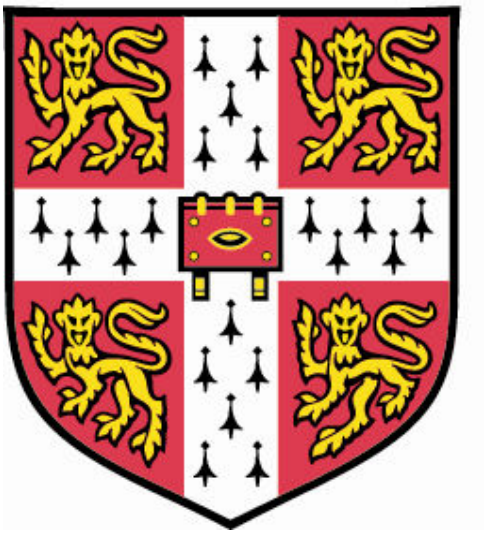


Copy Number Variation and Schizophrenia

by

Gloria Wing Chi TAM

Wellcome Trust Sanger Institute
Trinity College, University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

Nov 2008

Preface

This dissertation reports the work carried out at the Wellcome Trust Sanger Institute, between April 2005 and October 2008. It is submitted for the degree of Doctor of Philosophy, contains 248 pages (excluding bibliography and appendices), 58 figures and 20 tables and does not exceed the limit set by the Degree Committee.

This dissertation is my own work and contains nothing that is the outcome of work done in collaboration with others, except as specified in the text and below.

In section 3, oligonucleotide array hybridization experiments (for the *ABCA13* deletion) were performed in collaboration with Tomas Fitzgerald (Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK).

In section 4, whole-genome tiling path array hybridization experiments were performed in collaboration with Karen Porter (Wellcome Trust Sanger Institute), who performed approximately one third of the hybridization experiments. Affymetrix SNP array experiments and analysis were performed by the International Schizophrenia Consortium (Broad Institute, Boston, Cambridge MA) and the data was provided by Prof. Douglas Blackwood (Department of Psychiatric Genetics, University of Edinburgh, Edinburgh). Perl scripts for CNV genotyping were developed by Dr Richard Redon (Wellcome Trust Sanger Institute).

I hereby declare that I am the sole author of this dissertation and no part of the work contained in this dissertation has been previously submitted for any other degree.

_____

Gloria Wing Chi TAM

Abstract

Schizophrenia is a debilitating psychiatric illness affecting 1% of the population worldwide. The aetiology of schizophrenia is largely unknown, and deciphering schizophrenia genetics has remained a major challenge during the past decades in psychiatric research. In the past, visible alterations of the genome have been recognized as the underlying causes in a number of cognitive or behavioural defects. Structural chromosomal abnormalities, such as the 22q11 microdeletion and the Disrupted in Schizophrenia 1 (*DISC1*) translocation, were demonstrated to play a role in a proportion of schizophrenia cases. Furthermore, recent studies have identified a number of novel recurrent submicroscopic copy number changes significantly associated with schizophrenia (ISC 2008; Stefansson et al. 2008). This thesis describes a multi-faceted investigation to identify schizophrenia-related copy number variations (CNVs), defined as deletions and duplications larger than 1 kb in the genome.

As a first approach I performed array CGH on the whole-genome tiling path (WGTP) platform to screen for CNVs in three familial cases. Each pedigree consists of multiple patients affected with schizophrenia and other psychiatric illnesses. I identified a duplication on chromosome 1p36 common to all four affected members in one family, which was not identified in the normal HapMap controls (n=269). The CNV extends from the gene *H6PD* (Hexose-6-phosphate dehydrogenase precursor) to *SPSB1* (SPRY domain-containing SOCS box protein SSB-1). Using quantitative PCR, long range PCR and Fiber-FISH, I sequenced

the duplication breakpoint and delineated the structure of this potential pathogenic variant. Next, in a candidate-gene targeted approach I screened a multiplex schizophrenia family for CNVs in the gene *ABCA13* (ATP Binding Cassette Gene 13) at 7p12. I demonstrated the segregation of an intronic deletion with disease status.

Complementary to the family-based approach, I designed a population-based CNV study in schizophrenia versus matched control cohorts. Still using WGTP arrays, I performed a genome-wide screen for CNVs in 91 Scottish schizophrenia patients and 92 Lothian Birth Control DNA samples. In the WGTP dataset I identified a previously established schizophrenia-associated deletion at 15q11.2 (Stefansson et al. 2008) in a schizophrenia patient, near the *CYFIP1* (cytoplasmic FMR1 interacting protein 1 isoform) gene. I also identified a number of rare variants overlapping genes that are linked to various psychiatric diseases, including *SGCE* (sarcoglycan, epsilon), *OXTR* (Oxytocin) and *RCAN1* (Down Syndrome Critical Region 1). My results are consistent with recent reports demonstrating the role of rare CNVs in schizophrenia (ISC 2008, Walsh et al. 2008). In terms of common copy number variations, I genotyped 577 common CNVs using the WGTP data, and identified 31 candidates with putative bias in genotype distributions in cases versus controls. Two of these candidates, one at 3p26 and another at 15q13, were genotyped in an extended case-control cohort. Neither of them showed significant association with disease in the extended cohort.

The last approach was based on the hypothesis that CNVs could be linked to variations in learning, memory and brain function, both in the normal population and in psychiatric patients. The strategy involved a CNV screen on a set of proteins with important neuronal and synaptic functions. The NMDA receptor complex (NRC/MASC) was selected due to known roles of its components in cognitive and behavioural traits. Out of 186 NRC/MASC proteins, 20 of them showed CNVs in normal HapMap individuals. Four of these were linked to components of the core synaptic machinery, including a common CNV at *DLG1* (Discs, Large homolog 1 (Drosophila)). In addition, I investigated the multi-allelic variant at 17q21 near the gene N-ethylmaleimide-sensitive factor (*NSF*). I identified two major CNV blocks with interesting population bias, and identified for the first time a European-specific haplotype in an allelic variant known as H1.

Acknowledgements

Firstly, I would like to express my deepest gratitude towards my supervisors, Professor Seth Grant and Dr Nigel Carter. Their continuous support, advice and guidance have made my research life a lot more rewarding than it would have been. I am also particularly grateful to Dr Richard Redon, who has frequently inspired me on scientific ideas and technical knowledge, as well as communication skills and thought processes. My PhD studies were greatly enhanced by the generous support and patience from these three mentors.

I would also like to thank all members of the labs in Team 32 and Team 70, in particular Karen Porter and Tomas Fitzgerald, who have contributed to parts of the research work described in this thesis, as well as being supportive and joyful companions in the lab. I am also grateful to everyone who has provided me with advice, scientific discussions and technical support, including Dr Louie Van de Lagemaat, Dr Norboru Komiyama, Dr Fengtang Yang, Dr Jianxiang Chi (University of Oxford), Diane Rigler, Diana Rajan, Dr Douglas Strathdee, Kathryn Elsegood, Ellie Tuck, Dr Lianne Stanford and Charles Pettit.

I would like to express my appreciation to members of my thesis committee, Professor Nabeel Affara (Department of Pathology, University of Cambridge) and Professor Allan Bradley, for their scientific insights and constructive comments during my PhD reviews.

I would like to acknowledge my research collaborators at the University of Edinburgh, including Professor Douglas Blackwood, Professor Ian Deary, Dr. Walter Muir, Dr. Ben Pickard, Mary Malloy and Margaret Van Beck, for the provision of DNA samples, CNV data and the useful discussions on the clinical aspects of the project.

Special thanks go to the Wellcome Trust and Trinity College (University of Cambridge) for providing financial support. Last but not least, I am grateful to my family and all my friends for their endless love and support, and finally to Anson Ma for the wonderful years we have spent in Cambridge.

** Affiliation: Wellcome Trust Sanger Institute, unless otherwise specified in brackets.*

Gloria Tam

November 2008.

List of Figures

Figure 1.1 Types of genetic variants and their relative sizes……6

Figure 1.2 Comparative Genome Hybridization applied on a cancer cell line revealed amplification of the *myc* locus……10

Figure 1.3 Schematics of an array comparative genome hybridization experiment…. 13

Figure 1.4 Sensitivity and throughput of various CNV detection techniques…… 24

Figure 1.5 Non-allelic homologous recombination by low copy repeat…….. 29

Figure 1.6 Phenotypic effects of CNVs…… 31

Figure 1.7 Lifetime risk of developing schizophrenia based on relationship…….. 53

Figure 1.8 Genomic loci of the three novel recurrent deletions associated with schizophrenia…… 67

Figure 1.9 Schematic diagram of the synapse displaying known involvement of synaptic proteins in schizophrenia CNV loci…… 72

Figure 2.1 Amplified DNA samples after BioPrime Labelling Procedure……82

Figure 2.2 PCR genotyping of a deletion at 3p26 5' upstream of Close Homolog of L1 (*CHL1*) ……...92

Figure 2.3 Long Range PCR gel electrophoresis analysis of a tandem duplication at Chr1p36…….. 94

Figure 2.4 Dissociation curves analysis for different primer sets…… 97

Figure 2.5 A typical standard curve of the control primer…….. 97

Figure 2.6 Labelled DNA probes for Fluorescent In Situ Hybridisation…… 102

Figure 3.1 Three families analysed by whole-genome CNV screen…….. 108

Figure 3.2 qPCR validation of duplication 1p36 in Pedigree F-29……. 109

Figure 3.3 Delineation of 1p36.22 duplication structure by Fiber-FISH……...112

Figure 3.4 Sequencing breakpoints of 1p36.22 revealed repeat structures…….. 113

Figure 3.5 Array CGH detection of the 7p12 deletion for Patient 4398…….. 119

Figure 3.6 Array CGH profiles for all available members in Family 340……120

Figure 4.1 Case and control cohorts and CNV detection platforms used in our study…….. 128

Figure 4.2 Normalization and filtering steps applied to WGTP hybridization data before CNV analysis……130

Figure 4.3 WGTP array quality control indicators…….. 131

Figure 4.4 Frequency and types of CNVs and CNVRs in SCZ and LBC……... 133

Figure 4.5 Size and frequency distributions of CNVRs in Schizophrenia (SCZ) and Lothian Birth Control (LBC) Cohorts detected using the WGTP platform……... 134

Figure 4.6 CNVR gene content in the SCZ and LBC cohorts……..……...135

Figure 4.7 Correlation of CNV discovery rate with data quality in SCZ and LBC……. 136

Figure 4.8 WGTP data compared with 4 known schizophrenia CNV loci…….. 138

Figure 4.9 WGTP data detected a deletion at 15q11.2 in one patient……....140

Figure 4.10 WGTP CNV dataset validation strategies.……..……..... 142

Figure 4.11 Schizophrenia cohort-specific CNVRs containing brain-related or neuronal-related genes.……..……..……..……..…….. 149

Figure 4.12 Recurrent SCZ-specific CNVR regions detected by both WGTP and Affymetrix platforms.……..……..……..……..……...152

Figure 4.13 Vst scores to identify clones showing SCZ and LBC differentiation in the WGTP analysis……..……..……..……..……...159

Figure 4.14 Examples of bivariate clustering based on log2ratio of consecutive clones……..……..……..……..……....162

Figure 4.15 Three regions with significant difference of genotype distributions between SCZ and LBC as detected by bivariate clustering……..……..…… 164

Figure 5.1 A copy number polymorphism at chromosome 3p26.3 was detected by three BAC clones……..……..……..……..…….. 174

Figure 5.2 *CHL1* CNV genotypes from bivariate clustering……..…….. 175

Figure 5.3 Genomic location of the *CHL1* 5' CNV……..……..……...176

Figure 5.4 *CHL1* expression level against BAC clone log2ratio in HapMap samples..178

Figure 5.5 Genotyping of the *CHL1* 5' deletion polymorphism in HapMap samples… 179

Figure 5.6 *CHL1* gene expression against copy number of the *CHL1* 5' CNV in HapMap CEU samples……..……..……..……..……...180

Figure 5.7 The 15q13-14 genomic locus……..……..……..... 184

Figure 5.8 The three existing polymorphic structures of *CHRFAM7A* allele at 15q13-14……..……..……..……..……..…… 185

Figure 5.9 Copy number polymorphism at 15q13.2-13.3 spanning the gene *CHRFAM7A*……..……..……..……..……. 194

Figure 5.10 *CHRFAM7A* CNV genotype distributions in cases versus controls……... 195

Figure 5.11 Analysis of a typical Taqman qPCR experiment to determine *CHRFAM7A* genotypes. In each 96-well plate 40 samples (+2 controls) were genotyped……. 197

Figure 6.1 Schematic diagram of a glutamatergic excitatory synapse……..… 204

Figure 6.2 Log2ratio distributions for WGTP clones reporting CNV among 269 HapMap samples in 20 MASC regions……..……..……..…….. 207

Figure 6.3 CNVs affecting core components of the NRC/MASC signalling complex.. 212

Figure 6.4 Schematic representation of the 17q21 locus……..……..... 216

Figure 6.5 Array CGH genomic profiles of the 3 HapMap ethnic groups at chr17q21. 219

