-MD10 panels (Applied Biosystems). PCR reactions were performed in 96-well Costar (Thermowell) plates on MJ Research PTC-225 thermocyclers. The fluorescent labeling of primers, with 6-FAM, HEX, and NED phosphoramidites (Applied Biosystems), allowed both the pooling of panels of PCR products and, by means of ABI 373A and 377 sequencers (Applied Biosystems), their subsequent separation and detection on 5% polyacrylamide gels.

Data were extracted from gels by GENESCAN software (version 3.1) and were passed into the GENOTYPER program (version 2.0) for automated allele calling and manual genotype verification (Reed et al. 1994). Raw allele-size data were checked for inconsistencies, by GAS software (version 2.0) (A. Young, personal communication). Marker-allele frequencies were estimated within RECODE (version 1.4) (D. Weeks, personal communication), and Mega2 (version 2.2) (Mukhopadhyay et al. 1999; also see the Division of Statistical Genetics, Department of Human Genetics, University of Pittsburgh web site) was used for the creation of linkage files in a GENEHUNTER 2.0 (Kruglyak et al. 1996; also see the Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/pub/software/genehunter" web site) package. The Discovery Manager system (Genomica) was used for the storage of genotypic data.

Prior to statistical analyses, two data-verification steps were performed. Marker haplotypes were generated in a GENEHUNTER 2.0 (Kruglyak et al. 1996; also see the Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/pub/software/genehunter" web site) package, and all chromosomes showing an excessive number of recombination events were reex-

Table 3

Descriptive Statistics for Each Genome-Screen Phenotype for the Total Genome-Screen Sample, for Probands and for Cosibs

	Probands			Cosibs		
Statistic	ELS	RLS	NWR	ELS	RLS	NWR
Mean	76.44	86.22	91.86	85.08	94.29	99.62
Median	73	85	92	82	95	102
SD	15.41	17.62	20.69	15.59	17.97	16.02
Interquartile range	22.5	23.5	31.75	20.5	22	20
Skewness	.56	.31	−.02	.28	.07	−.58
Kurtosis	−.12	−.28	−.76	.04	.19	.09
Count	98	98	98	148	148	149

Table 4

Correlations between Phenotypes—in the Total Genome-Screen Sample, the Guy's Hospital Sample, and the Cambridge Sample

	ELStrans	RLStrans	NWRtrans
Total genome-screen sample:			
ELStrans	1.000		
RLStrans	.746	1.000	
NWRtrans	.538	.439	1.000
Cambridge sample:			
ELS trans	1.000		
RLS trans	.759	1.000	
NWR trans	.452	.436	1.000
Guy's Hospital sample:			
ELStrans	1.000		
RLStrans	.706	1.000	
NWRtrans	.597	.335	1.000

amined at the genotype level. Corrected data were then run through SIBMED (sibpair mutation and error detection) (Douglas et al. 2000; also see the Center for Statistical Genetics, University of Michigan web site), to identify possible genotyping errors or mutations. SIBMED uses a hidden Markov model to calculate posterior error probabilities for each sib-pair/marker combination, given all the available marker data, an assumed genotype-error rate (set at 1%), and a known genetic map. All genotypes highlighted by SIBMED were excluded from subsequent analyses.

Sex-averaged marker maps were from the Cooperative Human Linkage Center (see the CHLC Genetics Maps web site) and were supplemented with data from Généthon (Dib et al. 1996).

Information-content maps were produced for each chromosome, in a MAPMAKER/SIBS (version 2.0) (Kruglyak and Lander 1995; also see the Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/distribution/software/sibs" web site) package and were used to determine the markers used in a second round of genotyping, involving 100 microsatellites taken from the Généthon map (Dib et al. 1996) and from the ABI PRISM LMS2-HD5 panels (Applied Biosystems). This additional wave of markers allowed the elimination of gaps in both marker density and information. Final marker density was estimated as being <8 cM, for all chromosomes.

Linkage Analysis

The Haseman-Elston (HE) method (Haseman and Elston 1972) and the VC method (Amos 1994; Pratt et al. 2000) were used within a GENEHUNTER 2.0 (Kruglyak et al. 1996; also see the Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/pub/software/genehunter" web site) package, to calculate—by means of the ELStrans, RLStrans, and NWRtrans scores, as quantitative measures of language ability—both single-point and multipoint LOD scores for all autosomes. Additional multipoint HE and VC analyses were subsequently performed, with the WISC-III (Wechsler 1992) measure of PIQ, for all areas that showed suggestive linkage to a language trait.

GENEHUNTER 2.0 implements a traditional HE regression of squared phenotype differences (D^2) on estimated identity-by-descent (IBD) sharing (ν_i), for each sib pair, at a given genetic locus. At a QTL, the variance of D^2 (σ_i^2) is expected to be negatively correlated with the proportion of markers shared IBD (Haseman and Elston 1972).

For families with more than two children, all possible sib pairings were included in the HE analysis. No weighting of multiple sib pairs was used. Although this unweighted approach has been suggested to lead to false inflation of significance, as a result of dependence between pairs (Hodge 1984), simulations described below indicate that this is not the case in our data set (fig. 1*A*).

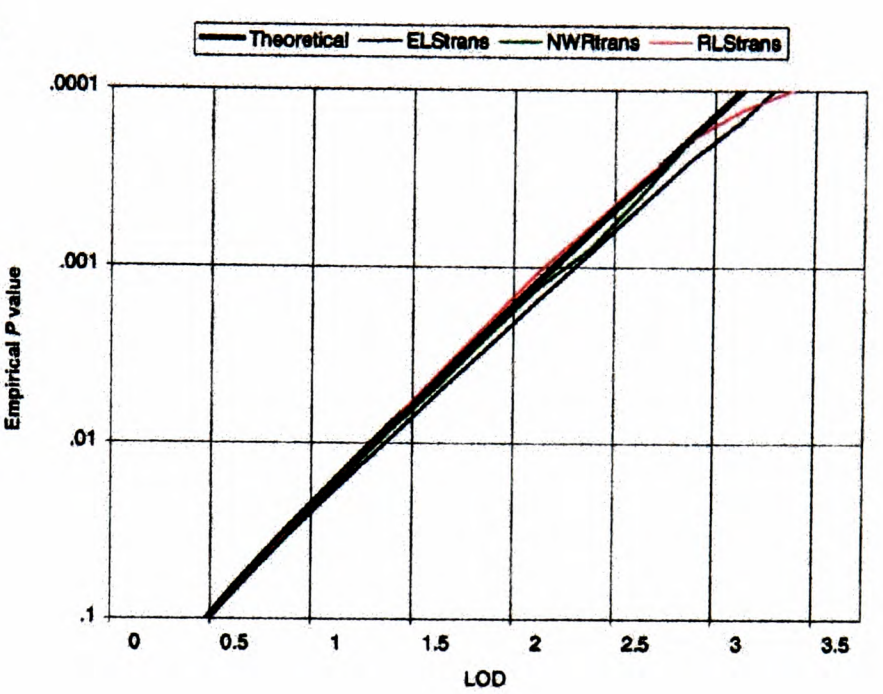
The VC method derives two maximum-likelihood models, both of which dissect the trait variability between siblings into major-gene (σ_a^2), polygenic (σ_p^2), and environmental (σ_e^2) variance components. Under the null hypothesis, it is assumed that there is no major-gene effect (i.e., $\sigma_a^2 = 0$), and in the alternative model the major-gene effect is unrestricted (i.e., $\sigma_a^2 \neq 0$). Comparison of the likelihood of these two models results in a likelihood-ratio estimate, and the theoretical significance of linkage effect can be assessed by a standard χ^2 test (Amos 1994). Empirical estimates of the significance of all VC results were derived by means of simulations, as described below. VC analysis was performed with a single mean and no dominance variance. No adjustment was made for multiple phenotypes. Regions of linkage were identified as those which, under all four types of analysis performed, exceeded thresholds for "suggestive" linkage that have been proposed by Lander and Kruglyak (1995).

X Chromosome

In the absence of a multipoint sex-linked VC method, linkage to the X chromosome was assessed by HE analysis only. Linkage analyses were performed within a MAPMAKER/SIBS (version 2.0) (Kruglyak and Lander 1995; also see the Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/distribution/software/sibs" web site) package, under an HE algorithm comparable to that used by GENEHUNTER 2.0, described above. In X-linked analysis, MAPMAKER/SIBS uses only male-male pairs.

Simulations

Deviations from assumptions made by both of the linkage methods described above (i.e., VC and HE) can



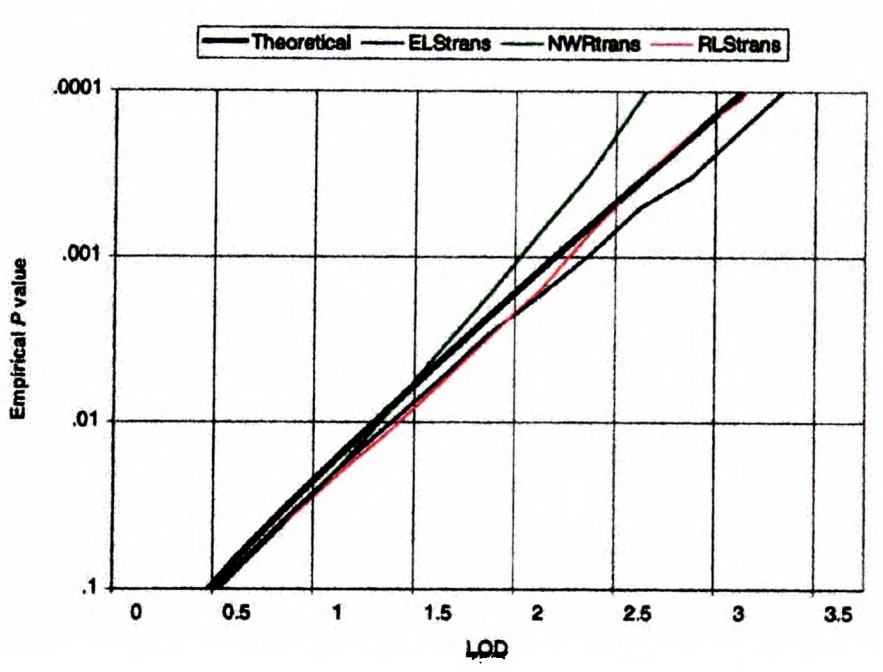


Figure 1 LOD-score–significance distributions for each measure used in the genome screen. The thicker black lines show the theoretical probability for any given LOD score, under the appropriate analyses; the colored lines show the phenotype-specific empirical probabilities for any given LOD score. Because of sample-specific deviations from assumptions critical to different analyses (see the "Subjects and Methods" section), empirical *P* values may differ from theoretical *P* values. Comparisons between empirical and theoretical probability distributions thus allow the quantification of these deviations for each phenotype. *A*, Relationship between HE LOD scores and pointwise significance of linkage in the total genome-screen sample. Simulations demonstrate that the unweighted HE approach does not lead to increased type I errors in our data set (i.e., the empirical distributions coincide with the theoretical distribution). *B*, Relationship between VC LOD scores and pointwise significance of linkage in the total genome-screen sample. Under VC analysis, ELStrans and RLStrans behave as predicted by theory, whereas the theoretical *P* values for NWRtrans are overly conservative.