Figure 6.6 Population differentiation at two major CNV blocks at 17q21……....220

Figure 6.7 High resolution oligo array CGH profiles for 20 individuals at chr17 41.5 Mb-42.2 Mb……..……..……..……..……..…….. 222

Figure 6.8 Quantitative PCR validation of $CNV_{KIAA1267}$ and CNV_{NSF}…… 224

Figure 6.9 Fiber-FISH experiment to visualize copy number of CNV_{NSF}…… 224

Figure 6.10 Comparing SCZ and LBC samples at $CNV_{KIAA1267}$ and CNV_{NSF}……. 227

Figure 6.11 Locations of derived and ancestral loci of segmental duplications at 17q21……..……..……..……..……..……228

Figure 7.1 Identification of CNVs as disease risk loci for further characterization….. 247

List of Tables

Table 1.1 Two most influential diagnostic guidelines for schizophrenia the DSM-IV and ICD-10……49

Table 1.2 Schizophrenia candidate genes with evidence from linkage and association analysis…… 56

Table 1.3 Three novel recurrent deletions associated with schizophrenia…… 68

Table 2.1 Cycling protocol for general PCR amplification…… 90

Table 2.2 Cycling protocol for long-range PCR amplification…… 93

Table 2.3 Reagent mixture for each Taqman quantitative PCR assay…… 95

Table 2.4 Cycling protocol for Quantitative PCR amplification…… 96

Table 3.1 Known genomic rearrangement at 1p36.22…… 115

Table 3.2 Schizophrenia patients with DNA analysed on custom arrays…… 119

Table 4.1 Validated SCZ-specific rare variants……143

Table 4.2 SCZ-specific rare variants with genes associated with psychiatric disorders. …… 146

Table 4.3 Recurrent SCZ-specific variants as detected in 297 SCZ samples…… 151

Table 4.4 Validating 5 top Vst-regions using custom-design oligonucleotide array…..160

Table 5.1 *CHL1* 5' CNV distribution determined by the original genome-wide array CGH data…… 182

Table 5.2 *CHL1* 5' CNV distribution in an extended case-control cohort determined by PCR genotyping…… 182

Table 5.3 Previous reports associating the 15q13-14 locus to schizophrenia and related psychiatric disorders…… 189

Table 5.4 *CHRFAM7A* CNV genotype distributions in the extended case and control cohorts…… 198

Table 6.1 CNVs detected at 20 genes encoding NRC/MASC signalling complex components……206

Table 6.2 CNV genotypes for all unrelated HapMap individuals at $CNV_{KIAA1267}$ and CNV_{NSF}…… 225

Table 6.3 SCZ and LBC genotype counts at $CNV_{KIAA1267}$ and CNV_{NSF}…… 227

List of Abbreviations

ADEOAD	Autosomal Dominant Form of Early-Onset Alzheimer Disease
Alu	a family of repeat elements named after the *AluI* restriction site
AS	Angelman Syndrome
ASD	Autism Spectrum Disorder
BAC	Bacterial Artificial Chromosome
bp	Base Pairs
CAA	Cerebral Amyloid Angiopathy
CD-CV	Common-Disease Common-Variant
CD-RV	Common-Disease Rare-Variant
CEU	HapMap DNA: Utah samples with European ancestry
CGH	Comparitive Genome Hybridisation
CHB	HapMap DNA: Chinese samples with Asian ancestry
CNP	Copy Number Polymorphism
CNV	Copy Number Variation
CNVR	Copy Number Variation Region
COS	Childhood-Onset
cR	combined ratio
Cy3	Indocarbocyanine
Cy5	Indodicarbocyanine
DECIPHER	Database of Chromosomal Imbalance
DGS	DiGeorge Syndrome
DGV	Database of Genomic Variants
DISC1	Disrupted In Schizophrenia 1
DNA	Deoxyribonucleic Acid

dNTP	Deoxynucleoside Triphosphate
DOP-PCR	Degenerate Oligonucleotide Primed Polymerase Chain Reaction
DSBs	Double Stranded Breaks
DSM-IV	Diagnostic and Statistic Manual of Mental Disorder- 4th Edition
ECS	Electroconvulsive Shocks
ESP	Clone-End Sequence-Pair
EtOH	Ethanol
FISH	Fluorescent In Situ Hybridisation
GAD	Genetic Association Database
G-banding	Giemsa-banding
GWAS	Genome-Wide Association Studies
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
INDEL	Insertions and Deletions in a chromosome
ISC	International Schizophrenia Consortium
JPT	HapMap DNA: Japanese samples with Asian ancestry
kb	Kilo Base (One Thousand Base Pairs)
LBC	Lothian Birth Control Cohort
LCR	Low Copy Repeat
LD	Linkage Disequilibrium
LINE	Long Interspersed Nuclear Element
LOD	Logarithm of Odds
LOH	Loss of Heterozygosity
LTP	Long-Term Potentiation
M	Molar
MAGUK	Membrane-Associated Guanylate Kinase

MAPH	Multiplex Amplifiable Probe Hybridisation
MAQ	Multiplex Amplicon Quantification
MASC	MAGUK Associated Signaling Complex
Mb	Mega Base
min	Minute
MLPA	Multiplex Ligation-Dependent Probe Amplification
MR	Mental Retardation
MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
NAHR	Non-Allelic Homologous Recombination
NHEJ	Non-Homologous End Joining
NMDA	N-Methyl-D-Aspartate
NRC	NMDA Receptor Complex
nt	Nucleotide
OMIM	Online Mendelian Inheritance In Man
OR	Odds Ratio
PBS	Phosphate-Buffered Saline
PCP	Phencyclidine
PCR	Polymerase Chain Reaction
PEM	Paired-End Mapping
PET	Poisitron Emission Tomography
PFGE	Pulsefield Gel Electrophoresis
PPI	Prepulse Inhibition
PSD	Postsynaptic Density
PWS	Prader Willi Syndrome

qPCR	Quantitaive PCR
RACE	Rapid Amplification of cDNA Ends
rcf	Relative Centrifugal Force
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
SCZ	Schizophrenia Cohort
SD	Segmental Duplication
SDe	Variability Measure for Array CGH Experiments
SINE	Short Interspersed Nuclear Element
SKY	Spectral Karyotype
SNP	Single Nucleotide Polymorphism
Tm	Melting Temperature
VCFS	Velo-Cardio-Facial Syndrome
VNTR	Variable Number Of Tandem Repeat
Vst	A varinace-based measure to compare quantitative data from different cohorts
WGTP	Whole Genome Tiling Path
YRI	HapMap DNA: Yoruba samples with African ancestry

CHAPTER 1 INTRODUCTION

1.1 COPY NUMBER VARIATION (CNV) AS A SOURCE OF GENETIC DIVERSITY 2

1.2 DETECTION OF COPY NUMBER VARIATION 7

1.2.1 Classical Cytogenetic Techniques for the Detection of Structural Variations 7
1.2.2 Array-based Comparative Genome Hybridization 11
1.2.2.1 BAC Array CGH 12
1.2.2.2 Oligonucleotide Array CGH 14
1.2.2.3 Choice of Reference DNA for Array Hybridization 15
1.2.3 Genotyping CNVs using Single Nucleotide Polymorphisms 16
1.2.3.1 Linkage Disequilibrium Based Tag-SNP Approach 16
1.2.3.2 CNV Genotyping Using SNP Arrays 17
1.2.4 Validation and Detection of CNV at Targeted Loci 19
1.2.4.1 Quantitative Fluorescent Real-time PCR (qPCR) 19
1.2.4.2 Multiplex Quantitative Fluorescent Real-time PCR 20
1.2.4.3 Other Methods 20
1.2.5 Genome Sequencing and CNV Detection 22
1.2.5 Genome Sequencing and CNV Detection 22
1.2.5.1 Clone End Mapping and Sequencing 22
1.2.5.2 Novel Sequencing Technology and Genome Comparison 23

1.3 MECHANISMS OF COPY NUMBER VARIATION GENERATION 25

1.3.1 Non-allelic Homologous Recombination 25
1.3.2 Non-homologous End Joining 26
1.3.3 Other Mechanisms 27
1.3.4 Insights from Breakpoint Mapping 27

1.4 BIOLOGICAL IMPACT OF COPY NUMBER VARIATION 30

1.4.1 Phenotypic Effect of CNVs 30
1.4.1.1 CNV and Expression 32
1.4.1.2 CNV and Gene Disruption 34
1.4.2 Phenotypic Variations and Evolution 35
1.4.2.1 CNV and Human Traits 35
1.4.2.2 Positive and Negative Selections on CNV 36
1.4.3 CNV and Disease 38
1.4.3.1 Genome Disorder 38
1.4.3.2 Rare CNVs in Mendelian Disease Traits 40
1.4.3.3 Common CNVs in Multifactorial or Complex Disease 41

1.5 SCHIZOPHRENIA 45

1.5.1 The Concept of Schizophrenia 45
1.5.2 Phenotypes and Diagnosis 47

1.5.2.1 Positive and Negative Symptoms 47
1.5.2.2 Endophenotypes 47
1.5.2.3 Course of Illness 48
1.5.2.4 Standardised Diagnostic Methods 48
1.5.3 Aetiology and Neurobiology 50
1.5.3.1 Dysfunction of Neurotransmitter Systems 50
1.5.3.2 Neurodevelopment and Neuropathology 52
1.5.4 The Genetics of Schizophrenia 53
1.5.4.1 Evidence of Genetic Contribution 53
1.5.4.2 A Search for Candidate Genes 54
1.5.4.3 Genes and Environment 57

1.6 CNV IN SCHIZOPHRENIA AND OTHER PSYCHIATRIC DISEASES 58

1.6.1 Early Studies on Chromosomal Abnormalities in Schizophrenia 58
1.6.1.1 DISC1 and Breakpoint Study in Schizophrenia 58
1.6.1.1 22q11 Microdeletion and Schizophrenia 60
1.6.2 Large Scale CNV Screen in Schizophrenia Patients 63
1.6.1.1 Summary of CNV Findings 63
1.6.2.1 Identification of Large Recurrent Schizophrenia Loci 65
1.6.2.2 Increased Mutation Burden of CNV in Schizophrenia Patients 69
1.6.2.3 Rare Variants Converging into Neurodevelopmental Pathways 70

1.7 SCOPE OF THESIS 77

CHAPTER 2 MATERIALS AND METHODS

2.1 ARRAY COMPARATIVE GENOME HYBRIDIZATION (ARRAY CGH) 80

2.1.1 Patient and Control DNA Samples 80
2.1.2 DNA Labelling 81
2.1.3 Sample Precipitation and Preparation 82
2.1.4 Array Hybridisation 83
2.1.5 Image Acquisition and Data Analysis 83

2.2 AGILENT OLIGO CUSTOM-DESIGNED ARRAY CGH 86

2.2.1 Custom Array Designs 86
2.2.2 DNA Labelling 86
2.2.3 Preparation of Labelled Genomic DNA 87
2.2.4 Array Hybridisation 87
2.2.5 Slide Washing 88
2.2.6 Image Acquisition and Data Analysis 89

2.3 POLYMERASE CHAIN REACTION (PCR) 90

2.3.1 PCR genotyping of the 3p26 Deletion near *CHL1* 90

2.3.2 Long Range Polymerase Chain Reaction (LR-PCR) 92
2.3.3 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR) 94
2.3.3.1 SYBR Green Method 94
2.3.3.2 Taqman Method with MGB Probes 95
2.3.3.3 Thermal Cycler and Reaction Condition 95
2.3.3.4 Standard Curve Generation 96
2.3.3.5 DNA Quantification and Data Analysis 98

2.4 FLUORESCENCE IN-SITU HYBRIDISATION (FISH) 99

2.4.1 Growing Cell lines 99
2.4.2 Preparation of Extended Chromatin Fibre Slides 99
2.4.3 Preparation of Fosmid Clone Insert DNA 100
2.4.4 Amplification and Labeling of DNA probes 101
2.4.4.1 GenomePlex® Whole Genome Amplification 101
2.4.4.2 DNA Fluorescent Labelling 101
2.4.4.3 Probe Fragmentation 102
2.4.5 Immunofluorescence and Image Acquisition 103

CHAPTER 3 FAMILIAL STUDY IN SCHIZOPHRENIA

3.1 WHOLE-GENOME CNV SCREEN IN FAMILIES 106

3.1.1 Whole-genome Array CGH Screen in Three Familial Cases 106
3.1.2 Characterization of the Rare Familial Duplication at 1p36.22 110
3.1.3 Known Genomic Rearrangements Near the 1p36.22 duplication 114

3.2 DELETION AT *ABCA13* IN AN EXTENDED FAMILY WITH SCHIZOPHRENIA 116

3.2.1 Evidence of Functional Mutations of *ABCA13* at 7p12.3 116
3.2.2 Oligo Array CNV Screen in an Extended Pedigree 118

3.3 CHAPTER SUMMARY AND DISCUSSION 122

CHAPTER 4 POPULATION-BASED CNV STUDY IN SCHIZOPHRENIA

4.1. EXPERIMENTAL DESIGN AND ARRAY DATA QUALITY CONTROL ..127

4.1.1 Case-Control CNV Screen Experimental Design 127
4.1.2 WGTP Array Data Quality Control 129

4.2 COPY NUMBER VARIATION DETECTION ON THE WGTP ARRAY 132

4.2.1 Distribution of CNVs and CNV Regions in Case and Control Cohorts..... 132
4.2.2 Bias of CNV Discovery Rate in SCZ Versus LBC 135

4.3 COMPARING WGTP DATA WITH KNOWN SCHIZOPHRENIA CNV REGIONS.....137

4.4 RARE VARIANTS SPECIFIC TO THE SCHIZOPHRENIA COHORT 141

4.4.1 Rare Variant Detection Using Consecutive Clone Calling Criteria..... 141
4.4.2 Validation of SCZ-Specific Rare Variants..... 142
4.4.3 Rare Variants in SCZ with Genes Involved in Psychiatric Disorders 144

4.5 RECURRENT SCZ-SPECIFIC VARIANTS IN EXTENDED COHORT 150

4.5.1 A CNVR at Down Syndrome Critical Region 1 (RCAN1/DSCR1)..... 153
4.5.2 A Variant Near Olfactomedin1 and other Recurrent CNVRs 154

4.6 FREQUENT COPY NUMBER VARIATIONS IN SCZ AND LBC 156

4.6.1 Variance-based Clone-by-Clone Cohort Comparison 156
4.6.2 CNV Genotyping with Bivariate Clustering..... 161

4.7 CHAPTER SUMMARY AND DISCUSSION.....165

CHAPTER 5 GENOTYPING TWO SCHIZOPHRENIA CNVS IN AN EXTENDED CASE CONTROL COHORT

5.1 CANDIDATE I: *CHL1* (CLOSE HOMOLOG OF L1) AT 3P26 170

5.1.1 Functional Significance of *CHL1*..... 170
5.1.2 Structure of the *CHL1* 5' Deletion Polymorphism 173
5.1.3 Correlation of *CHL1* CNV to mRNA Expression..... 177
5.1.4 Genotyping CHL1 CNV in a Larger Case-Control Cohort..... 181

5.2 CANDIDATE II: *CHRFAM7A* (CHRNA7-FAM7A FUSION GENE) AT 15Q13 183

5.2.1 Genomic Architecture at chromosome 15q13-14 183
5.2.2 Molecular Genetic Studies linking 15q13-14 to Schizophrenia 187
5.2.3 Functional Significance of *CHRNA7* and *CHRFAM7A* 191
5.2.4 Copy Number Polymorphism at the *CHRFAM7A* Region 192

5.3 CHAPTER SUMMARY AND DISCUSSION 199

CHAPTER 6 CNVS AND THE NMDA RECEPTOR COMPLEX

6.1 NMDA RECEPTOR COMPLEX, SCHIZOPHRENIA AND COGNITION .. 203

6.2 COPY NUMBER VARIATION AND THE NMDA RECEPTOR COMPLEX 205

6.3 CNV AT 17Q21 NEAR N-ETHYLMALEAMIDE-SENSITIVE FACTOR (NSF) 213

6.3.1 Introduction to the 17q21 locus near *NSF* 213
6.3.2 Known Genomic Structure of 17q21 215
6.3.3 Array CGH Data Reveals Two Major CNVs at 17q21 217
6.3.3.1 $CNV_{NSF:}$ Copy number Variant at 5' end of NSF 217
6.3.3.2 $CNV_{KIAA1267:}$ Copy number Variant at KIAA1267 218
6.3.4 Resolving CNVKIAA1267 with SNP and High Resolution Oligo Array Data 221
6.3.5 Validating CNVNSF and CNVKIAA1267 by qPCR and FISH 223
6.3.6 Genotyping HapMap Individuals for 17q21 Structural Variants 225
6.3.7 CNVNSF and CNVKIAA1267 in Schizophrenia Versus Control 226
6.3.8 Evolutionary History of CNVNSF and CNVKIAA1267 228

6.4 CHAPTER SUMMARY AND DISCUSSION 229

CHAPTER 7 GENERAL DISCUSSION

7.1 DESIGN OF CNV DISCOVERY AND ASSOCIATION STUDY 232

7.1.1 Multiple Approaches of CNV Study 232
7.1.2 Future Large-Scale CNV Studies 234
7.1.3 Enrichment of Schizophrenia Subtypes 235

7.2 UNDERSTANDING THE GENETIC MODEL OF SCHIZOPHRENIA 237

7.2.1 Rare Variants Versus Common Variants 237
7.2.2 Incomplete Penetrance and Expressivity 239
7.2.3 Aetiological Overlap of Schizophrenia with Other Psychiatric Diseases 239

7.3 CLINICAL RELEVANCE OF CNVS IN SCHIZOPHRENIA.....241

7.3.1 CNV Findings Translating into Disease Classification.....241
7.3.2 Genetic Counselling and Therapeutic Potential 242

7.4 THESIS SUMMARY 244

7.5 FUTURE DIRECTION 246

REFERENCE……249
APPENDICES……278

Chapter 1
Introduction

1.1 Copy Number Variation (CNV) as a Source of Genetic Diversity

Copy number variation (CNV) has long been studied as gene copy number difference among individuals at specific loci, for example in the α- and β- globin gene locus; or as genomic imbalance resulting in diseases such as Prader-Willi and Angelman Syndrome. Until recently, such rearrangements in the human genome were assumed to be limited in scale. Since 2004, however, research has revealed CNVs as a major source of human genetic variation (Iafrate et al. 2004; Sebat et al. 2004; Tuzun et al. 2005; Conrad et al. 2006; McCarroll et al. 2006; Redon et al. 2006). The unprecedented level of genetic diversity conferred by CNVs has opened a new chapter in our understanding of phenotypic variation, human evolution and disease susceptibility.

Copy number variation is defined as deletion and/or duplication of a DNA sequence larger than 1 kb in length (Feuk et al. 2006a; Freeman et al. 2006). It belongs to a spectrum of genetic variations ranging from the ubiquitous single nucleotide polymorphisms (SNPs), to fine-scale copy number changes such as small insertions and deletions (INDELs), microsatellite and minisatellite repeats, to larger scale structural variations such as inversions, translocations and CNVs (Bowcock et al. 1994; Armour et al. 1996; IHMC 2005; Feuk et al. 2006a; Mills et al. 2006; Conrad and Hurles 2007) (Figure 1.1). It is these variations and polymorphisms that constitute the dynamic human genome architecture and underlie the differences between individuals.

With remarkable advance in microarray technologies, and the availability of a complete human genome sequence, it became possible to obtain genome-wide maps of the locations and frequencies of CNVs. The first reports of wide-spread copy number variations or copy number polymorphisms (CNPs) (CNVs at >1% frequency) appeared in 2004. Iafrate *et al.* and Sebat *et al.* independently surveyed the human genome for copy

number changes, revealing the extent of this source of human genetic variation at a previously unanticipated level (Iafrate et al. 2004; Sebat et al. 2004). The two microarray-based CNV studies detected a total of 476 CNV loci from 75 individuals. Some of these loci affect genes that play important biological roles such as neurological functions and metabolism (Sebat et al. 2004). Since then, more copy number changes were mined from SNP genotyping data and clone paired-end sequencing data (Tuzun et al. 2005; Conrad et al. 2006; McCarroll et al. 2006).

In 2006, a comprehensive map of copy number variations was released based on 270 individuals from four different ethnic populations of European, Asian or African descents, who were originally included in the International HapMap Project. Redon *et al.* reported a total of 1447 copy number variable regions (CNVRs), corresponding to 360 Mb of human DNA sequences or 12% of the human genome (Redon et al. 2006). The CNVRs are enriched in functional categories such as cell adhesion, sensory perception of smell and of chemical stimulus, as well as neurophysiological processes. Distribution of CNVRs is non-uniform along the genome, with copy number changes preferentially clustered near segmental duplications (SDs), defined as duplicated sequences of >1 kb with 90% or more sequence identity in the reference human genome assembly (Bailey et al. 2002). Multi-allelic and complex CNVs are especially enriched in these SD regions. A number of following reports targeting specifically at segmental duplication regions also defined these loci as hotspots for chromosomal rearrangement and copy number variations (Sharp et al. 2005; Locke et al. 2006).

In the past years, reports of more comprehensive CNV discovery and characterization have come into view (de Smith et al. 2007; Pinto et al. 2007; Simon-Sanchez et al. 2007; Wang et al. 2007; Wong et al. 2007; Zogopoulos et al. 2007; Cooper et al. 2008). In

particular, recent studies on fine-scale CNV detection based on novel sequencing technologies and genome comparisons provide us with new perspectives on CNVs and structural variations (Khaja et al. 2006; Korbel et al. 2007; Levy et al. 2007; Kidd et al. 2008). These studies have revealed that individual human genomes are at least 0.5% different, with the order of 600-900 CNVs between any two individuals, contrary to our traditional view of human sharing 99.9% similarity (Korbel et al. 2008).

The phenomenon of copy number variations is not unique to the human genome. CNVs have been characterized in a number of mammalian species from the great apes (Locke et al. 2003; Perry et al. 2006; Wilson et al. 2006b; Perry et al. 2008) to rat and mouse (Li et al. 2004; Adams et al. 2005; Snijders et al. 2005; Egan et al. 2007; Guryev et al. 2008; She et al. 2008). CNV maps for the Drosophila genome was also recently published (Dopman and Hartl 2007; Emerson et al. 2008; Zhou et al. 2008), cataloguing differences in Drosophila's gene copy numbers, a concept that was developed in Drosophila genetics as early as the 1930s (Bridges 1936).

Whilst early CNV detection platforms mature and new technologies emerge, the discovery and characterization of CNVs has surfaced as an exciting branch of human genetics. Copy number changes were increasingly explored in the context of population demographics and human evolution (Conrad and Hurles 2007), with studies based on inter-population comparisons (Redon et al. 2006; Jakobsson et al. 2008) and inter-species evaluations (Locke et al. 2003; Perry et al. 2008). Population bias in copy numbers at specific loci have been demonstrated to play an important role in environmental adaptations, for example the evolution of CNV in the salivary amylase gene locus has been influenced by human diet (Perry et al. 2007); and the HIV-1 susceptibility CNV locus *CCL3L1* showed variable distribution in different populations

(Gonzalez et al. 2005). Of equal importance, CNV has brought a new dimension to disease genetics. Analogous to SNPs, an instrumental tool in discovering disease genes through linkage and association analysis, CNVs is a major source of genetic variation with potential clinical relevance in a number of diseases (see Chapter 1.4).

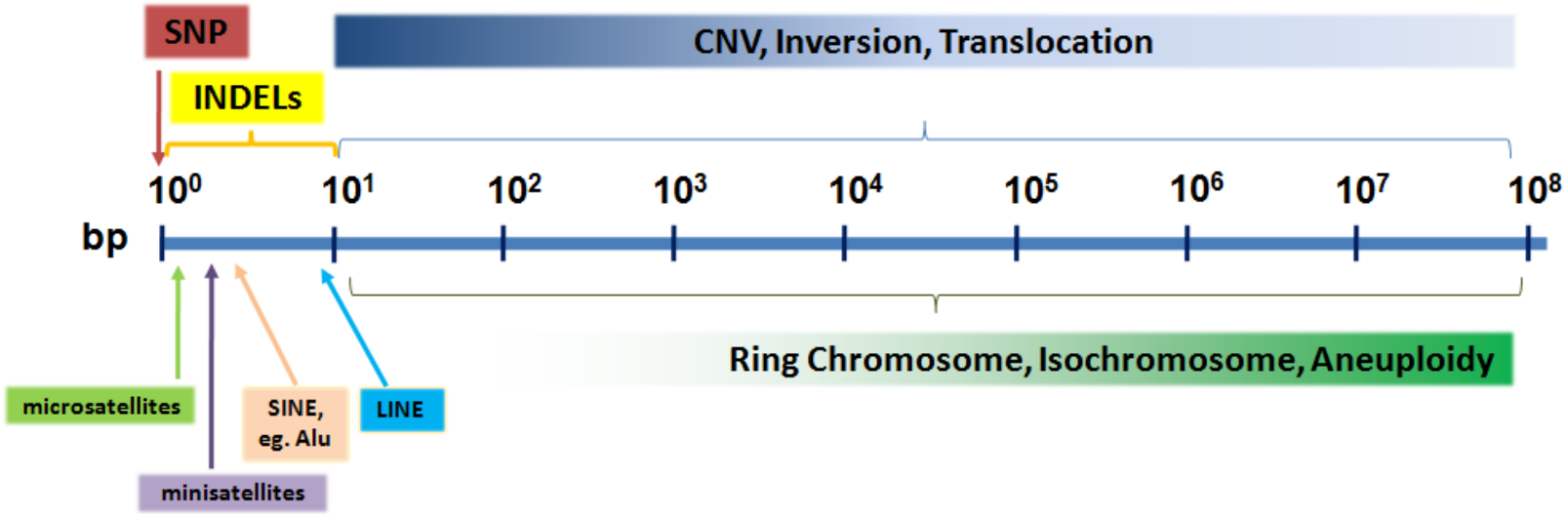


Figure 1.1 Types of genetic variants and their relative sizes. (SNP: Single Nucleotide Polymorphism; CNV: Copy Number Variation; LINE: Long Interspersed Nuclear Element; SINE: Short Interspersed Nuclear Element; Alu: a family of repeat elements named after the *Alu*I restriction site; bp: base pair)

1.2 Detection of Copy Number Variation

1.2.1 Classical Cytogenetic Techniques for the Detection of Structural Variations

Classical cytogenetics and traditional karyotype techniques such as Giemsa-banding (G-banding) (Craig and Bickmore 1993) have revealed gross structural variations as microscopically visible alterations. Cases of large-scale chromosomal rearrangements leading to Down syndrome (trisomy 21) and cri-du-chat syndrome (terminal 5p deletion) were discovered in the 1950-60s, many decades before the introduction of high-resolution techniques for rearrangement detection.