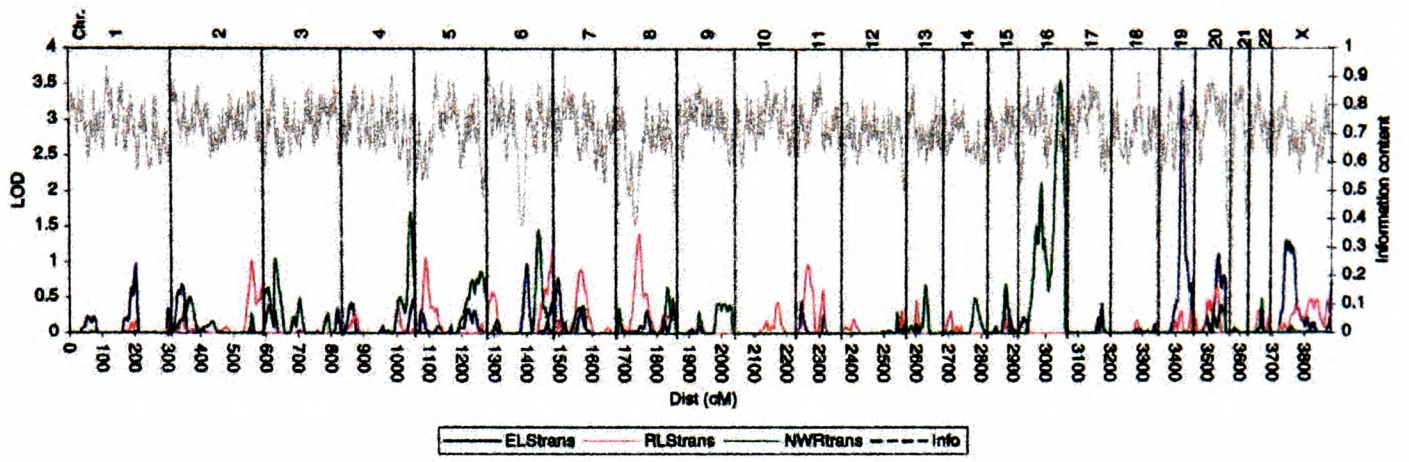


Figure 2 Genomewide plot of HE linkage to three language-related measures under multipoint HE analysis. Abbreviations for language measures are as in the Subjects and Methods section. The *X*-axis shows cumulative distance (in Haldane cM); chromosome numbers are displayed along the top of the graph. Genomewide information content was calculated by MAPMAKER/SIBS. The average information content across the genome is 71%. Note the magnitude of the linkages to chromosomes 16 and 19, in relation to the general background of the genome. VC analysis of the genome fully supported the HE results. All VC LOD scores >1.0 are reported in table 5.

lead to unpredictable variations in the relationship between nominal *P* values and LOD scores, resulting in both type I and type II errors. The VC method supposes the multivariate normality of data, and the unweighted HE method assumes statistical independence of all sib pairings in families with multiple sibships. We adjusted for any divergence from these assumptions, by performing simulations for each phenotype. This allowed an estimation of the empirical pointwise significance of LOD scores.

Pedigree structure and phenotype data were maintained for each family in the genome screen, and SIMULATE (J. Terwilliger, personal communication; also see the Rockefeller University "/software/simulate" web site) was used to generate random genotypes for a single marker with four equally frequent alleles (75% heterozygosity) within this framework. A total of 100,000 replications were run, and linkage was assessed for each, by both the VC approach and the unweighted HE approach.

As demonstrated by Fisher et al. (2002), the resulting LOD-score-significance distributions (fig. 1) can be taken to approximate that found at each point of a typical multipoint situation (where ~70%–80% of IBD information is extracted) and are therefore generally applicable for estimation of the pointwise significance of linkage peaks. Note that these empirical *P* values, although adjusted to account for measure-specific deviations from normality, still yield only pointwise estimates of significance; they are not adjusted to account for genomewide scanning.

Results

We found that, under our ascertainment criteria (i.e., a single language score >1.5 SD below that expected for age), 34.4% of siblings of probands could be classified as affected. If we assume a population prevalence of 4%, this gives a sibling risk ratio of 8.6 in the families that we studied.

Genomewide QTL analysis highlighted two prominent areas of linkage—one on chromosome 16 and one on chromosome 19 (figs. 2 and 3). Although several other regions were found to have LOD scores >1.0 (table 5), only the loci on chromosomes 16 and 19 exceeded the threshold (i.e., LOD score ≥2.2) that Lander and Kruglyak (1995) have proposed as being indicative of "suggestive" of linkage. Furthermore, they did so under all four types of analysis performed (i.e., multipoint HE and VC and single-point HE and VC) (table 5).

The locus on chromosome 16 was linked to the NWRtrans-measured trait and spans ~40 cM of 16q, from D16S515 to D16S520. Although the maximum LOD score (MLS) that HE yielded for this region reached 3.55, the VC analysis yielded a somewhat lower MLS, 2.57. However, empirical-probability distributions drawn from simulated data indicated a general deflation of VC LOD scores for the NWRtrans-measured trait (fig. 1*B*). In 100,000 simulations, a VC LOD score >2.57 was seen only eight times (i.e., pointwise empirical $P = .00008$) and thus is consistent with the HE result (empirical $P = .00003$) and verges on the threshold (i.e., $P = .00002$) that Lander and Kruglyak (1995) have proposed as being indicative of "significant" linkage. Furthermore, chromosome 16 yielded the most significant single-point result for the genome (D16S516; LOD score 2.77), and single-point LOD scores >1.5 were seen for a cluster of three markers directly under the peak of linkage: markers

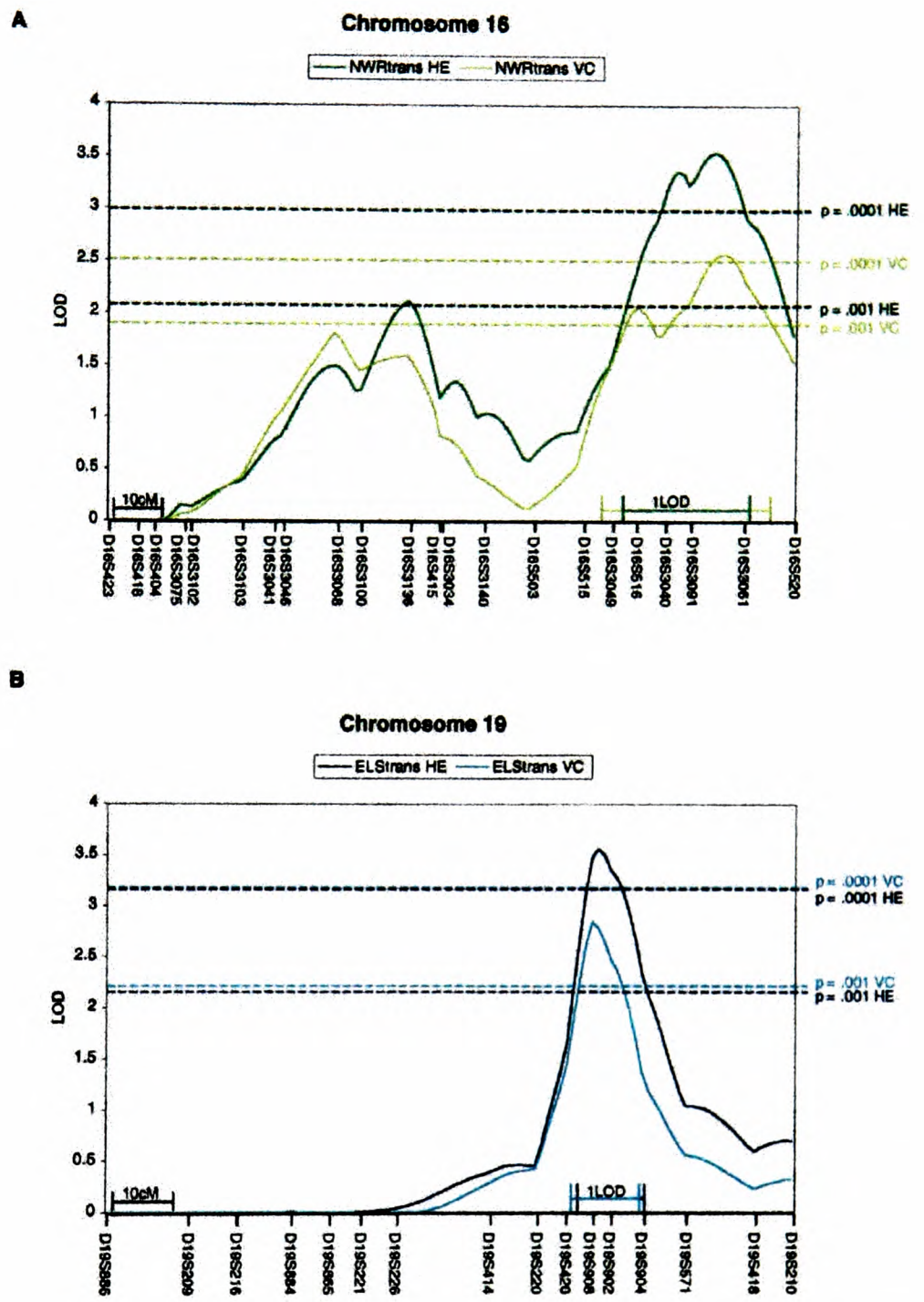


Figure 3 Suggestive linkage to chromosomes 16 and 19. The *X*-axis shows positions of the markers typed; a 10-cM (Haldane) bar is given for reference. *A*, Linkage to chromosome 16, under both the HE method and the VC method, for NWRtrans. For ELStrans and RLStrans, the LOD scores remained <0.42 and <0.27, respectively, for the entire chromosome (data not shown). A 1-LOD interval is shown for both the HE peak and the VC peak, by the dark-green and light-green bars, respectively. Light-green lines show the LOD-score thresholds for empirical *P* values of .001 and .0001 for NWRtrans under VC analysis; dark-green lines show the same LOD-score thresholds under HE analysis. *B*, Linkage to chromosome 19, under both the HE method and the VC method, for ELStrans. For RLStrans and the NWRtrans, the highest LOD scores were 0.33 and 0.20, respectively, for the entire chromosome (data not shown). A 1-LOD interval is shown for both the HE peak and the VC peak, by the dark-blue and light-blue bars, respectively. Light-blue lines show the LOD-score thresholds for empirical *P* values of .001 and .0001 for ELStrans under VC analysis; dark-blue lines show the same LOD-score thresholds under HE analysis.