Fluorescence in situ hybridisation (FISH) further increased the resolution of cytogenetic techniques to 104–106 bp, extending the reach of traditional karyotyping to the detection of submicroscopic copy number changes. FISH (Bauman et al. 1980) is an in situ hybridisation technique in which a labelled probe of specific DNA sequences (e.g. a Bacterial Artificial Chromosome (BAC)/fosmid clone) is hybridized to a preparation of metaphase chromosomes or interphase DNA, usually attached to a glass slide. The chromosomal DNA and probe mixture was then denatured, allowing the single-stranded probe and single stranded DNA to re-anneal, with the probe hybridizing to the complementary sequences on the DNA and reformed a double stranded molecule. Following hybridization, unbound probes were washed away, and the hybridized probes were visualized directly if they were tagged with fluorochromes (e.g. Cyanine (Cy) or Alexa Fluors dyes), or detected by antibodies against hapten-tagged probes, or affinity agents such as avidin or streptavidin if probes were labelled with the biotin and digoxigenin systems (Langer, 1981).

FISH techniques have evolved throughout the years. For instance, improvement on probe labelling has led to the development of multi-probe FISH and spectral karyotype (SKY) (Schrock et al. 1996). These multi-colour hybridizations allow visualizing of rearrangements involving different chromosomes. On the other hand, the DNA preparation method has also progressed such that extended chromatin fibres (as a replacement for interphase or metaphase preparation) were fixed onto glass slides for high-resolution detection of submicroscopic changes, a technique known as Fiber-FISH (Florijn et al. 1995). This allows visualization of small copy number changes, for example deletion gaps or tandem duplications down to the size of a fosmid clone (40 kb) or even smaller. The technique can also be used to solve more complex rearrangements, as in the case of the expansion of CCL3L1 and related segmental duplication in human and chimpanzees.

Along with the development of FISH, complementary techniques such as comparative genome hybridization (CGH) (Kallioniemi et al. 1992) were developed for the detection of structural rearrangement. In CGH, differentially labelled test and reference DNA were simultaneously hybridized to chromosome spreads. Unlabelled Cot-I DNA was also co-hybridized to block DNA repeats. The hybridisation was then detected with two fluorochromes, and genomic regions of gains or losses would be detected as changes in the ratio of intensities of the two fluorophores along the chromosome. The technique was widely used in the analysis of tumor malignancies and constitutional chromosomal aberrations (Forozan et al. 1997).

The major advantage of CGH is that it allowed whole-chromosome or whole-genome surveys of chromosomal rearrangement and aberrations, compared to previous target-specific approaches. Figure 1.2 shows the image of a CGH experiment applied on a

cancer cell line when the technique first developed. High-level amplification of the *myc* locus at 8q24 was revealed as an elevated ratio of test/reference signal intensity. Nevertheless, such CGH approaches remained limited in resolution by the use of metaphase chromosomes as DNA hybridization targets. Therefore, higher resolution detection would still require laborious locus-by-locus techniques, such as Southern blot analysis (Southern 1975) or pulse field gel electrophoresis (PFGE) (Schwartz et al. 1983; Herschleb et al. 2007) (see section 1.2.5).

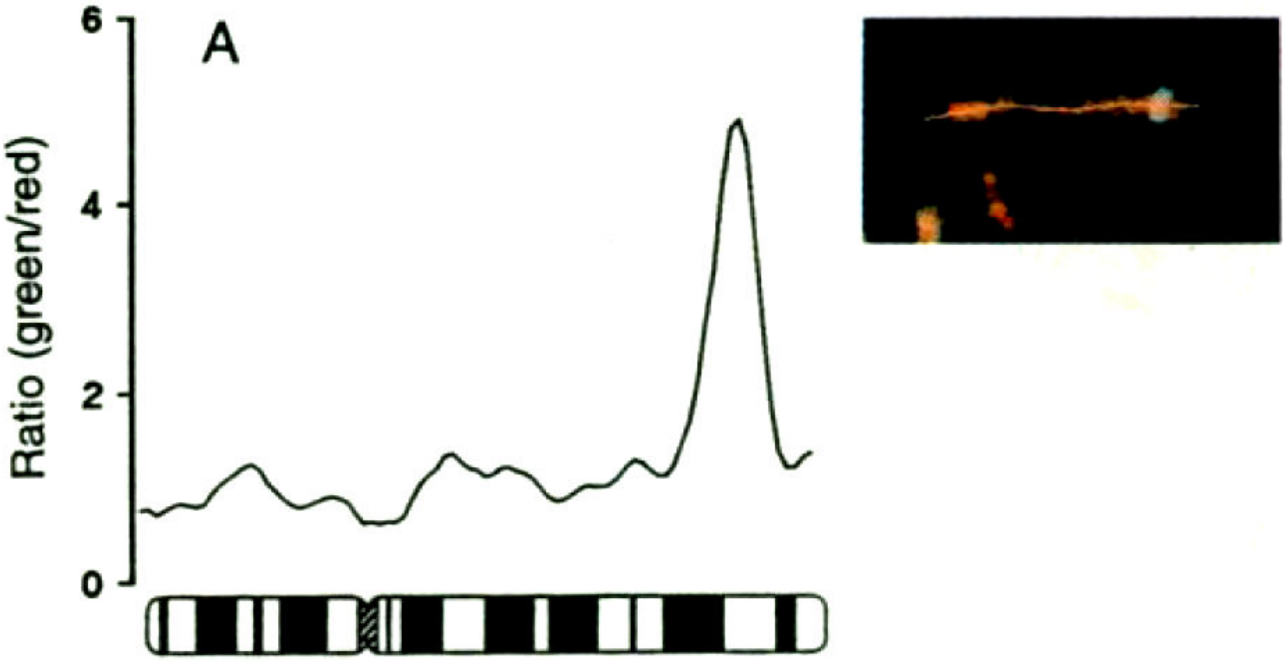


Figure 1.2 Comparative Genome Hybridization applied on a cancer cell line revealed amplification of the *myc* locus. (Diagram reproduced from Kallioniemi *et al.* 1992, *Science* **258**(5083))
Kallioniemi *et al.* performed a CGH experiment on cell line COLO 320HSR (cancer cell line) (labelled in green) versus a normal reference cell line (labelled in red). Signal from the two fluorophores was captured and analyzed by digital image analysis system, which estimated the ratio of intensities of the two fluorophores along chromosome 8 (left). The *myc* locus at chromosome 8q24 showed a high green-to-red ratio (right), consistent with known high-level amplification of *myc* in the cancer cell line.

1.2.2 Array-based Comparative Genome Hybridization

Array-based comparative genome hybridisation (array CGH), a modification to the traditional CGH technique, greatly enhanced the resolution and dynamic range for copy number detection. By substituting chromosome targets in a traditional CGH experiment by a matrix of defined nucleic acid target sequences spotted on glass chips, Solinas-Toldo *et al.* developed the prototype of a CGH array (Solinas-Toldo et al. 1997). Pinkel *et al.* further implemented the platform to screen for copy number changes in human breast cancer, revealing previously undetected aberrations (Pinkel et al. 1998). The array CGH platform has since opened many new opportunities to assess copy number rearrangements associated with human disease and genetic diversity.

A typical CGH array (refer to Fig. 1.3) consists of mapped DNA sequences mechanically spotted or directly synthesized onto microscope glass slides. Various array approaches employed different sources of DNA sequences, which could be broadly classified as genomic inserts (such as BAC, cosmid or fosmid clones (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Snijders et al. 2001), cDNA clones (Pollack et al. 1999), genomic Polymerase Chain Reaction (PCR) products (Dhami et al. 2005) or oligonucleotides (Urban et al. 2006). For each hybridization experiment, uniquely labelled subject and control DNA are co-hybridised with Cot-1 blocking agent (to suppress signal from common repetitive sequences). The test and reference DNA signal intensity was recorded for all probes on the array. Significant deviation from the 1:1 test/reference for a probe (or a series of consecutive probes) would be interpreted as DNA copy number changes (Pinkel and Albertson 2005).

1.2.2.1 BAC Array CGH

The first generation of CGH arrays, and indeed the first genome-wide human CNV screens, was predominantly based on Bacterial Artificial Clones immobilized onto coated glass slides (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Snijders et al. 2001; Fiegler et al. 2003; Iafrate et al. 2004; Redon et al. 2006). BAC arrays started out targeting specific regions of the genome, or tiling the genome at an average resolution of ~1 Mb. With the availability of the overlapping sequencing clone contigs generated for the public domain of the Human Genome Project (Lander et al. 2001; IHGSC 2004), BAC arrays can now be synthesized with over 30,000 features at a tiling path resolution of ~80-150 kb (Ishkanian et al. 2004; Fiegler et al. 2006). In terms of DNA for array spotting, traditionally, substantial effort was required in bacterial culture handling to extract enough DNA from the BAC clones (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Albertson and Pinkel 2003). The problem was later circumvented by applying DNA amplification methods such as rolling circle replication (Buckley et al. 2002), linker adaptor PCR (Klein et al. 1999; Snijders et al. 2001) or degenerate oligonucleotide primers PCR (DOP-PCR) (Fiegler et al. 2003). The whole-genome tiling path (WGTP) array platform was generated using the DOP-PCR strategy (Telenius et al. 1992), with BAC DNA amplified using three different, specifically designed degenerate oligonucleotide primers. Complete amplification of the clone DNA was achieved, and contamination from *E. coli* host vector DNA was minimized. In summary, whole genome tiling path BAC arrays provide a rapid, sensitive and reliable platform for the detection of genomic imbalances (Fiegler et al. 2006).

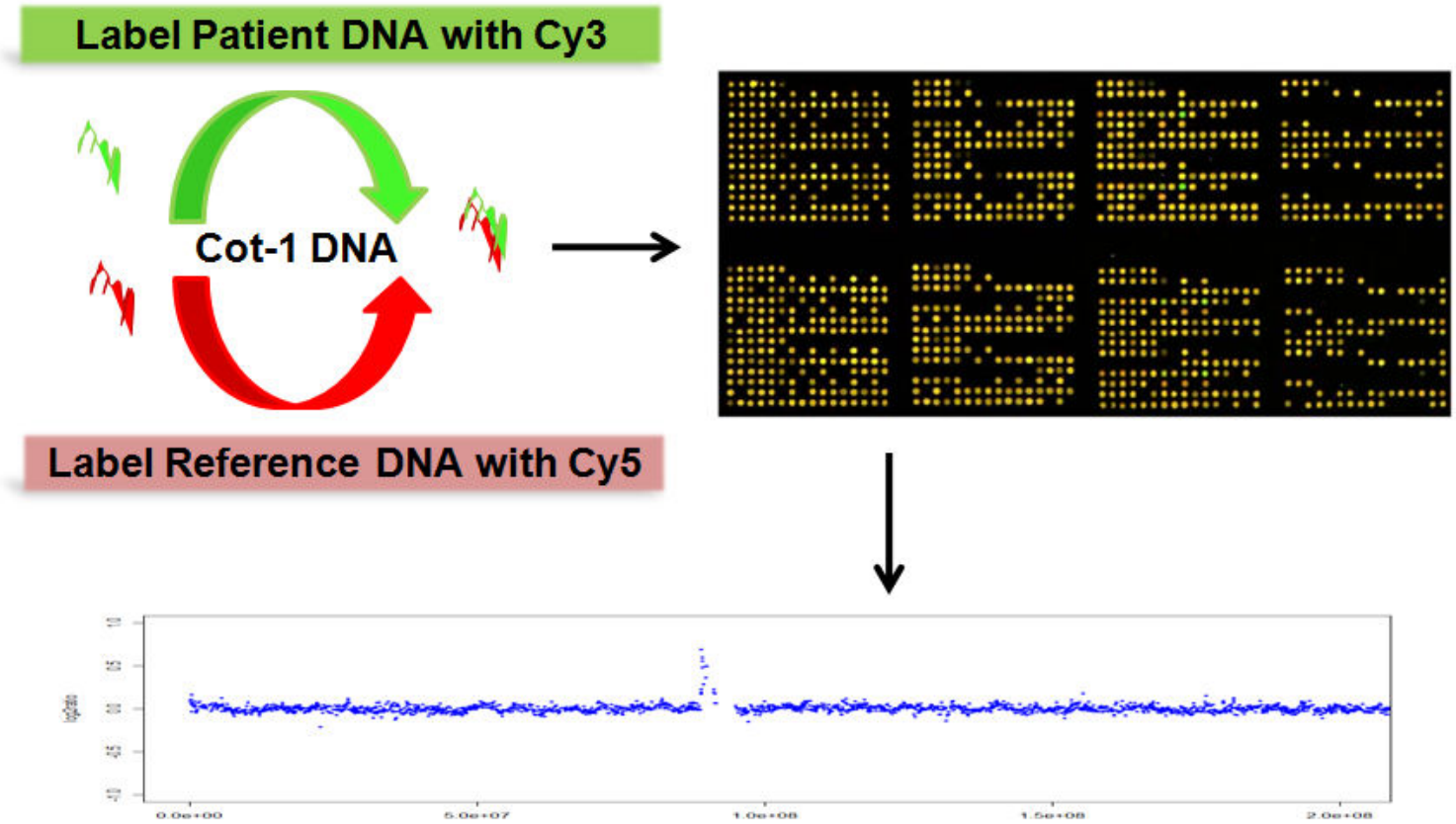


Figure 1.3 Schematics of an array comparative genome hybridization experiment.
The test (green) and reference (red) DNAs are differentially labelled and hybridised to a microarray slide in the presence of Cot-1 DNA. Duplications and deletions are identified as deviations in the signal intensity ratios from the two fluorochromes.

1.2.2.2 Oligonucleotide Array CGH

An alternative to BAC arrays are oligonucleotide-based arrays, which are mostly supplied commercially (e.g. from Roche Nimblegen and Agilent Technologies). These arrays are synthesized in-situ with ~20-80-mer oligonucleotides which form the probes or features for CNV detection. Array designs can be optimized to avoid highly repetitive regions, but with the option of covering low copy repeats such as segmental duplications where CNVs are abundant. Designs can also be customized to target clinically relevant regions for disease studies (Urban et al. 2006; Baldwin et al. 2008).

Early oligo arrays were generally of poorer signal-to-noise ratio compared to BAC arrays, thus resulting in more variable reported signals for CNV detection (Carter 2007). Furthermore, due to the higher costs involved in commercial arrays and reagent purchase, oligo arrays were initially applied mostly for validation (Conrad et al. 2006; Locke et al. 2006; Wong et al. 2007) or breakpoint mapping (Sharp et al. 2005; Gribble et al. 2007). Advancement in technologies, such as the use of digital mask photolithography (Nuwaysir et al. 2002), allow oligo arrays to be constructed at much higher density (Urban et al. 2006), providing considerably better resolution and precision for CNV detection . With improved signal-to-noise ratio, enhanced reproducibility and quality control and decreasing cost per feature, oligo arrays are now recognized as a tool for accurate, high-resolution CNV detection (Ylstra et al. 2006). Examples of the latest whole-genome oligo arrays for human CNV detection include the Agilent 244A (with ~244.000 features) and Nimblegen HG18 WG Tiling CGH 2.1M (with 2.1 million features).

Other related array CGH platforms include representational oligonucleotide microarray analysis (ROMA). The platform consists of oligonucleotide probes from the human genome sequence, hybridized with "representations" of the test and reference genome,

which is prepared by cleaving the genomes with restriction enzymes, followed by differential PCR amplifications (Lucito et al. 2003). This technique was used in an early CNV discovery study (Sebat et al. 2004) as well as a few recent reports of CNV association with psychiatric diseases (Sebat et al. 2007; Walsh et al. 2008).

1.2.2.3 Choice of Reference DNA for Array Hybridization

In array CGH studies, co-hybridization of test and reference genome allows reported signal ratios to be subjective and independent of the probe concentration of each spots, or of variations within array (Carter 2007). The choice of reference, however, is inconsistent within the research community, making direct comparison of array CGH results difficult to interpret. In the early days, pools of reference DNA were used to dilute the effect of CNVs unique to individuals (Iafrate et al. 2004; Sebat et al. 2004). However, this gives imprecise signal intensity ratio especially at genomic regions that harbour complex and multi-allelic copy number changes. The current recommended practice is to perform array hybridization to a single, well-characterized reference genome (Scherer et al. 2007). In our laboratory we routinely use a well-characterized trio-offspring DNA from the HapMap collection, NA10851 (Conrad et al. 2006; McCarroll et al. 2006; Redon et al. 2006), for all array CGH studies.

1.2.3 Genotyping CNVs using Single Nucleotide Polymorphisms

The analysis of single nucleotide polymorphism is a well-established tool for genetic studies for the following reasons. First, over recent decades, a large amount of effort has been put into the discovery, validation and characterization of SNPs in the human genome (Altshuler et al. 2000; Reich et al. 2003; Hinds et al. 2005). Secondly, the HapMap Project provides a catalogue of well-characterized SNPs in four major ethnic populations (IHMC 2005), an important resource for assessing genetic variants in the context of population association studies and evolution. Thirdly, high-throughput array technologies for SNP genotyping have been developed (of particular success are commercial arrays from Affymetrix and Illumina Inc.).

1.2.3.1 Linkage Disequilibrium Based Tag-SNP Approach

One way of utilizing SNPs for CNV investigation is to "tag" common copy number changes, or CNPs, with surrounding SNPs based on linkage disequilibrium. A number of studies have investigated the linkage-disequilibrium properties of CNPs. Early studies indicated that deletion polymorphisms are generally in strong linkage disequilibrium and segregate on ancestral SNP haplotypes (Hinds et al. 2005; McCarroll et al. 2006). Later studies suggest that although some CNVs are in appreciable linkage disequilibrium with nearby markers, accurate genotypes can only be captured for a minority of CNPs tested (Redon et al. 2006). Results were far less robust for CNPs in complex regions such as segmental duplications, which may partially be due to the scarcity of SNP markers surrounding those regions (Locke et al. 2006). Other reasons such as high recombination rate in regions of CNV or high rate of spontaneous recurrence of CNVs may also be possible (Lee and Lupski 2006). In conclusion, current genome-wide technologies are limited to tag CNVs using nearby polymorphic SNP markers (Eichler et al. 2007; McCarroll and Altshuler 2007).

1.2.3.2 CNV Genotyping Using SNP Arrays

Arrays originally designed to genotype SNPs in genome-wide linkage association studies can be used to reliably estimate copy number changes. Affymetrix arrays, for example, were synthesized with 25-mer of matched and mismatched probe pairs to target each SNP under investigation. For array hybridization, a single test DNA is digested with restriction enzyme(s), which was then ligated with adaptors for universal DNA amplification. Hybridization signals of probes at each locus were then compared to those from a single or a group of references hybridized on the same array type, from which CNV calls were generated. Furthermore, CNVs (in particular deletion polymorphisms) can be inferred from regions with extended loss of heterozygosity (LOH), non-Mendelian inconsistency among families and enrichment of Hardy-Weinberg disequilibrium (Conrad et al. 2006; McCarroll et al. 2006).

The first SNP-array based CNV studies started off using the Affymetrix GeneChip Human 10K (Herr et al. 2005), slowly advancing to high-resolution and more robust analysis as SNP array chips and associated bioinformatics tools matured. The Affymetrix GeneChip 500K array, for instance, was used for CNV detection as a complementary platform to BAC array in the HapMap CNV study published in 2006 (Komura et al. 2006; Redon et al. 2006). An apparent advantage of such an array platform is its versatility- SNPs, CNVs and other clinically useful data such as uniparental disomy (copy number neutral LOH) (Friedman et al. 2006; Peiffer et al. 2006) can all be mined from data generated from the same chip. A major drawback of the early SNP arrays (e.g. Affymetrix 250K/500K or Illumina HumanHap300) was uneven probe spacing, with particularly sparse probes at regions near segmental duplication, and at repeat-rich centromeres and telomeres, perhaps due to problems they create against robust SNP genotyping (Carter 2007). To enhance the power of CNV detection, the most recent SNP genotyping platforms have

included non-polymorphic probes specifically selected for their genomic positions and for linear response to copy number changes. Examples of the latest platforms are Affymetrix Genome-wide Human SNP array 5.0 and 6.0, the latter with ~906,600 SNPs and ~946,000 CNV probes; and Illumina Human1M-Duo Bead Chip, with ~1.2 million markers and 1.5 kb median marker spacing, with probes designed to target known CNV regions and gaps between HapMap SNPs. These platforms have generated a new paradigm for genome-wide disease association studies, integrating both SNP and CNV assessments in the work-flow of such investigations (Korn et al. 2008).

1.2.4 Validation and Detection of CNV at Targeted Loci

Quantitative or semi-quantitative measurements of copy number variations at targeted loci are required for a number of reasons: i) as independent platforms to validate array-based CNV discoveries; ii) to convert analog signal-intensity ratios from array-based CNV detection to discrete CNV genotypes; iii) to precisely map breakpoints of CNV regions; and/or iv) to develop low-cost, reliable assays for large-scale genotyping in a large number of samples, for example in disease association studies or in clinical diagnostic settings. Traditional methods such as FISH, Southern Blot, PFGE and long range PCR serve some of these purposes, but they remain laborious, technically demanding, and low-throughput therefore unsuitable for larger scale studies. A number of complimentary techniques have evolved along with the progress of CNV discoveries.

1.2.4.1 Quantitative Fluorescent Real-time PCR (qPCR)

QPCR was first developed for quantification of RNA (Higuchi et al. 1992; Heid et al. 1996). In qPCR assays, input DNA quantity was monitored with dyes (e.g. SYBR-green) or dual-labelled probes (e.g. Taqman^{TM} probes), and fluorescence was recorded in real-time during the PCR amplification reaction. For instance, Taqman^{TM} assays draw on the 5'-3' exonuclease activity of Taq polymerase, which cleaves the dual-labeled probes attached to DNA, releasing a reporter dye (e.g. FAM) which emits a fluorescence signal during PCR amplification. The technique was first used in the context of gene copy number changes assessing the hemizygosity associated with tumor, and of the gene dosage at hereditary neuropathy with liability to pressure palsies and Charcot Marie Tooth disease (Laurendeau et al. 1999; Wilke et al. 2000). In the absence of post-PCR manipulation, qPCR has the advantage of being fast, semi-automated with accurate quantification. Furthermore, a minimal amount of input DNA is needed and therefore the method is

scalable to high-throughput screening (Hoebeeck et al. 2007), although multiplex assays are still difficult to optimize (Conrad and Hurles 2007). It is now widely used as CNV validation strategy and for clinical diagnosis (Weksberg et al. 2005; Qiao et al. 2007; Malakho et al. 2008).

1.2.4.2 Multiplex Quantitative Fluorescent Real-time PCR

More quantitative PCR methods have been developed to increase throughput and accuracy of traditional qPCR. For example multiplex ligation-dependent probe amplification (MLPA) (Schouten et al. 2002), Multiplex Amplicon Quantification (MAQ) and multiplex amplifiable probe hybridization (MAPH) (Armour et al. 2000; White et al. 2002) are routinely used for copy number quantification in research as well as diagnostic settings. These multiplex approaches allow PCR amplification to be performed using universal primer sets, with different probes specifically bound for different target amplicons generating specific fluorescent peaks for semi-quantitative detection of gene dosage. All these methods facilitate parallel screening of a large number of samples across a large number of putative CNV loci at affordable costs and high reliability.

1.2.4.3 Other Methods

Other methods for CNV dosage quantification include novel techniques based on competitive PCR. In competitive PCR, a control target of known concentration is co-amplified with an unknown test competitor sequence, and the concentration of the test sample is inferred by comparison with the control. The difference between the competitive sequences could be as small as a single nucleotide mismatch. Interspecies PCR, for example, exploits the large number of SNPs or dispersed repeat element differences between human and chimpanzee as competitive sequences for CNV dosage detection (Armour et al. 2007; Williams et al. 2008). These techniques have demonstrated robust

and inexpensive measurement of copy numbers at the 22q11 loci (Williams et al. 2008) and beta-defensin locus (Hollox et al. 2008) .

1.2.5 Genome Sequencing and CNV Detection

The availability of the human genome sequence (Lander et al. 2001; Venter et al. 2001; IHGSC 2004) was a pre-requisite for the success of CNV discoveries. Genomic insert clones (e.g. BACs) or oligonucleotides on whole-genome arrays are tiling representations of the reference genome sequence (Shendure and Ji 2008). SNP genotyping arrays, another major CNV detection platform, were developed with the International HapMap Project (IHMC 2005) playing an instrumental role. Nevertheless, both the human genome assembly and the International HapMap Project are not perfect, and the reference genome sequence tends to conceal or collapse regions of the genome most likely to harbor copy number variants, for example, regions at or near segmental duplications. Technologies for deep sequencing and comparison of more individual genomes are therefore likely to reveal more about copy number variations.

1.2.5.1 Clone End Mapping and Sequencing

In clone end mapping, a genomic sequence is fragmented and subcloned into circular clone (e.g. fosmid) vectors to create a genomic library. The ends of these genomic insert are sequenced from universal primers in the vector, and these are mapped to a reference genome. End sequence pairs that are discordant in terms of length (suggesting insertions/deletions) or orientation (suggesting inversion) are recorded (Eichler et al. 2007). In 2005, Tuzun *et al.* reported 297 structural variant events between two individual genomes based on fosmid end sequence mapping of a DNA sample NA15510 against the reference genome sequence (Tuzun et al. 2005). Two years later, Korbel *et al.* reported a study based on higher-resolution method known as paired-end mapping (PEM) on two individual genomes (NA15510 and NA18505), detected over 1000 insertions/deletions (Korbel et al. 2007). More recently, clone-end sequence-pair (ESP)

maps of eight human genomes were released, revealing even more novel CNVs and refining previously known structural variations (Kidd et al. 2008).