Table 5

LOD Scores >1.0

Chromosomal Region	MLS[a] Multipoint (Empirical P[b]) HE	VC	Single Point[c]	Score[d]
1q24	.98 (.01924)	1.16 (.01246)	1.25 (D1S218)	ELStrans
2q36	1.03 (.01468)	1.52 (.00574)	.72 (D2S338)	RLStrans
3p24	1.05 (.01534)	1.64 (.00238)	1.42 (D3S1266)	NWRtrans
4q35	1.70 (.00296)	.89 (.02632)	1.00 (D4S1535)	NWRtrans
5p15	1.06 (.01310)	1.32 (.00950)	2.12 (D5S416)	RLStrans
5q34	.86 (.02506)	1.76 (.00160)	1.76 (D5S1960)	NWRtrans
6q25	1.46 (.00522)	1.07 (.01554)	1.23 (D6S441)	NWRtrans
7q11	.90 (.02146)	1.24 (.01544)	.81 (D7S669)	RLStrans
8p12	1.41 (.00508)	1.22 (.01220)	1.56 (D8S260)	RLStrans
8q24	.66 (.04378)	1.40 (.00534)	1.10 (D8S272)	NWRtrans
15q22	.69 (.03990)	1.05 (.01620)	1.19 (D15S153)	NWRtrans
16q24	3.55 (.00003)	2.57 (.00008)	2.77 (D16S516)	NWRtrans
19q13	3.55 (.00004)	2.84 (.00027)	2.49 (D19S908)	ELStrans
20q13	1.11 (.01414)	1.43 (.0063)	.93 (D20S171)	ELStrans
Xp11	1.30			ELStrans

[a] According to Kruglyak and Lander's (1995) guidelines, suggestive linkage corresponds to a pointwise significance of .0007, and significant linkage corresponds to a pointwise significance of .00002. Regions of "linkage" were identified as those which, under all types of analysis performed, exceeded thresholds for suggestive linkage (Lander and Kruglyak 1995); values considered to be suggestive are underlined.

[b] Calculated by simulations, as described in the "Subjects and Methods" section.

[c] Values are for the marker with the highest single-point (by either HE or VC) LOD score in the area of linkage and are provided as a guide to the level of single-point support given to the multipoint results.

[d] Phenotype for which the highest LOD score is seen; other traits may also show linkage to the same region, although at lower levels.

D16S516 (2.77), D16S3040 (2.24), and D16S3091 (1.95) (table 5).

The locus on chromosome 19 was linked to the ELStrans-measured trait and covers ~30 cM of 19q, from D19S220 to D19S418 (fig. 3). This QTL was evident under both the HE and the VC multipoint analyses (HE LOD score 3.55; VC LOD score 2.84) and was supported by single-point analysis, in which two adjacent markers showed linkage to the ELStrans-measured trait, with LOD scores >1.5 (D19S908 [LOD score 2.49]) and D19S902 [LOD score 1.74]) (table 5). Simulations indicated that the ELStrans measure behaves as predicted by theory and, therefore, that empirical P values can be taken as being representative of nominal P values for linkages to the ELStrans-measured trait and, therefore, to the locus on chromosome 19 (fig. 1).

In both the chromosome 16 region of linkage and the chromosome 19 region of linkage, the LOD score for PIQ was never >0.15 (data not shown), indicating that both of these loci are likely to reflect language-specific influences, as opposed to general-intelligence effects.

To clarify the contribution that each group made to the two linkage peaks, we divided the genome-screen sample into its constituent Guy's Hospital and Cambridge samples and reanalyzed chromosomes 16 and 19. Figure 4 shows that the Guy's Hospital sample and the Cambridge sample contribute equally to both peaks of linkage.

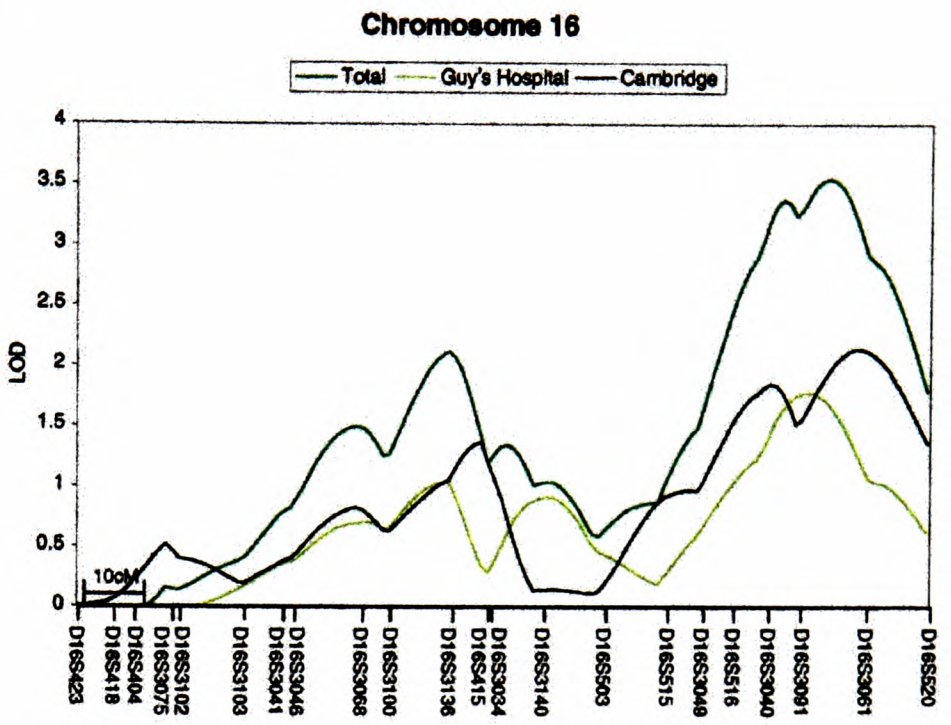
Interestingly, we found no evidence for linkage to chromosome 7q, the location of both SPCH1 (the family-KE linkage) (MIM 602081) and AUTS1 (the autism chromosome 7 linkage) (MIM 209850). At D7S486, the peak (LOD score 6.22) of linkage in family KE, our single-point LOD score remained <0.001, for all three phenotypes.

Our sample contained a male:female ratio of ~3:2, which is consistent with the male predominance reported in previous studies (Stevenson and Richman 1976). However, we found no strong evidence for a major sex-specific locus in the HE analysis of the X chromosome. Although, for ELStrans, a LOD score of 1.30 was found on Xp, LOD scores for all other measures remained <0.5, across the entire X chromosome.

Discussion

We have reported here the first molecular genetic study of typical SLI. We implicate two novel loci, on chromosomes 16 and 19, that are found to influence language-related traits. Evidence for these QTLs has been drawn from four complementary analyses (i.e., multipoint HE and VC and single-point HE and VC), and

A



B

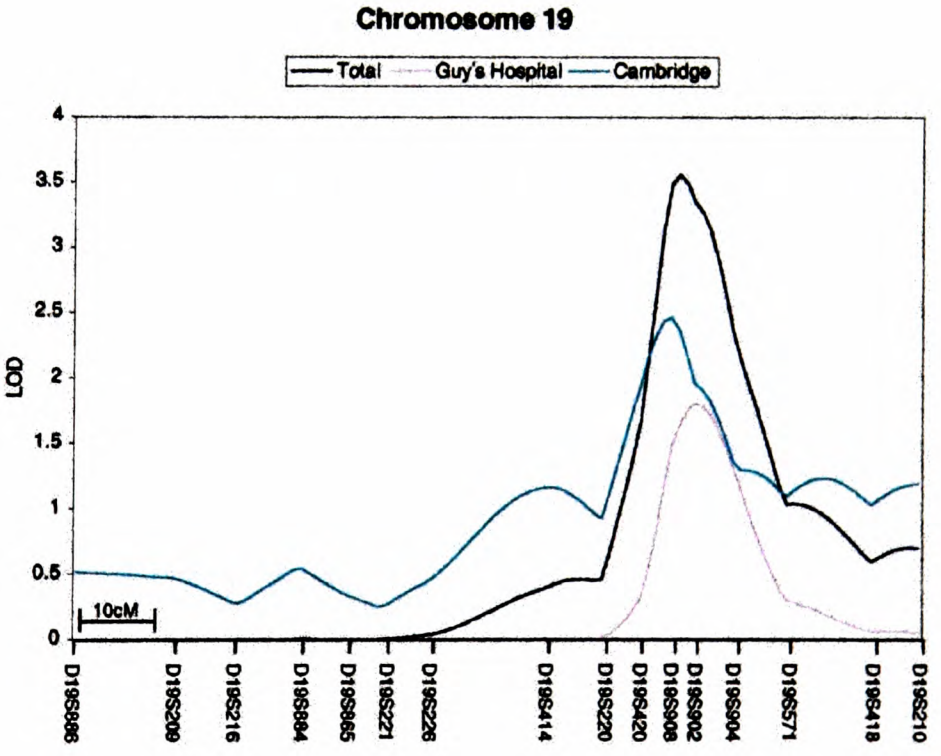


Figure 4 Linkage to chromosomes 16 and 19, based on the Guy's Hospital sample and the Cambridge sample. The format of the graph is as in figure 3. *A*, Linkage to chromosome 16, in the combined genome-screen sample, the Guy's Hospital sample, and the Cambridge sample, for NWRtrans. Traces are shown for HE analysis only. *B*, Linkage to chromosome 19, in the total genome-screen sample, the Guy's Hospital sample, and the Cambridge sample, for ELStrans. Linkage is independently demonstrated both in the Guy's Hospital sample and in the Cambridge sample. Traces are shown for HE analysis only.

both loci have shown to be relevant in the Guy's Hospital sample and the Cambridge sample.

One important feature of this study is the use of quantitative measures of generalized language abilities. The lack of consensus as to the etiological basis of SLI often makes the derivation of a consistent qualitative affection status unfeasible. The use of quantitative traits circumvents this issue and, in complex cognitive disorders, has been demonstrated to provide a suitable means of investigation of underlying genetic effects (Cardon et al. 1994; Fisher et al. 1999; Gayán et al. 1999). A quantitative-trait approach does, however, create its own issues, perhaps the most pertinent of which is the selection of phenotypes for the appraisal of disorder severity. In the diagnosis of SLI, both ICD-10 and DSM-IV guidelines require a substantial discrepancy between PIQ and verbal abilities. Although the enforcement of discrepancy scores acts to aid the elimination of general IQ effects, these scores are generally felt to result in an overly restrictive phenotype, which is susceptible to compound errors. Also relevant to the current study is the finding that discrepancy scores show only a minimal level of heritability and, hence, may not reflect the underlying genetic influences involved in SLI (Bishop et al. 1995). For the genome screen, we therefore chose to employ broad phenotype batteries, alongside a single specific measurement of phonological short-term memory. All three traits have been demonstrated to be significantly heritable and good predictors of language abilities (Semel et al. 1992; Bishop et al. 1995, 1999).

The importance of phenotype selection has been confirmed by the results of the genome screen in the current study. Intriguingly, only minimal linkage is seen to measures of receptive language abilities—the strongest RLStrans result was seen for chromosome 2q and peaked at 1.52—a result consistent with the previously reported lack of probandwise concordance between twins with only receptive language impairments (Bishop et al. 1995). In contrast, measurements of expressive language skills and phonological short-term memory, both of which have been demonstrated to be subject to strong genetic influence (Bishop et al. 1995, 1999), yielded the two most significant linkage results in the genome screen. These two loci provide the only areas of suggestive linkage in the entire genome. The background level was generally low, with few regions yielding LOD scores >1.5 (table 5).