1.2.5.2 Novel Sequencing Technology and Genome Comparison

Computational comparisons between published genome assemblies have provided us more insights on copy number variations (Khaja et al. 2006; Levy et al. 2007). With the advancement in technologies, a number of novel sequencing methods were developed on top of the traditional Sanger sequencing technique (Sanger 1981). In particular, "second generation" sequencing methods based on cyclic-array sequencing have already been realized as commercial products, for instance in 454 sequencing (Roche Applied Science), Solexa technology (Illumina Inc) and the SOLiD platform (Applied Biosystems) (Shendure and Ji 2008). These massively parallel DNA sequencing technologies will allow complete genome sequencing and re-sequencing of more individuals at an affordable cost (Service 2006; Eichler et al. 2007), unveiling the true extent of CNVs in the human genome.

Figure 1.4 summarizes the sensitivity and the throughput of various CNV detection techniques we have discussed so far.

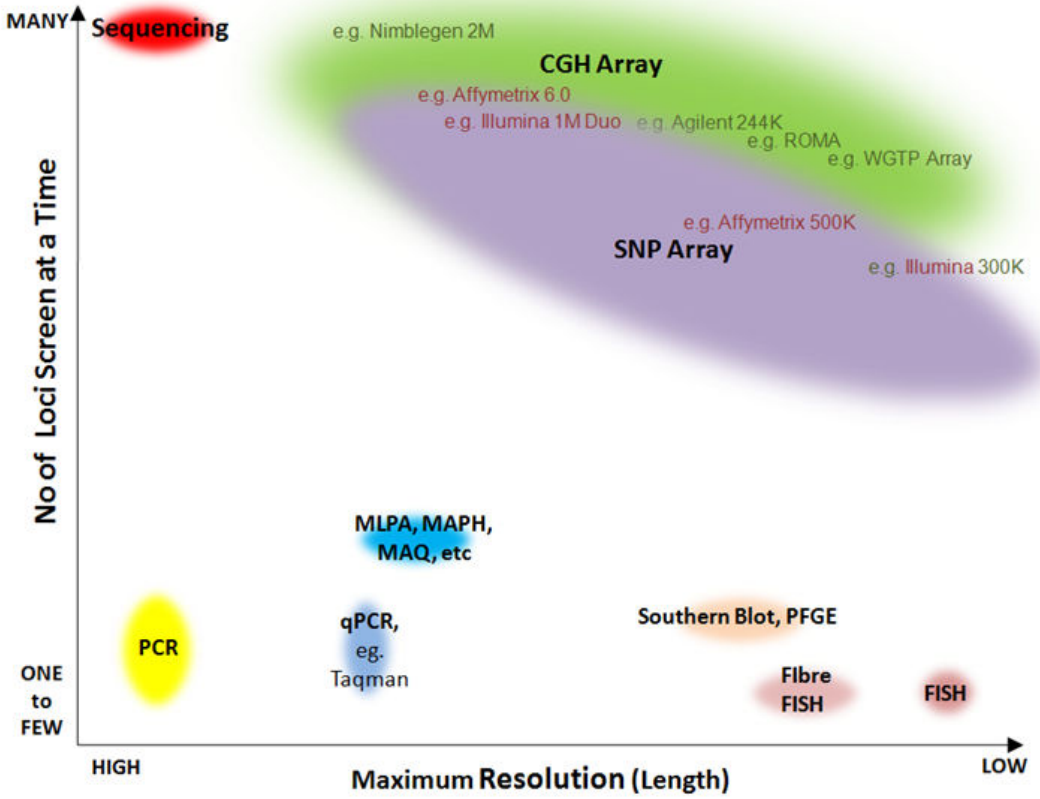


Figure 1.4 Sensitivity and throughput of various CNV detection techniques. x-axis: maximum resolution of the technique (indicate sensitivity); y-axis: number of loci that can be screened at a time (indicate throughput and scalability).
PCR: polymerase chain reaction; qPCR: quantitative PCR; MLPA: multiplex ligation-dependent probe amplification; MAPH: multiplex amplifiable probe hybridization; MAQ: Multiplex Amplicon Quantification; PFGE: pulse field gel electrophoresis; FISH: fluorescence in-situ hybridization; SNP Array: single nucleotide polymorphism genotyping array; CGH Array: Comparative Genome Hybridization array

1.3 Mechanisms of Copy Number Variation Generation

Understanding the underlying mutational processes for CNV will provide information on how this type of genetic variant emerges, and should yield important insights into the genomic distribution, evolution and frequency of CNVs in the population. There are two major mechanisms for CNV generation, namely non-allelic homologous recombination (NAHR) and non-homologous end joining (NHEJ). Both are cellular mechanisms intended for maintaining the integrity of DNA by repairing DNA double stranded breaks (DSBs).

1.3.1 Non-allelic Homologous Recombination

The evidence of NAHR being a major mechanism of CNV generation came from observations that copy number changes frequently fall in close proximity to repeat sequences, for instance segmental duplications (Sharp et al. 2005) or Alu repeats (Lupski 2006; de Smith et al. 2008). Early examples came from genomic disorders, a class of diseases resulting from alteration of dosage-sensitive genes, usually by large-scale microdeletion or microduplication of the genome (Stankiewicz et al. 2003; Lee and Lupski 2006). These genomic disorders, such as Velocardiofacial Syndrome, Williams Buren Syndrome and Charcot Marie Tooth Disease, frequently have breakpoints clustering at highly homologous segmental duplicons or low copy repeats (LCRs). It was suggested that the LCRs act as substrates for NAHR, mediating the rearrangements (Stankiewicz and Lupski 2002). It was also hypothesized that recurrent CNVs were most frequently results of NAHR (Lee and Lupski 2006).

Copy number changes mediated by NAHR may occur at homologous sequences either on the same or the non-homologous chromosome (inter-chromosomal or

intrachromosomal NAHR). Apart from LCRs, more divergent repeats such as SINEs (eg. Alu), LINEs and human endogenous retroviruses (HERVs) can all act as substrates for NAHR (Roy et al. 2000; Fredman et al. 2004; Hurles 2005), although such events are less likely to be recurrent (Cooper et al. 2007).

During meiotic NAHR, misalignments or unequal cross-over of the homologous sequences generate germline rearrangements such as duplications, deletions and inversions (Figure 1.5), and be detected as structural variations using techniques as discussed in section 1.2. Somatic rearrangement may also be generated in a similar manner (Cook and Scherer 2008), although this class of CNV is less well-studied (Piotrowski et al. 2008).

Depending on the distribution and participation of the homologous sequences, copy number changes generated by NAHR could be simple deletion or duplication (Figure 1.5a & b); or more complex rearrangement, such as tandemly duplicated arrays (Hurles 2005), as in the case of the opsin locus for red-green color vision (Neitz and Neitz 1995) (Figure 1.5d); or other complicated structural variations involving multiple homologous duplicons.

1.3.2 Non-homologous End Joining

NHEJ is an alternative mechanism for repairing DSBs in cells. When random DSB occur at regions with no extensive homologous sequences to act as the repair template for NAHR, the broken ends of the DSB may be rejoined by nucleases removing the nucleotides, and the Pol X family of DNA polymerases filling in the missing nucleotides. NHEJ is an error-prone repair mechanism, with frequent gains and losses of nucleotides

at the junctions (Lieber et al. 2003). Much less is known about NHEJ compared to NAHR (Lee et al. 2007). The mechanism is suggested to be mediated by microhomologies (< 25 bp homology) (Lieber et al. 2003). It is also known that NHEJ is more prevalent in unstable (or fragile) regions of the genome, for example in the subtelomeric regions (Nguyen et al. 2006; Kim et al. 2008b). The mechanism has been implicated in a number of genomic disorders (Inoue et al. 2002; Shaw and Lupski 2005)

1.3.3 Other Mechanisms

Recently, a few novel mechanisms of CNV generation have been suggested, which may generate interest for further characterization. One example is a mechanism involving DNA stalling and template switching during mitotic replication, as observed in Pelizaeus-Merzbacher disease, an X-linked genomic disorder (Lee et al. 2007). Another mechanism involves both homologous and non-homologous recombination, as seen in the methyl-CpG-binding protein 2 (*MECP2*) duplication on chromosome Xq28 (Bauters et al. 2008).

1.3.4 Insights from Breakpoint Mapping

A recent paper systematically investigating the breakpoints of CNVs (Kim et al. 2008b) provided new insights into CNV generation mechanisms. In the study, 534 (fosmid-paired end sequenced) CNV breakpoints were subjected to bioinformatics analysis to look for co-localization with various classes of repeats. In particular, they reported that 28% of CNVs were colocalized with segmental duplication, a result consistent with earlier studies (Sharp et al. 2005; Redon et al. 2006; Cooper et al. 2007) albeit at a lower rate. In terms of signatures with other NAHR substrates, CNV breakpoints were not shown to be significantly associated with Alu elements, contrary to a previous study

showing significant association (Cooper et al. 2007). On the other hand, CNV breakpoints were shown to co-localize with LINE elements (>20% of breakpoints) and microsatellite repeats (3% of breakpoints) (Kim et al. 2008b).

Moreover, Kim *et al.* showed evidence for NHEJ as a major mechanism of CNV generation. 40% of the breakpoints showed microhomologies, implicating NHEJ involvement. Another 14% demonstrates micro-insertions (base pairs of gains/losses) at breakpoint junctions, which is also suggestive of NHEJ mediation (Kim et al. 2008b).

In summary, the non-random distribution of CNV, its co-localization with segmental duplications and its clustering at telomeric and centromeric regions, probably reflect the underlying mechanism of CNV generation (Stankiewicz and Lupski 2002; Hurles 2005; Lupski 2007). Whether such mutational bias plays a particular role in human evolution and diseases remain to be investigated.

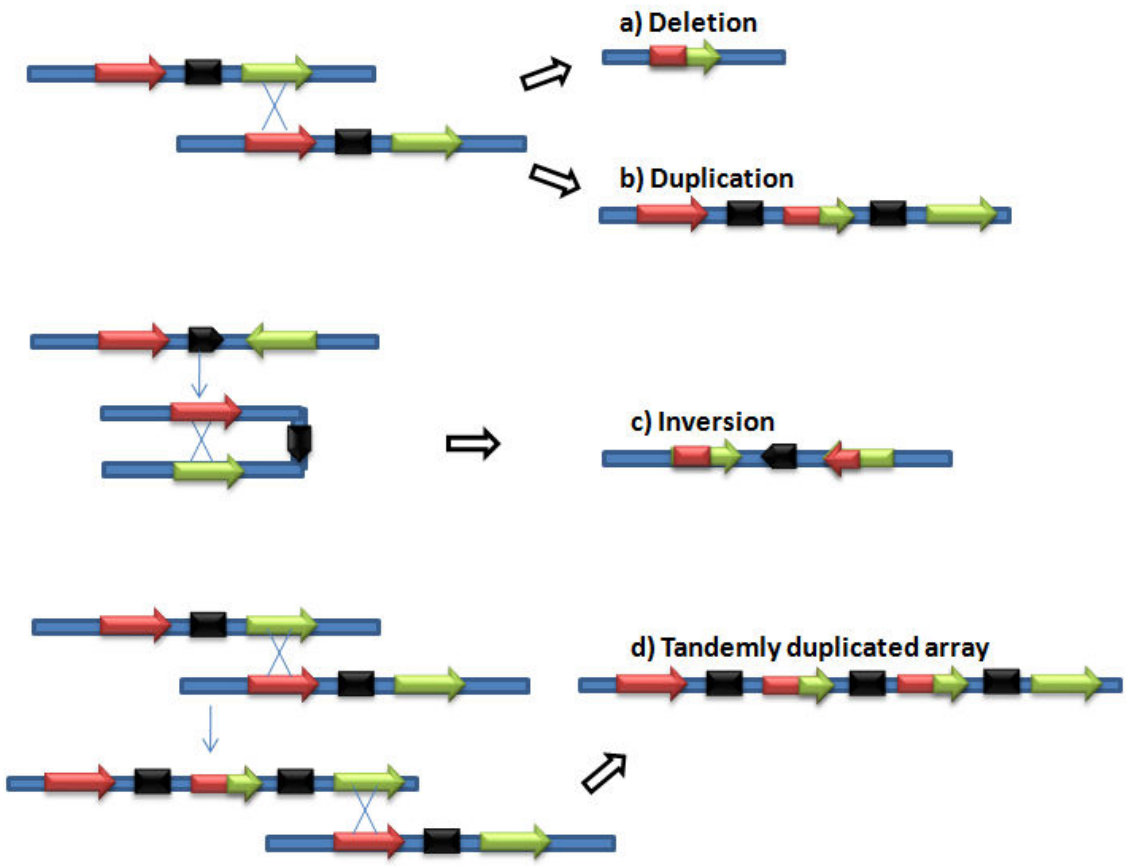


Figure 1.5 Non-allelic homologous recombination by low copy repeat. Examples of mechanisms generating different structural variants: **a)** deletion; **b)** duplication; **c)** inversion; **d)** tandemly duplicated array

1.4 Biological Impact of Copy Number Variation

1.4.1 Phenotypic Effect of CNVs

Genetic variants of all sizes and types could confer phenotypic effects on individuals via two broad mechanisms: a) perturbing gene expression, by acting on transcription, splicing, or translation and stability; or b) modify the function of a protein, for example by altering the protein structure. Based on these two mechanisms, below is a non-exhaustive list of how copy number variations could confer functional consequences (Lupski and Stankiewicz 2005; Feuk et al. 2006b; Eichler et al. 2007; Shelling and Ferguson 2007):

- Deletion and duplication may act on dosage-sensitive genes, with gene copy number changes resulting in alteration of gene expression;
- Rearrangement may alter the regulatory elements for a nearby gene, with possible long-range effect (Kleinjan and van Heyningen 2005) which changes gene expression at a distance;
- CNVs may disrupt a gene, e.g. by interrupting protein coding sequences, causing functional loss or modification;
- Such rearrangements may also generate novel fusion products, for example by fusing different protein domains, by inserting protein coding sequences to the proximity of regulatory elements, or by deleting exons generating novel splice variants;
- The change of expression of a gene involved in CNV may upset the stoichiometry, or the balance, of a macromolecular complex or network, translating the CNV effect to other genes and proteins (Korbel et al. 2008);
- Deletions may unmask mutations or functional SNPs in the remaining allele

- CNVs may affect microRNAs in the genome, which in turn may lead to alterations in gene expression

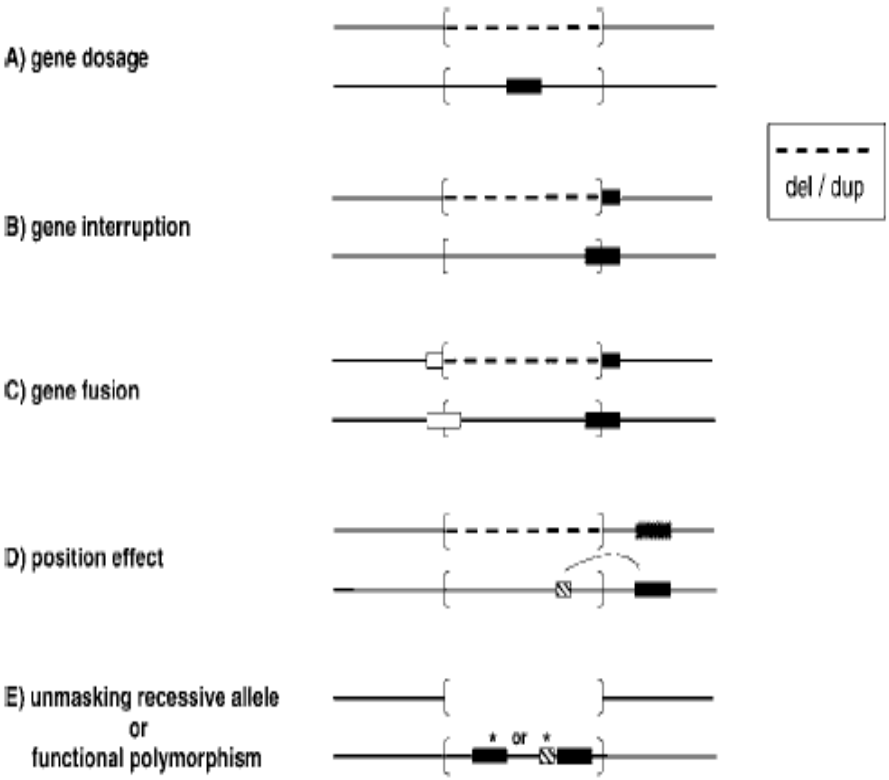


Figure 1.6 phenotypic effects of CNVs (diagram adapted from Lupski and Stankiewicz, 2005). Examples of CNVs affecting phenotypes through: **a)** gene dosage; **b)** gene interruption; **c)** gene fusion; **d)** position effect; **e)** unmasking recessive allele

1.4.1.1 CNV and Expression

The effect of CNV on gene dosage is most prominent in genomic disorders. In the example of the Charcot Marie Tooth locus (CTM-1A) on chromosome 17, for example, haploinsufficiency of the dosage-sensitive gene peripheral myelin protein 22 (*PMP22*) causes Hereditary Neuropathy with Liability to Pressure Palsies (HNPP), whereas reciprocal duplication results in trisomic expression of *PMP22*, leading to another neuropathy, the Charcot Marie Tooth Disease (Chance and Fischbeck 1994).

Compared to simple deletions or duplications, multi-allelic CNVs may give subtler yet more variable effects on the range of gene expression by altering dosage levels for individuals in the population. In the case of the human salivary amylase gene, *AMYI*, the diploid copy number of this starch hydrolysis gene varies from 2 to >15 (Iafrate et al. 2004; Redon et al. 2006), with populations having higher starch diet generally found to have higher *AMY1* copy number and vice versa (Perry et al. 2007). It was demonstrated that this CNV locus affects gene expression at both the transcriptional and translational levels, with copy numbers shown to correlate with mRNA and *AMY1* protein level in human saliva samples (Perry et al. 2007).

As a genome-wide survey of the impact of CNV on expression patterns, Stranger *et al.* examined mRNA levels in lymphoblastoid cell lines from 210 unrelated HapMap individuals from 4 different ethnic groups, in combination with CNV data generated from the same HapMap samples (Redon et al. 2006; Stranger et al. 2007). 8.75% of variation in expression levels of 972 genes could be attributed to copy number changes (Stranger et al. 2007). In some cases, the effects of CNVs on gene expression were exhibited across all 4 ethnic groups, while in other cases they remain population-specific. Distant regulation or positional effect were also reported, with > 50% of the expression probes

associated with CNVs being away from the CGH clone encompassing the CNV, some of them as far as 2 Mb apart (Stranger et al. 2007). One example detected was the changes in mRNA expression of UDP-glucuronosyltransferase 2B17 (*UGT2B17*), which vary with copy number (McCarroll et al. 2006; Stranger et al. 2007). Variability of the gene was previously shown to associate with testosterone level, male insulin sensitivity, fat mass and prostate cancer (Xue et al. 2008).

A simple, direct model for the effect of CNV on gene expression suggests proportional relationship between the two, with expression increasing as gene dosage increases. This probably explains the effects of many CNV loci, for instance the nearly linear relationship of protein expression with salivary amylase gene copy number (Perry et al. 2007). Negative relationship has also been suggested, in which higher copy number decreases expression level. For instance a small duplication downstream of the proteolipid protein gene *PLP1* was proposed to cause gene silencing by positional effect, thus lowering gene expression (Lee et al. 2006). However, considering the intricate and convoluted human biological networks and pathways, it is possible that some CNVs confer more complex effects on gene expression. Recently, Mileyko et al. examined this by modelling the effect of four common gene regulatory motifs: “positive feedback”, “bistable feedback”, “toggle switch” and “repressilator” (Mileyko et al. 2008). They proposed that a small change in gene copy number within these motifs could completely switch a biological network from one alternative steady state to another, and this change in equilibrium could result in large and unanticipated changes in gene expression.

Nonetheless, some CNVs are phenotypic benign variants (as marked by the numerous CNVs in apparently normal populations) (Cooper et al. 2007). This suggests that changes in copy number may not necessarily translate into gene expression,

phenotypes or human traits. In many cases, CNVs are neither disease-causing variants nor variants with adverse phenotype. These apparently "neutral" CNVs may be explained by i) the genomic location of CNV (e.g. being located in gene-poor regions); ii) the insensitivity of the CNV genes to dosage; or iii) the effect of dosage compensation (Birchler et al. 2005), a mechanism which would balance out gene expression changes resulting from CNVs. In conclusion, the severity of phenotypes caused by a given CNV via gene expression is a combination of the dosage sensitivities of genes affected by the CNV, plus their interaction with other genetic and environmental variants.

1.4.1.2 CNV and Gene Disruption

Since not all genes are dosage sensitive, an alternative mechanism that CNV might confer phenotypes and diseases is by disrupting genes and modifying protein function. In a recent genome-wide CNV screen in schizophrenia patients, Walsh *et al*. suggests that CNV breakpoints in the case cohort disrupt many more genes that are involved in neurodevelopmental pathways than in the control cohort (Walsh et al. 2008). One example was a ~400 kb deletion disrupting the *ERBB4* gene, which encodes a type I transmembrane tyrosine kinase receptor. Using 3' rapid amplification of cDNA ends (RACE), the authors cloned the mutant transcript, which was found to lack the intracellular kinase domain of the receptor. This mutant allele could potentially have a dominant negative effect, with evidence from a previous mouse model. CNVs that disrupt genes can potentially have major phenotypic effects, forming a class of variants with high penetrating effect (Walsh et al. 2008). With more CNV breakpoints being mapped by high-resolution techniques and the resulting allelic sequences and structures characterized, we will gain a more complete understanding of the effects of copy number variants.

1.4.2 Phenotypic Variations and Evolution

1.4.2.1 CNV and Human Traits

Over the years, through a combination of genome-wide association or single locus approaches, single nucleotide polymorphisms have been shown to influence a diverse range of normal human traits, from the ability in speech and language (e.g. *FOXP2*) (Lai et al. 2001), to variations in height (Gudbjartsson et al. 2008; Lettre et al. 2008), even to sprinting ability in athletic performance (e.g. *ACTN3*) (North et al. 1999; Yang et al. 2003). Human copy number variations were demonstrated to encompass more base pairs of changes between individuals than all SNPs combined together (Redon et al. 2006). With growing numbers of CNV discovery studies, a major question arising is what impact does this ubiquitous class of genetic variant have on normal human phenotypic trait (Nguyen et al. 2006). A clear example comes from the aforementioned *AMY1* locus, in which copy number was shown to correlate with the amount of starch-digesting amylase protein in human saliva (Perry et al. 2007). The geographical variation of copy number at this locus was proposed to be shaped by the evolution of diet in high-starch eating populations (such as Japanese) compared to low-starch eating populations (such as Biaka) (Perry et al. 2007).

Another argument comes from the X-linked opsin gene locus dictating red-green color vision (Deeb 2005). The locus comprises red (L-opsin) and green (M-opsin) photopigment genes and a “locus control region” regulating gene expression. Variable copy numbers were detected in the L-opsin gene (0-4 haploid copy), M-opsin gene (0-7 haploid copy) and the locus control region unit (0 or 1 haploid copy). It is a combination of the copy number of these three units, sequence variation and unequal cross-over between homologous L-opsin and M-opsin that determines the phenotype of color vision, even to the extreme of color blindness (Cooper et al. 2007).

In addition, copy number at the cytochrome P450 gene *CYP22A6* has been implicated in altered nicotine metabolism and smoking behaviour (Rao et al. 2000), and a deletion polymorphism at UDP-glucuronosyltransferase 2B17 (UGT2B17) was demonstrated to correlate with level of testosterone excretion (Jakobsson et al. 2006). The *UGT2B17* CNV was also shown to have major geographic variation, with evidence of positive selection in the East Asia population (Xue et al. 2008).

1.4.2.2 Positive and Negative Selections on CNV

CNVs at specific loci (e.g. AMY1) have been suggested as being involved in human adaptive evolution (Perry et al. 2007). On a genome-wide level, several studies have proposed that natural selection processes may have influenced and shaped CNV genomic distribution and gene content (Nguyen et al. 2006; Cooper et al. 2007; Nguyen et al. 2008).

Duplications have long been suggested as substrates on which selection processes can act. A duplicated gene and its regulatory elements could subsequently be modified for new functions, facilitating a species' diversification and evolution (Lynch and Conery 2000; Dermitzakis and Clark 2001; Korbel et al. 2008). Some of these innovations may have been favoured in positive selection processes. In contrast, negative selection could also have shaped the characteristics of CNVs, and can easily be envisaged in deletion scenarios where deleterious loss-of-function alleles could be generated. In support of this, we see an enrichment of recent common CNVs (those that do not overlap with SDs) in gene-poor regions (Cooper et al. 2007). Array CGH studies further suggested that deletions in the genome are more biased away from the morbid OMIM (Online Mendelian Inheritance in Man) genes than duplications are (Redon et al. 2006).

One could hypothesize that the current distribution of CNVs in the human genome has been driven by past positive selection and adaptive evolution. Alternatively, from a more neutralist's point of view, CNVs could be under reduced purifying selection (Nguyen et al. 2008). This model suggests that variants are enriched in nonessential genes, with fixation of slightly deleterious substitutions, while the "neutral" CNVs were under genetic drift.