Furthermore, although all three phenotypes were found to be moderately correlated in our sample (table 4), at both peaks of linkage a discordance between all traits was apparent. Linkages to chromosomes 16 and 19 were seen to be specific to NWRtrans and ELStrans, respectively, with no corresponding peaks seen for the other measures (see fig. 3). However, studies of dyslexia (Grigorenko et al. 1997; Fisher et al. 1999) indicate that the dissection of a complex trait in such a simple manner is not always appropriate and that inferences relating specific loci to distinct components of language impairment should be viewed with caution (Fisher et al. 1999).

In genome screens for complex traits, it is not uncommon to see a shift between the original peak and replication peaks or to find linkage to phenotypes other than that originally reported (Cardon et al. 1994; Grigorenko et al. 1997; International Molecular Genetic Study of Autism Consortium 1998; Fagerheim et al. 1999; Fisher et al. 1999; Gayán et al. 1999; Phillipe et al. 1999; Risch et al. 1999). Thus, the independent reproduction of the loci on chromosomes 16 and 19, in both the Guy's Hospital sample and the Cambridge sample, was particularly striking. The observation of linkage—in exactly the same region and to the same phenotypes—across two separate groups with such different origins provides further endorsement for the QTLs reported here.

The only previously reported linkage to SLI is that of the SPCH1 region on chromosome 7q in family KE (Fisher et al. 1998). Our genome screen shows no evidence for linkage to this area, indicating that it is unlikely to play a significant role in cases of typical SLI. However, given the heterogeneity of the disorder, it remains possible that a subset of individuals in the current study may harbor mutations in *FOXP2*, and it should be stressed that mutation analysis of the gene will be necessary to assess the full impact of this locus in our subjects with SLI.

The overlap between the SPCH1 and AUTS1 chromosome 7 linkages (International Molecular Genetic Study of Autism Consortium 1998; Lai et al. 2000) has fueled much debate with regard to the relationship between the genetic and phenotypic overlap of SLI and autism (Folstein and Mankoski 2000; Vincent et al. 2000; Warburton et al. 2000). Although some autistic children may develop language that is normal in terms of vocabulary, grammar, and phonology, they invariably encounter difficulties with the use of language in a social context (i.e., pragmatic language). It is estimated that one-third of autistic children never develop speech at all (Rapin 1997). In addition, the prevalence of autism in the siblings of children affected by SLI has often been reported to be increased compared with that in the general population (3%:0.1%) (Hafeman and Tomblin 1999). A recent genome screen for loci involved in autism has implicated a 19q region that is coincident with the chromosome 19 peak reported here (Lui et al. 2001). However, the autistic sample that showed the greatest evidence of linkage to this 19q locus (MLS 1.70) was from the narrow diagnostic group (i.e., that excluding children who may overlap into the SLI spectrum). We found no additional evidence for linkage to any other

major loci (i.e., on chromosomes 2, 7, and 15) that previous studies had found to be associated with autism (reviewed by Lamb et al. 2000).

Another disorder that shows significant comorbidity with SLI is dyslexia (Bishop and Adams 1990; Catts 1993). The strong links between both dyslexia and SLI and phonological impairments have often led to the speculation that language impairments and reading disabilities may represent different manifestations of similar neurological deficits (Snowling et al. 2000). Relatives of probands affected by dyslexia experience an increased risk of language impairment (Gallagher et al. 2000), whereas studies of children selected for language impairments often report a high incidence of literacy problems (Tallal et al. 1989). However, the exact relationship between the two disorders remains undetermined. We found no evidence for linkage to regions of chromosomes 2, 6, 15, or 18, the regions that previously implicated by genetic mapping studies of dyslexia (Cardon et al. 1994; Grigorenko et al. 1997; Fisher et al. 1999, 2002; Gayán et al. 1999). Further independent studies involving large sample sizes will be necessary to elucidate any common genetic mechanisms underlying the phenotypic overlaps between both autism and dyslexia and SLI.

QTL genome screens such as the one that has been reported here are crucial to the study of disorders such as SLI, since they neither make no prior assumptions about the basis of the disease nor target specific chromosomal regions for analysis. The current study has provided an overview of the whole genome with respect to language-related phenotypes and has highlighted two loci that appear to have a significant genetic effect on the development of SLI. This work represents the first major step in the clarification of the genetic mechanisms behind SLI, which may lead to a better understanding of the processes involved in language acquisition while also facilitating better diagnosis and treatment of individuals with language impairments.

Acknowledgments

We would like to thank all the families who have participated in the study, as well as the professionals who continue to make this study possible. We thank Michael Rutter, for initial discussions on the genetics of SLI; Lon Cardon, Dan Weeks, and Clyde Francks, for their statistical advice; all members of the Monaco lab, for their support and advice during the past 3 years; Jonathan Roiser, for his PEP work; all at the Newcomen and CLASP centers, for their involvement in the project; and Jane Addison, Claire Poppy, Deborah Jones, Tilly Storr, and Til Utting-Brown, for their assistance with data collection and management. All laboratory work and the collection of Guy's families were funded by The Wellcome Trust. CLASP is funded by The Wellcome Trust, British Telecom, The Isaac Newton Trust, a National Health Service (NHS) Anglia & Oxford Regional R&D Strategic Investment Award, and an NHS Eastern Region R&D Training Fellowship Award. D. F. Newbury is funded by an Medical Research Council Studentship, and A. P. Monaco and D. V. M. Bishop are Wellcome Trust Principal Fellows.

Appendix

Organizations (members) of the SLI Consortium are as follows: The Wellcome Trust Centre for Human Genetics (D. F. Newbury, J. D. Cleak, Y. Ishikawa-Brush, A. J. Marlow, S. E. Fisher, and A. P. Monaco), CLASP (C. M. Stott, M. J. Merricks, I. M. Goodyer, and P. F. Bolton), Newcomen Centre, Guy's Hospital (L. Jannoun, V. Slonims, and G. Baird), School of Epidemiology & Health Science, The University of Manchester (A. Pickles), Department of Experimental Psychology, University of Oxford (D. V. M. Bishop), Human Communication and Deafness, School of Education, The University of Manchester (G. Conti-Ramsden), and Department of Child Health, Aberdeen University (P. J. Helms).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Center for Statistical Genetics, University of Michigan, http://www.sph.umich.edu/statgen/software (for SIBMED)

CHLC Genetics Maps, http://lpg.nci.nih.gov/html-chlc/ChlcMaps.html

Division of Statistical Genetics, Department of Human Genetics, University of Pittsburgh, http://watson.hgen.pitt.edu/mega2.html (for Mega2 version 2.2)

Généthon, http://www.genethon.fr/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for AUTS1 [MIM 209850], *FOXP2* [MIM 606354], and SPCH1 [MIM 602081])

Rockefeller University "/software/simulate" web site, ftp://linkage.rockefeller.edu/software/simulate (for SIMULATE)

Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/distribution/software/sibs" web site, ftp://ftp-genome.wi.mit.edu/distribution/software/sibs (for MAPMAKER/SIBS)

Whitehead Institute for Biomedical Research/MIT Center for Genome Research "/pub/software/genehunter" web site, http://www-genome.wi.mit.edu/ftp/pub/software/genehunter/ (for GENEHUNTER 2.0)

References

Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54: 535–543

Baddeley A, Gathercole S, Papagno C (1998) The phonological loop as a language learning device. Psychol Rev 1:158–173

Baddeley A, Wilson BA (1993) A developmental deficit in short-term phonological memory: implications for language and reading. Memory 1:65–78

Beitchman JH, Brownlie EB, Inglis A, Wild J, Mathews R,

Schachter D, Kroll R, Martin S, Ferguson B, Lancee W (1994) Seven-year follow-up of speech/language-impaired and control children: speech/language stability and outcome. J Am Acad Child Adolesc Psychiatry 33:1322–1330

Bishop DVM, Adams C (1990) A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. J Child Psychol Psychiatry 31:1027–1050

Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P (1999) Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. J Speech Lang Hear Res 42: 155–168

Bishop DV, Edmundson A (1986) Is otitis media a major cause of specific developmental language disorders? Br J Disord Commun 21:321–338

Bishop DV, North T, Donlan C (1995) Genetic basis of specific language impairment: evidence from a twin study. Dev Med Child Neurol 37:56–71

——— (1996) Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. J Child Psychchol Psychiatry 37:391–403

Cantwell DP, Baker L (1987) Prevalence and type of psychiatric disorder and developmental disorders in three speech and language groups. J Commun Disord 20:151–160

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279

Catts HW (1993) The relationship between speech-language impairments and reading disabilities. J Speech Hear Res 36: 948–958

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Douglas JA, Boehnke M, Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. Am J Hum Genet 66:1287–1297

Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA (1999) A new gene (DYX3) for dyslexia is located on chromosome 2. J Med Genet 36:664–669

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayán J, Olson RK, Pennington BF, Smith SD, DeFries JC, Stein JF, Monaco AP (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 30 (in press)

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet 64:146–156

Fisher SE, Vargha-Khadem, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168–170

Folstein SE, Mankoski RE (2000) Chromosome 7q: where autism meets language disorder? Am J Hum Genet 67:278–281

Gallagher A, Frith U, Snowling MJ (2000) Precursors of literacy delay among children at genetic risk of dyslexia. J Child Psychol Psychiatry 41:202–213

Gathercole SE, Willis C, Baddeley AD, Emslie H (1994) The children's test of nonword repetition: a test of phonological working memory. Memory 2:103–127

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64:157–164

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet 60:27–39

Hafeman L, Tomblin B (1999) Autistic behaviours in the siblings of children with specific language impairment. Mol Psychiatry Suppl 4:S14

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. Hum Mol Genet 7: 571–578

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Lai CSL, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED, Vargha-Khadem F, Monaco AP (2000) The *SPCH1* region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am J Hum Genet 67:357–368

Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A novel forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:465–466

Lamb JA, Moore J, Bailey A, Monaco AP (2000) Autism: recent molecular advances. Hum Mol Genet 9:861–868

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Law J, Boyle J, Harris F, Harkness A, Nye C (1998) Screening for speech and language delay: a systematic review of the literature. Health Technol Assess 2:1–184

Lewis BA, Thompson LA (1992) A study of developmental speech and language disorders in twins. J Speech Hear Res 35:1086–1094

Lui J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D, Lord C, Iversen P, Hoh J, the Autism Genetic Resource Exchange Consortium, Ott J, Gillam TC (2001) A genomewide screen for autism susceptibility loci. Am J Hum Genet 69:327–340

Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE (1999) Mega2, a data-handling program for facilitating genetic linkage and association analyses. Am J Hum Genet Suppl 65:A436

Neils J, Aram DM (1986) Family history of children with developmental language disorders. Percept Mot Skills 63: 655–658

Phillippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman E, Zappella M, Aschauer H, van Malldergerme L, Penet C, Feingold J, Brice A, Leboyer M, the Paris Autism Research International Sibpair Study (1999) Genome-wide scan for autism susceptibility genes. Hum Mol Genet 8:805–812

Pratt SC, Daly MJ, Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. Am J Hum Genet 66:1153–1157

Rapin I (1997) Current concepts: autism. N Engl J Med 337: 97–104

Reed PW, Davies JL, Copeman JB, Bennet ST, Palmer SM, Pritchard LE, Gough SC, et al (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nat Genet 7:390–395

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

Semel EM, Wiig EH, Secord W (1992) Clinical Evaluation of Language Fundamentals–Revised. Psychological Corporation, San Antonio

Snowling MJ, Adams JW, Bishop DV, Stothard SE (2001) Educational attainments of school leavers with a preschool history of speech-language impairments. Int J Lang Commun Disord 36:173–183

Snowling M, Bishop DV, Stothard SE (2000) Is preschool language impairment a risk factor for dyslexia in adolescence? J Child Psychol Psychiatry 41:587–600

Stevenson J, Richman N (1976) The prevalence of language delay in a population of three-year-old children and its association with general retardation. Dev Med Child Neurol 18:431–441

Stott CM, Merricks MJ, Bolton PF, Goodyer IM. Screening for speech and language disorders: the reliability, validity and accuracy of the general language screen. Int J Lang Commun Disord (in press)

Stromswold K (1998) Genetics of spoken language disorders. Hum Biol 70:297–324

Tallal P, Ross R, Curtiss S (1989) Familial aggregation in specific language impairment. J Speech Hear Disord 54:167–173

Tomblin JB, Buckwalter PR (1998) Heritability of poor language achievement among twins. J Speech Lang Hear Res 41:188–199

Tomblin JB, Records NL, Zhang X (1996) A system for the diagnosis of specific language impairment in kindergarten children. J Speech Hear Res 39:1284–1294

Vincent JB, Herbrick J-A, Gurling HMD, Bolton PF, Roberts W, Scherer SW (2000) Identification of a novel gene on chromosome 7q31 that is interpreted by a translocation breakpoint in an autistic individual. Am J Hum Genet 67:510–514

Warburton P, Baird G, Chen W, Morris K, Jacobs BW, Hodgson S, Docherty Z (2000) Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. Am J Med Genet 96: 228–234

Wechsler D (1992) Wechsler Intelligence Scale for Children–Third UK Edition. Psychological Corporation, London

Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N (1992) Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci USA 89: 5847–5851

Am. J. Hum. Genet. 70:1318–1327, 2002

Report

FOXP2 Is Not a Major Susceptibility Gene for Autism or Specific Language Impairment

D. F. Newbury,[1,*] E. Bonora,[1,*] J. A. Lamb,[1] S. E. Fisher,[1] C. S. L. Lai,[1] G. Baird,[2] L. Jannoun,[2] V. Slonims,[2] C. M. Stott,[4] M. J. Merricks,[4] P. F. Bolton,[4] A. J. Bailey,[3] A. P. Monaco,[1] and the International Molecular Genetic Study of Autism Consortium[†]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford; [2]Newcomen Centre, Guy's Hospital, and [3]Centre for Social, Genetic and Developmental Psychiatry and Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London; and [4]University of Cambridge, Developmental Psychiatry Section, Cambridge, United Kingdom

The *FOXP2* gene, located on human 7q31 (at the *SPCH1* locus), encodes a transcription factor containing a polyglutamine tract and a forkhead domain. *FOXP2* is mutated in a severe monogenic form of speech and language impairment, segregating within a single large pedigree, and is also disrupted by a translocation in an isolated case. Several studies of autistic disorder have demonstrated linkage to a similar region of 7q (the *AUTS1* locus), leading to the proposal that a single genetic factor on 7q31 contributes to both autism and language disorders. In the present study, we directly evaluate the impact of the *FOXP2* gene with regard to both complex language impairments and autism, through use of association and mutation screening analyses. We conclude that coding-region variants in *FOXP2* do not underlie the *AUTS1* linkage and that the gene is unlikely to play a role in autism or more common forms of language impairment.

Autism is a neurodevelopmental disorder characterized by deficits in reciprocal social interaction and communication, accompanied by repetitive and stereotyped behaviors and interests (World Health Organization 1993; American Psychiatric Association 1994).

Specific language impairment (SLI) is defined as a significant deficit in language development that exists despite adequate educational opportunity and normal nonverbal intelligence. A diagnosis of SLI is made after ruling out the presence of other conditions, such as autism (Tomblin et al. 1996).

Although autism and SLI are generally accepted to be clinically distinct, the boundaries between the two conditions are not always clear, and there remains a group of children who show social and/or language difficulties yet fail to meet strict diagnostic criteria for either disorder. Some have argued for the formation of a "semantic-pragmatic" classification for these "borderline autistic/language impaired" individuals (Rapin and Allen 1983; Bishop and Rosenbloom 1987).

Language deficits form a major component of the autism diagnostic criteria, and, in general, autistic individuals tend to experience more-severe linguistic impairments than are associated with SLI alone (Lord et al. 1994). Autistic children typically make few spontaneous remarks, produce stereotyped utterances, and make only minimal use of gesture (Tager-Flusberg et al. 2001). Although some autistic children may develop acceptable skills in terms of vocabulary, grammar, and phonology, they invariably retain fundamental difficulties with the use of language in a social context (i.e., pragmatics) (Mawhood et al. 2000). A substantial proportion of autistic children completely fail to develop language at all (Rapin 1997; Tager-Flusberg et al. 2001). In contrast, the types of language problems seen in SLI tend to be more heterogeneous. Children affected by SLI show a wide range and severity of deficits with respect to the articulation of speech sounds, verbal expression, and comprehension of speech (Bishop 1994; Conti-Ramsden et al. 1997). Pragmatic impairments are usually absent or mild.

Received November 21, 2001; accepted for publication January 24, 2002; electronically published March 13, 2002.

Address for correspondence and reprints: Dr. A. P. Monaco, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom. E-mail: anthony.monaco@well.ox.ac.uk

* The first two authors contributed equally to this work.

† A list of members of the Consortium can be found in the Acknowledgments section.

0002-9297/2002/7005-0022$15.00

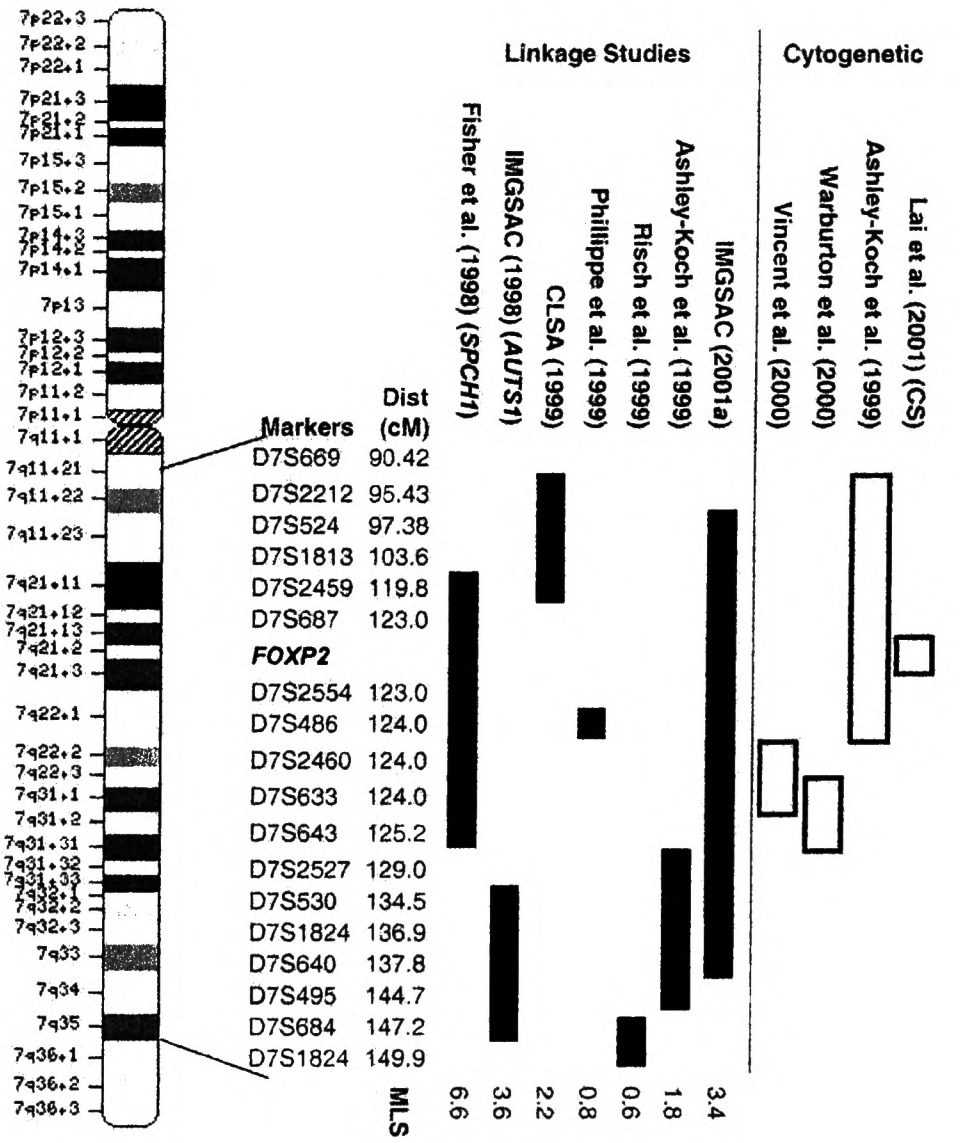


Figure 1 Autism and language studies of chromosome 7q. The chromosome 7 ideogram shows the order and map distance of markers used in various studies. Blackened boxes show approximate positions of regions highlighted by linkage studies. Unblackened boxes represent approximate positions of breakpoints in cytogenetic studies. Each MLS shown was obtained in the region highlighted for the appropriate linkage study. The methods of LOD estimation varied between studies. Note that the *SPCH1* linkage has been directly attributed to the *FOXP2* gene.

There is now a large amount of evidence, from family and twin studies, indicating a strong role for genetic factors in both autism (Folstein and Rutter 1977; Steffenburg et al. 1989; Bolton et al. 1994; Bailey et al. 1995) and SLI (Lewis and Thompson 1992; Bishop et al. 1995; Tomblin and Buckwalter 1998). However, it is accepted that each of these conditions is complex in nature, with several loci interacting to produce a genetic liability to disease onset (Pickles et al. 1995).

Recent advances in technology and statistical genetics have allowed the completion of several genomewide scans using sibling pairs affected by autism. The first of these studies yielded a maximum LOD score (MLS) of 3.55 in a 40-cM region on the long arm of human chromosome 7, between markers D7S530 and D7S684, in a subset of families from the United Kingdom (the *AUTS1* locus [MIM 209850]; fig. 1) (International Molecular Genetic Study of Autism Consortium [IMGSAC] 1998). The involvement of this *AUTS1* locus in autism is further supported by several independent linkage investigations with differing degrees of significance and varying chromosomal locations. (fig. 1) (Ashley-Koch et al. 1999; Collaborative Linkage Study of Autism 1999; Phillippe et al. 1999; Risch et al. 1999; IMGSAC 2001*a*).

Concurrent studies of the KE family, a unique three-generation pedigree with a severe monogenic speech and language disorder, independently yielded strong evidence for linkage to a similar region of 7q31, between markers D7S2459 and D7S643 (the *SPCH1* locus [MIM 602081]; fig. 1) (Fisher et al. 1998; Lai et al. 2000). The KE phe-

notype is characterized by severe orofacial dyspraxia, which impedes complex articulatory movement, accompanied by extreme impairments in both expressive and receptive language skills. There is also evidence of nonverbal deficits in some individuals (Vargha-Khadem et al. 1995). Although the affected members of the KE family show no autistic features and do not meet strict diagnostic criteria for SLI, the overlap between the *SPCH1* and *AUTS1* loci raised the question of whether a single gene on 7q might be involved in both autism and SLI (Folstein and Mankoski 2000). Such a hypothesis is strengthened by cytogenetic studies of individuals with chromosome 7 abnormalities (fig. 1). Ashley-Koch et al. (1999) described a family with a pericentric inversion of the long arm of chromosome 7 (inv[7][q22-q31.2]), transmitted from an unaffected mother to all three of her children. Interestingly, two of the three siblings in this family were affected by autism, and the third presented with a severe expressive-language disorder. Vincent et al. (2000) characterized a translocation transmitted from an unaffected mother to an autistic child (t[7;13][q31.2;q21]) and mapped the breakpoint within a highly conserved, brain-expressed gene of unknown function (*RAY1*) between markers D7S2460 and D7S633. Warburton et al. (2000) described two unrelated individuals, one with autism and a second with a severe expressive-language impairment, both of whom showed de novo abnormalities involving breakpoints on chromosome 7q31 (inv[7][p12.2;q31.3] and t[2;7][p23;q31.3], respectively). Finally, Lai et al. (2001) described a child (referred to as "CS") affected by a severe orofacial dyspraxia and language deficits similar to those seen in the KE family, with a de novo translocation (t[5;7][q22;q31.2]) mapping to the *SPCH1* region.

Recently, the gene mutated in the KE family was identified as *FOXP2* (MIM 605317) (Lai et al. 2001). The *FOX* genes encode a large family of transcription factors, all of which possess a winged-helix—or forkhead box ("fox")—DNA-binding domain. The known sequence of *FOXP2*, as reported by Lai et al. (2001), is organized into 19 exons (2 of which are alternatively spliced), and the major splice form encodes a 715-residue protein containing a characteristic fox domain (exons 12–14) and a 40-residue polyglutamine tract (exons 5 and 6). The polyglutamine repeat is encoded by a mixture of CAG and CAA codons and has been demonstrated to be stable in normal individuals (Lai et al. 2001). The mutation identified in the KE family is a G→A transition in exon 14 that cosegregates with the speech and language disorder in the KE pedigree. This nonsynonymous change results in an arginine-to-histidine substitution at a highly conserved residue within the fox domain (Lai et al. 2001). Furthermore, the *FOXP2* gene was directly disrupted by the chromosomal breakpoint of the unrelated translocation patient, CS. Lai et al. (2001) suggested that the KE and CS phenotypes may be caused by haploinsufficiency of *FOXP2* at a key stage of embryogenesis, which results in the abnormal development of neural structures important for speech and language.

Clearly, there is strong support for the role of chromosome 7q31 in the etiology of both autism and language disorders. However, questions remain with regard to the relevance of *FOXP2* within more common and genetically complex forms of language impairment, and it is still a matter of debate as to whether the phenotypic and genetic overlaps between autism and SLI are caused by the same or by different loci. The present study therefore presents the characterization of *FOXP2* within samples of patients with SLI and autism, with two aims. The first is to assess the relevance of the *FOXP2* gene within forms of language impairment more common than those found in the KE family and in the translocation patient CS, and the second is to directly evaluate the hypothesis that the overlap in *SPCH1* and *AUTS1* mapping da-

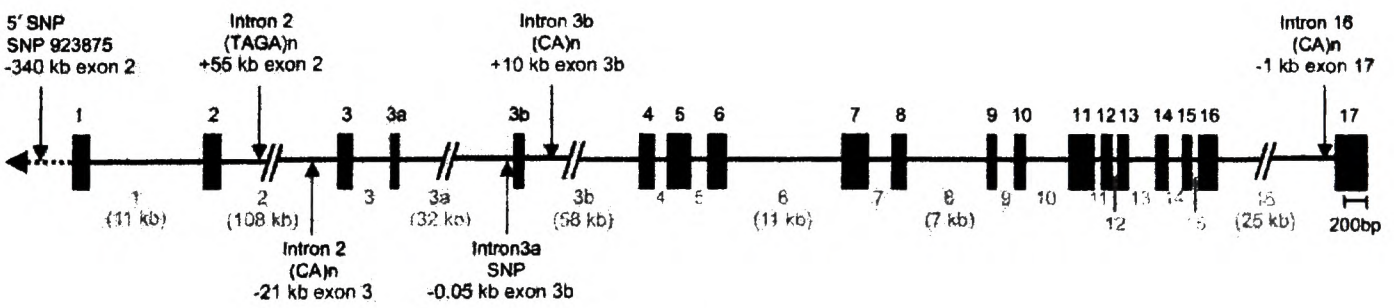


Figure 2 Schematic of *FOXP2* (adapted with permission from Lai et al. [2001]). Numbers in black indicate exon numbers. Numbers in grey indicate intron numbers as used in table 3. All exons are shown to scale. Introns are shown to scale with each other, and the sizes of all introns >5 kb are given in brackets (in kb). Positions of all microsatellites and SNPs used for association analysis are indicated by arrows, and distances (in kb) are given from the nearest coding exon. Exons 5 and 6 contain a polyglutamine encoding tract; exons 12–14 contain the forkhead (fox) domain; exons 3a and 3b are alternatively spliced; the KE mutation is found in exon 14; the CS translocation breakpoint is between exon 3b and exon 4.

Table 1

Association of *FOXP2* with Intronic Microsatellites: TDT within Autistic Families

Micosatellite[a] and Allele (Allele Size in bp)	Frequency[b] (%)	No. of Alleles								
		Paternal			Maternal			Combined		
		Transmitted	Not Transmitted	χ^2	Transmitted	Not Transmitted	χ^2	Transmitted	Not Transmitted	χ^2
Intron 2: (TAGA)n:										
3 (458)	50.7	41	52	1.30	56	47	.79	113	115	.02
2 (454)	25.9	48	36	1.71	38	42	.20	101	93	.33
4 (462)	16.5	30	25	.45	37	38	.01	70	66	.12
Other	6.8	14	20		10	14		24	34	
Intron 2: (CA)n:										
7 (195)	39.3	48	60	1.33	56	75	2.76	109	140	3.86
2 (185)	23.7	44	30	2.65	50	54	.15	96	86	.55
3 (187)	22.8	35	41	.47	48	37	1.42	86	81	.15
Other	14.2	37	33		41	29		78	62	
Intron 3b: (CA)n:										
7 (435)	30.2	46	44	.04	42	48	.4	93	97	.08
8 (437)	21.1	31	28	.15	36	36	.00	70	67	.07
9 (439)	14.9	20	34	3.63	26	21	.53	49	58	.76
6 (433)	13.4	28	19	1.72	23	19	.38	51	38	1.90
Other	20.7	37	37		36	39		74	77	
Intron 16: (CA)n:										
2 (225)	45.5	71	51	3.28	42	54	1.50	116	108	.29
8 (239)	27.2	38	37	.01	34	24	1.72	75	64	.87
7 (237)[c]	10.9	10	31	10.76	12	18	1.20	<u>22</u>	<u>49</u>	<u>10.27</u>
Other	16.2	36	36		35	27		73	65	

Note.—Empirical P values for transmission of all alleles was >.1 for alleles of all microsatellites, with the exception of intron 3b ([CA]n), for which $P = .08$.

[a] For positions of microsatellites, see figure 2.

[b] Alleles with a frequency <10% are grouped as "other."

[c] Allele-specific TDT for allele 7 combined (*underlined*): $P = .01$ (uncorrected).

ta reflects the involvement of a single gene—that is, *FOXP2*.

For the present study, 169 multiplex families with autism (857 individuals) were selected. These individuals represent the complete IMGSAC cohort (IMGSAC 2001*b*) and thus include those used in the original identification of *AUTS1* (IMGSAC 1998). Families were collected over four successive stages, and affection status was assessed using a variety of standardized tests. In brief, parents from families identified through an initial screen were administered the Autism Diagnostic Interview–Revised (ADI-R [Lord et al. 1994]) and the Vineland Adaptive Behaviour Scales (Sparrow et al. 1984). Potential cases were assessed using the Autism Diagnostic Observation Schedule (ADOS [Lord et al. 1994]) or the ADOS-Generic (ADOS-G [Lord et al. 2000]). Where possible, psychometric evaluation was conducted using Raven's progressive matrices (Raven 1989) or the Mullen Scales of Early Learning (Mullen 1995), as well as the British Peabody Picture Vocabulary Test III (Dunn and Dunn 1997) or an appropriate translation. A physical examination was undertaken and allowed the exclusion of children with signs of tuberous sclerosis. Where possible, affected individuals were karyotyped, and those found to have any chromosome abnormalities, including fragile X, were excluded.

In addition, we selected 43 families with language impairment (210 individuals) who form a subset of the SLI Consortium (SLIC) genome-screen sample (SLIC 2002). Probands were recruited by the Newcomen Centre at Guy's Hospital, London, through three special schools for language disorders and through Afasic, a support organization for people with developmental and language impairments. All probands, either currently or in the past, had language skills >1.5 SD below the nor-

Table 2

Association of *FOXP2* with Intronic Microsatellites: QTDT within Language-Impaired Families

Microsatellite	χ^2		
	ELS	RLS	NWR
Intron 2: (TAGA)n	2.38	.95	7.88
Intron 2: (CA)n	7.14	9.60	8.89
Intron 3b: (CA)n	8.23	10.18	7.69
Intron 16: (CA)n	2.58	1.30	3.11

Note.—All P values were >.1.

Table 3

Polymorphisms Detected in *FOXP2*

Intron/Exon[a] and Position[b]	Change	Frequency in Individuals with[c]	
		Autism ($n = 48$)	SLI ($n = 43$)
Intron 3:			
−102 bp from exon 3a	T/A	9	11
Intron 3a:			
−32 bp from exon 3b	A/G	0	1
−48 bp exon3b[d]	T/C	15	19
−68 bp from exon3b	G/A	1	0
Intron 5:			
+17 bp from exon5	T/G	8	10
Intron 5/Exon 6:			
...	Ins CAGCAG	0	1
Intron 11:			
+9 bp from exon11	T/C	1	1
−80 bp from exon12	A/G	0	1
Intron 13:			
+30 bp from exon13	C/G	1	2
Intron 14:			
+24 bp from exon14	T/C	0	1
−44 bp from exon15	T/G	2	0
−58 bp from exon15	T/C	3	0

[a] For intron numbers, see figure 2.

[b] Position is given in relation to nearest exon; "−" denotes that the SNP is found 5′ to the exonic sequence, and "+" denotes that the SNP is found 3′ to the exonic sequence.

[c] Frequency of heterozygotes within the sample tested.

[d] Polymorphism used to type individuals for association analysis.

mative mean for their chronological age, on the receptive and/or expressive scales of the Clinical Evaluation of Language Fundamentals (CELF-R) battery (Semel et al. 1992). Any proband or sibling found to have a nonverbal IQ of <80 was excluded from the sample. Additional exclusion criteria included an International Classification of Diseases–10th Revision/Diagnostic and Statistical Manual of Mental Disorders–4th Edition diagnosis of childhood autism.

The SLIC study used this sample in an investigation of three quantitative measures of language abilities, none of which showed linkage to chromosome 7q (SLIC 2002). A comprehensive description of the relevant cohorts can be found in reports by IMGSAC (2001*b*) and SLIC (2002).

To begin, we used RepeatMasker software (BCM Search Launcher Web site) to identify four novel intronic polymorphic microsatellites (fig. 2) that lie within BACs (NH0095P09, RG250D13, and NH563O05) covering the *FOXP2* sequence. Fluorescently labeled primers were designed (sequences are available from the authors on request) and were used to amplify all four microsatellites in available members of the families described above. The products were genotyped on ABI377 sequencing machines (PE Applied Biosystems), as described elsewhere (SLIC 2002), and were tested for association, as follows.

Children from the autistic families were classified as "affected" or "unknown" and were tested for association through use of a transmission/disequilibrium test (TDT) within the sib_tdt program from the ASPEX package version 2.2 (The ASPEX Linkage Analysis Package ftp site). This program calculates probabilities for χ^2 statistics by permuting parental alleles while fixing the identity-by-descent status of siblings within a family, thereby allowing the use of multiple siblings within a nuclear family.

We found no evidence for association within the autism group at the marker level ($P > .05$; table 1). Although allele-specific TDT indicated a weak association ($P = .01$) between the 237-bp allele of the intron 16 microsatellite and autistic disorder, the excess of nontransmitted alleles in this case (22 transmitted and 49 nontransmitted; see table 1) corresponds to a protective effect. Furthermore, after appropriate correction for multiple testing, this association was rendered nonsignificant.

The heterogeneity of the SLI phenotype rendered the derivation of a consistent affection status impractical. We therefore employed three quantitative measurements of language abilities to examine *FOXP2* association within the SLI cohort. The CELF-R was used to derive scores of expressive and receptive language abilities (ELS and RLS, respectively) (Semel et al. 1992). Each score is calculated from performance on three subtests designated to be primarily receptive or expressive in nature. The exact combination of individual tests used is dependent upon subject age. Additive raw scores from each segment are transformed to derive standardized scores with a mean of 100 and an SD of 15 in the general population calibration sample (Semel et al. 1992).

In addition, a test of nonword repetition (NWR) was used to assess phonological short-term memory (Gathercole et al. 1994). In this test, subjects are required to repeat tape-recorded nonsensical words of increasing length and complexity (e.g., "brufid" and "contramponist"). Studies show that individuals with current language impairments, as well as those who had language difficulties in early childhood which later resolved, perform poorly on this test (Gathercole et al. 1994; Bishop et al. 1999). In addition, it has been suggested that performance on the nonword repetition task is the best index of disorder in the KE pedigree (Vargha-Khadem et al. 1998).

Association was evaluated by the QTDT (quantitative transmission/disequilibrium test) program (Abecasis et al. 2000; QTDT Home Page), which employs a variance-components model that partitions association into between- and within-family components. The QTDT program includes a permutation framework, which allows the derivation of empirical *P* values for the sample being

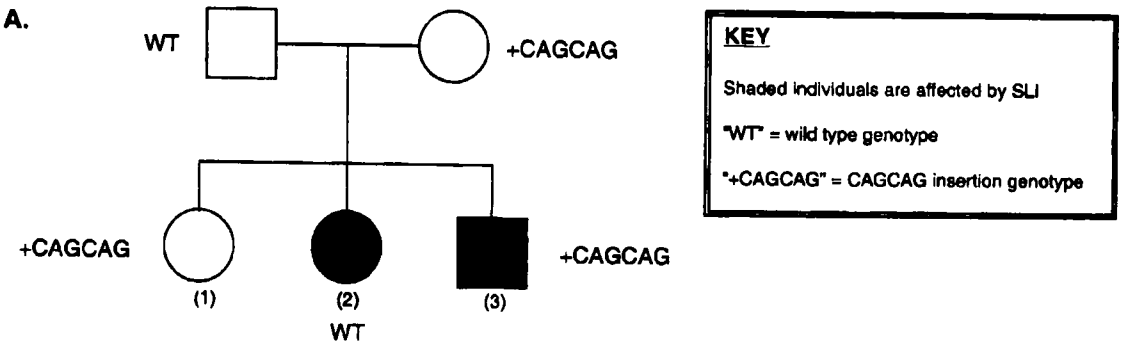


Figure 3 *FOXP2* exon 6. *A*, Family 43 pedigree. Both the mother and father have no reported history of language problems. Child 1 is clinically normal, child 2 is clinically affected, and child 3 has a reported language delay but is too young to test formally. *B*, The CAGCAG insertion in *FOXP2* exon 6. The boxed area represents the possible site for CAGCAG insertion. Underlined bases represent primers for exon 5. Exonic sequence is represented by capital letters; intronic sequence is represented by lowercase letters.

evaluated. This corrects for small sample sizes or deviations of quantitative traits from multivariate normality. We performed 1,000 simulations for each of the above traits and again found no evidence for association ($P > .05$) (table 2).

Despite the lack of positive findings from the association analyses, it remained possible that some individuals may harbor *FOXP2* mutations that were indiscernible at the whole-sample level. We therefore initiated a mutation screen of the *FOXP2* coding sequence by means of denaturing high performance liquid chromatography (DHPLC) (Kuklin et al. 1997).

DHPLC analysis was performed for all 43 probands of the families with SLI described above. For the autism cohort, a subset of 48 affected individuals were selected who contributed to the linkage peak on 7q. This enriched for individuals who are likely to carry etiological variants at the *AUTS1* locus.

Primers, taken from the Lai et al. (2001) study and available on request, were used to amplify all 19 *FOXP2*

Table 4

SNP Association Analyses of *FOXP2*: TDT within Autistic Families

		No. of Alleles								
		Paternal			Maternal			Combined		
SNP[a]	Frequency (%)	Transmitted	Not Transmitted	χ^2	Transmitted	Not Transmitted	χ^2	Transmitted	Not Transmitted	χ^2
Intron 3a	55.6	54	45	.82	58	57	.01	137	127	.38
5′ (923875)	48.5	31	39	.91	35	23	2.48	91	87	.09

NOTE.—All P values were >.1

[a] For SNP positions, see figure 2.

Table 5

SNP Association Analyses of *FOXP2*: QTDT within Language-Impaired Families

SNP[a]	Frequency (%)	χ^2 ELS	RLS	NWR
Intron 3a	51.0	.07	.81	.65
5′ (923875)	40.6	.28	2.03	.03

NOTE.—All *P* values were >.1

[a] For SNP positions, see figure 2.

exons and surrounding intron-exon boundaries, by means of a touchdown PCR protocol (PCR Protocol for WAVE Machine Web site). DHPLC analysis was performed using the WAVE DNA Fragment Analysis System (Transgenomic), and fragments that showed a variant elution pattern were directly sequenced.

We identified a total of 11 sequence variants, all of which were single-base substitutions within intronic regions (table 3). No changes were seen within exons 12–14, which contain the fox domain.

In one family with SLI (family 43), we identified a CAGCAG insertion within a polyglutamine stretch at the intron/exon border of exon 6 (fig. 3). This region represents the longest stretch of pure CAG repeats within the polyglutamine region and thus is the most likely place for an expansion to occur. We genotyped all available members of this family, but the expansion did not cosegregate with language impairment (fig. 3). In addition, the insertion may actually fall within intronic sequence—the exact position of the insertion remains unclear, because of the repetitive nature of this stretch of DNA—and it does not alter the reading frame of the sequence. It therefore appears likely that this expansion simply represents a rare polymorphism that is not relevant to the SLI phenotype.

The DHPLC analysis identified a common single-nucleotide polymorphism (SNP) within intron 3a of the *FOXP2* gene (fig. 2). This intron 3a polymorphism involves a T→C transition that destroys an *Afl*II site within the sequence and therefore could be directly typed via a restriction enzyme assay. One hundred sixty-nine families with autism were again tested through use of the sib_tdt program from the ASPEX package (v2.2), and 43 families with SLI were tested within the QTDT program (Abecasis et al. 2000), as described above. We found no evidence for association with this SNP, within either the autism or the SLI groups (tables 4 and 5).

Although the Lai et al. (2001) article presented the entire coding region of the *FOXP2* gene, the transcription start site could not be defined, and northern analyses indicated the existence of additional 5′ and/or 3′ untranslated exons that remain uncharacterized. Given the wide expression pattern of *FOXP2* (Lai et al. 2001), we postulated that it was likely to represent a housekeeping gene and therefore initiated a search for CpG islands in the genomic sequence immediately upstream of the *FOXP2* gene (CpG Islands Web site). Complementary bioinformatic analyses (e.g., Nix [MRC Human Genome Mapping Project Resource Center Web site] and Promoter Inspector [Genomatix Web site], and Ensembl [Ensembl Genome Server]) indicated that the closest CpG island lay 340 kb upstream of the present *FOXP2* coding sequence.

Similarity searches demonstrated that this sequence was highly homologous (83% identity) to a CpG-rich region upstream of the mouse *Foxp2* gene. Furthermore, this murine sequence was directly linked to the *Foxp2* coding region in three independent ESTs (AW490098, BB660527, and BB656124), indicating that it is transcribed as part of the mouse *Foxp2* mRNA.

As a final verification step, we therefore typed an SNP adjacent to this CpG island (SNP 923875) (dbSNP Home Page), through use of a restriction enzyme assay. We found no evidence for association at this 5′ SNP, in either the autism or the SLI groups (tables 4 and 5).

In the absence of any mutation or association evidence to suggest otherwise, we must therefore conclude that *FOXP2* is unlikely to play a major role in the onset of autism or SLI. As a corollary, since the autism cases studied here included those originally used in the detection of the *AUTS1* locus and the sample was enriched for individuals who showed linkage to 7q31, we can conclude that the *SPCH1* and *AUTS1* loci are attributable to different genes that, coincidentally, lie in similar positions on chromosome 7q.

Finally, it would appear that the role of *FOXP2* in speech and language disorders does not generalize to more common and genetically complex forms of language impairment within our SLIC cohort. However, it is worth noting that although the probands with SLIC were chosen to represent a diverse range of impairments spread over many linguistic domains, it remains possible that *FOXP2* variations may be involved in specific and distinct forms of speech and language impairments not represented within our sample.

Acknowledgments

Consortium members are as follows: *United Kingdom*—Centre for Social, Genetic and Developmental Psychiatry and Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London: Sarah Palferman, Nicola Matthews, Martha Turner, Janette Moore, Amaia Hervas, Anne Aubin, Simon Wallace, Janine Michelotti, Catherine Wainhouse, Alina Paul, Elaine Thompson, Ramyani Gupta, Claire Garner, Marianne Murin, Christine Freitag, Nicola Ryder, Emily Cottington, Jeremy Parr, Greg Pasco, Andrew Pickles, Michael Rutter, and Anthony Bailey; Wellcome Trust Centre for Human Genetics, University of Oxford: Janine A. Lamb, Gabrielle Barnby, Pat Scudder, Elena Bonora, Angela Marlow, and Anthony P. Monaco; Newcomen Centre, Guys Hospital, London: Gillian Baird and Anthony Cox; Regional Genetics Centre, Division of Med-

ical and Molecular Genetics, Guys Hospital, London: Zoe Docherty, Pamela Warburton, Elizabeth P. Green, and Stephen J. Abbs; Flemming Nuffield Unit, Newcastle: Ann Le Couteur, Helen R. McConachie, and Tom Berney; Neuropsychology Department, Newcastle General Hospital: Thomas P. Kelly; Developmental Psychiatry Section, University of Cambridge Clinical School: Petrus J. De Vries, Emma Gaitonde, and Patrick F. Bolton; Booth Hall Childrens Hospital, Manchester: Jonathan Green, Anne Gilchrist, and Jane Whittacker; and European Centre for Collection of Animal Cell Cultures, Salisbury: Bryan Bolton and Ros Packer. *Italy*—Dipartimento di Biologia, Universita di Bologna: Elena Maestrini, Francesca Blasi, and Elena Bacchelli. *The Netherlands*—AZU, Department of Child and Adolescent Psychiatry, Utrecht: Herman Van Engeland, Maretha V. De Jonge, Chantal Kemner, and Judith Timp. *Germany*—Deutsches Krebsforschungszentrum, Department of Molecular Genome Analysis: Sabine M. Klauck, Kim S. Beyer, Sabine Epp, and Annemarie Poustka; Deutsches Krebsforschungszentrum, Department of Biostatistics: Axel Benner; and J.W. Goethe–Universität, Department of Child and Adolescent Psychiatry, Frankfurt: Fritz Poustka, Dorothea Rühl, Gabriele Schmötzer, Sven Bölte, and Sabine Feineis-Matthews. *France*—Unité de Diagnostic et Evaluation de l'Autisme, Hôpital la Grave, Toulouse: Eric Fombonne, Bernadette Rogé, Jeanne Fremolle-Kruck, Catherine Pienkowski, and Marie-Thérèse Tauber. *Denmark*—Videnscenter Og Centre For Autisme, Virum: Lennart Pedersen, Torben Isager, Gunna Eriksen, and Demetrious Haracopos; John F. Kennedy Instituttet, Glostrup: Karen Brondum-Nielsen; Biophysics Group, Danish Technical University: Rodney M. J. Cotterill. *Greece*—Department of Psychological Paediatrics, Agia Sophia Childrens Hospital, Athens: John Tsiantis and Katerina Papanikolaou. *United States*—Department of Psychiatry, University of Chicago, Chicago: Catherine Lord, Christina Corsello, Stephen Guter, Bennett Leventhal, and Edwin Cook; UCLA Center for Neurobehavioural Genetics, Los Angeles: Susan L. Smalley, Stanley F. Nelson, Amy Liu, Janet Miller, Martha Dedricks, Lisa Chrzanowski, Julia Bailey, James McGough, and Jennifer Levitt; Child Study Center, Sterling Hall of Medicine, Yale University, New Haven: David Pauls, Fred Volkmar, Joel Bregman, Ami Klin, and John Alsobrook; and Department of Human Genetics, University of Pittsburgh, Pittsburgh: Daniel E. Weeks.

We would like to thank all the families who have participated in the study, as well as the professionals who continue to make this study possible. Thanks to all at the Newcomen Centre, for their assistance with data collection and management, and to all members of the Monaco group, for their support and advice. This work has been funded by The Wellcome Trust, the U.K. Medical Research Council, BIOMED 2 grant CT-97-2759, European Commission Fifth Framework Grant QLG2-CT-1999-0094, Telethon-Italy grant E.1007, the Janus Korczak Foundation, Deutsche Forschungsgemeinschaft, Foundation France Télécom, Conseil Régional Midi-Pyrénées, Danish Medical Research Council, Sofiefonden, the Beatrice Surovell Haskells Fond for Child Mental Health Research of Copenhagen, Danish Natural Science Research Council grant 9802210, National Institute of Child Health and Development grant 5-P01-HD-35482, and National Institutes of Health grants MO1 RR06022 GCRC NIH, NIH K05 MH01196, and K02 MH01389. D.F.N. is funded by a Medical Research Council Studentship, E.B. is funded by a University of Oxford Graduate Prize Studentship, and A.P.M. is a Wellcome Trust Principal Research Fellow.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

ASPEX Linkage Analysis Package, The, ftp://lahmed.stanford.edu/pub/aspex/index.html
BCM Search Launcher, http://searchlauncher.bcm.tmc.edu:9331/seq-util/seq-util.html (for RepeatMasker)
CpG Islands, http://www.ebi.ac.uk/emboss/cpgplot/
dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/
Ensembl Genome Server, http://www.ensembl.org/Homo_sapiens/
Genomatix: PromoterInspector, http://anthea.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *AUTS1* [MIM 209850], *SPCH1* [MIM602081], and *FOXP2* [MIM605317])
PCR Protocol for WAVE Machine, http://www.well.ox.ac.uk/genomics/wave.html (for DHPLC PCR protocol)
QTDT Home Page, http://www.well.ox.ac.uk/asthma/QTDT/index.html
U.K. MRC Human Genome Mapping Project Resource Center, http://www.hgmp.mrc.ac.uk/ (for Nix)

References

Abecasis GR, Cardon LR, Cookson WOC (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (DSM-IV). American Psychiatric Association, Washington, DC

Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh MB, Aylsworth AS, Vance JM, Gilbert JR, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (1999) Genetic studies of autistic disorder and chromosome 7. Genomics 61:227–236

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

Bishop DV (1994) Is specific language impairment a valid diagnostic criteria? Genetic and psycholinguistic evidence. Phil Trans R Soc Lond B Biol Sci 346:105–111

Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P (1999) Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. J Speech Lang Hear Res 42: 155–168

Bishop DVM, North T, Donlan C (1995) Genetic basis of specific language impairment: evidence from a twin study. Dev Med Child Neurol 37:56–71

Bishop DVM, Rosenbloom L (1987) Classification of childhood language disorders. In: Yule W, Rutter M (eds) Lan-

guage development and disorders. MacKeith Press, London, pp 16–41
Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, Rutter M (1994) A case-control family history study of autism. J Child Psychol Psychiat 35:877–900
Collaborative Linkage Study of Autism (1999) An autosomal genome screen for autism. Am J Med Genet 88: 609–615
Conti-Ramsden G, Crutchley A, Botting N (1997) The extent to which psychometric tests differentiate subgroups of children with SLI. J Speech Lang Hear Res 40:765–777
Dunn LM, Dunn LM (1997) Peabody picture vocabulary test. American Guidance Service, Circle Pines, MN
Fisher SE, Vargha-Khadem, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168–170
Folstein S, Rutter M (1977) Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry 18:297–321
Folstein SE, Mankoski RE (2000) Chromosome 7q: where autism meets language disorder? Am J Hum Genet 67: 278–281
Gathercole SE, Willis C, Baddeley AD, Emslie H (1994) The children's test of nonword repetition: a test of phonological working memory. Memory 2:103–127
International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. Hum Mol Genet 7: 571–578
——— (2001*a*) Further characterisation of the autism susceptibility locus *AUTS1* on chromosome 7q. Hum Mol Genet 10:973–982
——— (2001*b*) A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. Am J Hum Genet 69:570–581
Kuklin A, Munson K, Gjerde D, Haefele R, Taylor P (1997) Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. Genet Test 1:201–206
Lai CSL, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Curtis Jamison D, Green ED, Vargha-Khadem F , Monaco AP (2000) The *SPCH1* region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am J Hum Genet 67:357–368
Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A novel forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519–523
Lewis BA, Thompson LA (1992) A study of developmental speech and language disorders in twins. J Speech Hear Res 35:1086–1094
Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 30:205–223
Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview–Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Aut Dev Disord 24:659–685
Mawhood L, Howlin P, Rutter M (2000) Autism and developmental receptive language disorder—a comparative follow-up in early adult life I: cognitive and language outcomes. J Child Psychol Psychiat 41:547–559
Mullen EM (1995) Mullens scales of early learning. American Guidance Service, Circle Pines, MN
Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M, and the Paris Autism Research International Sibpair Study (1999) Genome-wide scan for autism susceptibility genes. Hum Mol Genet 8:805–812
Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, Rutter M (1995) Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. Am J Hum Genet 57:717–726
Rapin I (1997) Current concepts: autism. N Engl J Med 337: 97–104
Rapin I, Allen D (1983) Developmental language disorders: nosologic considerations. In: Kirk U (ed) Neuropsychology of language, reading and spelling. Academic Press, New York, pp 155–184
Raven J (1989) Standard progressive matrices. Australian Council for Educational Research, Victoria
Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, et al (1999) A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65: 493–507
Semel EM, Wiig EH, Secord W (1992) Clinical evaluation of language fundamentals–revised. Psychological Corporation, San Antonio
The SLI Consortium (2002) A genomewide scan identifies two novel loci involved in specific language impairment. Am J Hum Genet 70:384–399
Sparrow S, Balla D, Cicchetti D (1984) Vineland adaptive behaviour scales. American Guidance Service, Circle Pines, MN
Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, Bohman M (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiat 30:405–416
Tager-Flusberg H, Joseph R, Folstein S (2001) Current directions in research on autism. Ment Retard Dev Disabil Res Rev 7:21-29
Tomblin JB, Buckwalter PR (1998) Heritability of poor language achievement among twins. J Speech Lang Hear Res 41:188–199
Tomblin JB, Records NL, Zhang X (1996) A system for the diagnosis of specific language impairment in kindergarten children. J Speech Hear Res 39:1284–1294
Vargha-Khadem F, Watkins K, Alcock KJ, Fletcher P, Passingham RE (1995) Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc Natl Acad Sci USA 92:930–933
Vargha-Khadem F, Watkins K, Price CJ, Ashburner J, Alcock KJ, Connelly A, Frackowiak RSJ, Friston KJ, Pembrey ME, Mishkin M, Gadian DG, Passingham RE (1998) Neural basis of an inherited speech and language disorder. Proc Natl Acad Sci USA 95:12695–12700
Vincent JB, Herbrick J-A, Gurling HMD, Bolton PF, Roberts W, Scherer SW (2000) Identification of a novel gene on chro-

mosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. Am J Hum Genet 67:510–514

Warburton P, Baird G, Chen W, Morris K, Jacobs BW, Hodgson S, Docherty Z (2000) Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. Am J Med Genet 96: 228–234

World Health Organization (1993) The ICD-10 classification for mental and behavioural disorders: diagnostic criteria for research. World Health Organization, Geneva