Examining the functional categories of CNV gene content may provide a glimpse into the underlying selection processes. A number of studies have suggested that CNVs significantly co-localize with genes that have environmentally responsive functions, such as sensory perception and immunity (Nguyen et al. 2006; Redon et al. 2006). Such enrichment could reflect these functional classes being less deleterious, therefore remaining in the genome by reduced purifying selection (Nguyen et al. 2008). On the other hand, proteins within these biological categories were well-documented in mammalian adaptive evolutions (Cooper et al. 2007), and may have acted as substrates for natural selection. To this end, CNVs affecting environmental response genes have recently been demonstrated having involvement in a number of infectious and autoimmune diseases, such as HIV-1 susceptibility, glomerulonephritis and Crohn's Disease (Gonzalez et al. 2005; Aitman et al. 2006; Fellermann et al. 2006).

In summary, natural selection may differentially act on a spectrum of CNVs with variable phenotypic consequences, from beneficial, neutral, benign to deleterious (e.g. disease-causing CNVs). Further characterization of CNVs in a range of phenotypic and disease contexts will expand on our rudimentary knowledge about this source of genetic variants.

1.4.3 CNV and Disease

One of the major challenges in biology is to unravel mechanisms underlying human diseases. Studies in copy number variations have indicated its role in a number of sporadic diseases, Mendelian disorders as well as some multifactorial and complex disease phenotypes. The earliest evidence of such came from a group of diseases known as 'genomic disorders'.

1.4.3.1 Genome Disorder

Genomic Disorders are defined as diseases caused by genomic rearrangements affecting dosage sensitive genes (Stankiewicz and Lupski 2002; Lupski 2007). They are classic models of DNA sequences being altered as copy number variations resulting in adverse human phenotypes (Lupski and Stankiewicz 2005). Well-known examples are Williams Beuren Syndrome (WBS) [del(7)(q11.23q11.23)], Velocardiofacial/ DiGeorge Syndrome (VCFS/DGS) [del(22)(q11.2q11.2)], Prader–Willi (PWS) [pat del(15)(q11.2q13)] or Angelman syndrome (AS) [mat del(15)(q11.2q13)], and Charcot Marie Tooth Disease (CMT) with reciprocal Hereditary Neuropathy with Liability to Pressure Palsies (HNPP) (Lupski 2006). These classical examples are typically large microdeletions and duplications (of several Mb) (Inoue and Lupski 2003), as they were discovered via traditional cytogenetic technique (e.g. karyotyping). Recent advance in CNV detection techniques have revealed a growing number of novel genomic disorders, such as the 15q13.3 microdeletion syndrome (~1.5 Mb) associated with mental retardation and seizures (Sharp et al. 2008), and the 17q21.31 microdeletion syndrome (~500 kb) associated with mental retardation and developmental delay (Koolen et al. 2006; Sharp et al. 2006; Shaw-Smith et al. 2006).

Genomic disorders have provided us with important insights regarding CNVs and human diseases: First, these syndromes have revealed the plasticity of the human genome: alterations of megabases of DNA sequences affecting genes with important biological functions can apparently be tolerated in human. In some cases, such as Charcot Marie Tooth Disease, phenotypes of the disorder could be narrowed down to a single gene (*PMP22*) within the large rearrangement (Lupski et al. 1991; Patel et al. 1992). Furthermore, several genomic disorders are the result of reciprocal deletions/duplications of the same loci, demonstrating that varying dosage of the same gene(s) could result in diverse phenotypes (Lupski and Stankiewicz 2005).

Secondly, genomic disorders provide us insights into the underlying mutational mechanism of copy number changes. Some cases of genomic disorders may cluster to breakpoints of LCRs or segmental duplications (Lupski and Stankiewicz 2005), suggesting NAHR as one of the disease mechanisms. Realizing genomic disorders may reflect the underlying genomic architecture, several research groups have interrogated specifically the SD-rich regions of the genome, facilitating discovery of further genomic disorders and diseases loci (Sharp et al. 2006). Other higher-order genomic architecture, such as nearby inversion polymorphism, may also predispose the disease locus to rearrangement. This has been demonstrated in 17q12.13 microdeletion syndrome, where a parental H2 inversion is necessary for disease transmission (Koolen et al. 2006; Shaw-Smith et al. 2006; Koolen et al. 2008), and has been suggested in Sotos Syndrome (Kurotaki et al. 2005), another genomic disorder.

Furthermore, these disorders demonstrated the interaction of CNVs with other genetic factors in diseases. PWS and AS, for example, are derived from the paternal and maternal microdeletions of the same genomic locus, resulting in distinct disease

phenotypes (Lupski 2006). This reveals an interconnection between CNVs and epigenetic factors (imprinting mechanism).

Finally, the majority of genomic disorders involve nervous system disorders and neuropathies, providing us invaluable insights into diseases of the brain. For example, these mental retardation syndromes were found to be comorbid with various psychiatric symptoms (eg. PW/AS with autism (Veltman et al. 2005), VCFS/DGS with schizophrenia (Murphy and Owen 2001)), suggesting genes altered in these loci may play multiple roles resulting in overlapping behavioral and cognitive phenotypes, or vice versa the diverse phenotypes could be the consequences of the perturbation of a common biological pathway via the alteration of common gene(s).

1.4.3.2 Rare CNVs in Mendelian Disease Traits

In contrast to genomic disorders, which in many cases are sporadic diseases resulting from *de novo* CNVs (Lupski 2007), the involvement of CNVs in diseases can also be illustrated with rare inherited copy number changes. These inherited CNVs act in the same way as Mendelian mutations do in diseases (Estivill and Armengol 2007). Below are examples originated from two neurodegenerative disorders.

Triplication or duplication of the alpha-synuclein (*SNCA*) locus was shown to cause Parkison's diseases (Singleton et al. 2003; Chartier-Harlin et al. 2004). Singleton *et al.* first demonstrated triplication of the SNCA to hereditary early-onset parkinsonism with dementia. Compared to the triplication cases, the duplication kindred identified by Chartier-Harlin *et al.* seemed to show milder disease progression closely resembling idiopathic Parkinson's, with late onset and no cognitive decline or dementia, suggestive of a dosage effect in disease progression. Alpha-synuclein protein is a major component of Lewy bodies, the "pathological hallmark of Parkinson's Disease" (Singleton et al.

2003). Examination of brain tissue showed increased copy number correlates with the level of soluble alpha-synucelin, and an even more prominent increase in the form of aggregated, insoluble deposition (Miller et al. 2004) .

In the second example, duplication of the amyloid precursor protein (*APP)* on chromosome 21 was demonstrated to cause autosomal dominant form of early-onset Alzheimer disease (ADEOAD) with cerebral amyloid angiopathy (CAA) (Rovelet-Lecrux et al. 2006). From their study, Rovelet-Lecrux *et al.* reported that 5 out of 65 families of ADEOAD with CAA showed duplication at the *APP* locus, estimating the frequency of this duplication to be 8% (5 of 65) in the disease. This is also consistent with previous reports suggesting that *APP* duplication coincide with patients with Alzheimer disease (Delabar et al. 1987; Sleegers et al. 2006). Alterations in the *APP* gene probably lead to accumulation of amyloid precursor protein which results in neurodegeneration (Rovelet-Lecrux et al. 2006), highlighting *APP* as a dosage sensitive gene. The copy number change at the *APP* locus was also put into context of SNP mutations in the same gene, where the CNV is estimated to correspond to half of SNP missense mutations leading to disease (Rovelet-Lecrux et al. 2006).

1.4.3.3 Common CNVs in Multifactorial or Complex Disease

Common diseases involving multiple genetic and environmental factors are in principle more susceptible to copy number variations, which could alter gene dosage without disrupting protein functions and therefore result in high phenotypic plasticity (McCarroll and Altshuler 2007). Furthermore, many multifactorial diseases, such as a number of complex psychiatric disorders, were shown to be heritable with a substantial genetic contribution, whilst the underlying genetic factors remain largely elusive (Burmeister et al.

2008). The involvement of CNVs in complex diseases is therefore an area under intense investigation.

There is now considerable evidence of complex diseases under the influence of copy number variants. These include a number of infectious and autoimmune diseases, cancer, as well as psychiatric diseases (discuss in section 1.6). Below I discuss three examples in detail, namely the human chemokine gene *CCL3L1* in HIV-1 susceptibility and progression (Gonzalez et al. 2005); the Fc-receptor gene *FGCR3B* in lupus and related autoimmune diseases (Aitman et al. 2006; Fanciulli et al. 2007), and the β-defensin gene cluster (including *DEFBA3 & DEFB2*) in Psorasis and Colonic Crohn's Disease (Fellermann et al. 2006; Hollox et al. 2008).

In 2005, Gonzalez *et al.* reported probably the first evidence of copy number variants stably transmitted to contribute to complex diseases (Gonzalex, 2005). The authors studied the distribution of chemokine gene-containing segmental duplications on chromosome 17, which include genes *CCL3L1*, *CCL4L1* and *TBC1D3*, in 1064 humans from 57 populations and 83 chimpanzees. The copy number of the SD has previously been demonstrated to regulate the production of human chemokine *CCL3L1* (Townson et al. 2002). Gonzalez *et al.* described a geographic variation of the *CCL3L1*-containing SD, with African populations possessed a significantly higher copy number than non-Africans. Furthermore, within populations, a lower gene dose relative to the average copy number in the population was demonstrated to confer HIV-1 susceptibility, as well as the increased risk of progression to AIDS or death. *CCL3L1* encodes a ligand for the receptor CCR5, which in turn interacts with the HIV-1 virus glycoprotein (gp) 120 (Gonzalez et al. 2005).

For the role of CNV in autoimmune disease, Aitman *et al.* demonstrated that copy number changes of the Fc receptor *FGCR3B* and its ortholog in rat *Fcgr3-rs* could play a role in immunologically mediated glomerulonephritis (Aitman et al. 2006). In the rat, a 226 bp deletion of the *Fcgr3-rs* allele in the Wistar Kyoto rat strain was shown to increase susceptibility to crescentic glomerulonephritis. The same study also reported a lower copy number of *FGCR3B* in human patients with lupus glomerulonephritis or lupus erythematosus. In a subsequent study *FGCR3B* was reported to contribute to another two systemic autoimmune diseases: microscopic polyangiitis and Wegener's granulomatosis (Fanciulli et al. 2007). In the immune system, Fc receptors play important roles in the activation and modulation of immune responses. Furthermore, in a separate study, low copy number of the complement component C4, a different effector protein in the immune system, was also shown to involve in systemic lupus erythematosus (Yang et al. 2007).

Research also suggested that copy number of the beta-defensin segmental duplication on chromosome 8 is involved in infectious and inflammatory diseases. The segmental duplication containing *DEFB4* (encoding the protein human beta-defensin 2, hBD-2) and a number of other DEFB genes at the 8p23 β-defensin locus shows variable copy number from 2-12 copies per diploid genome (Hollox et al. 2008). Lower copy number was suggested to predispose to Crohn's disease of the colon (Fellermann et al. 2006), while higher copy number was associated with the risk to psoriasis, a skin inflammatory disease (Hollox et al. 2008). β-defensins are small antimicrobial peptides which play a proinflammatory role (Hollox et al. 2008).

The above studies provided important insights regarding the role of CNVs in complex disease genetics: first, in these complex diseases, copy number variations, and

frequently multi-allelic CNVs (ranging from 2 to >10, e.g. in *CCL3L1* and *DEFB4)*, could play a role in disease susceptibility; secondly, in particular diseases (e.g. *CCL3L1* in HIV-1 susceptibility) it was the copy number in relation to the population average, rather than the copy number *per se*, that confer disease susceptibility (Eichler et al. 2007). This suggests the context of population and ethnicity matters, and that other genetic or environmental factors could act in conjunction with CNVs to confer disease susceptibility; thirdly, the same CNV loci were involved in multiple complex diseases, indicating common disease mechanisms and pathways (e.g. *FGCR3B* in multiple autoimmune diseases); and finally, such a role of multi-allelic CNVs in complex diseases may not be easily captured even in large linkage analysis, as modelled bioinformatically by Hollox *et al.* in the case of psoriasis (Hollox et al. 2008).

In summary, it is now evident that copy number variations in genes could be direct risk factors for a number of human diseases. This list of CNV disease loci is likely to expand tremendously, as results from further genome-wide association studies or analysis of family pedigrees provide systematic analysis of the role of copy number variations.

1.5 Schizophrenia

1.5.1 The Concept of Schizophrenia

In 1896, the German physician Emil Kraepelin first described schizophrenia as a mental illness with a discrete disease entity, which he referred to as “dementia praecox” (dementia of early life) (Kraepelin 1919). The term “schizophrenia”, was later introduced in 1911 by the Swiss psychiatrist Eugen Bleuler (Bleuler 1911). The word derived from its Greek roots “schizo” (split) and “phrene” (mind). It was a refinement to “dementia praecox” (Kahn and Pokorny 1964), since it was recognized that the disease did not necessarily include a dementia process, and could sometimes occur late as well as early in life .

Schizophrenia (OMIM181500) is a debilitating psychiatric illness with a prevalence of 1% worldwide (Jablensky et al. 1992). The disease is characterized by psychotic symptoms such as delusion, hallucination, and disorganized thinking, together with cognitive and social impairment (Andreasen 1995). Schizophrenia usually has its onset in the second or third decade of the patient’s life (Hafner et al. 1998; Hafner and an der Heiden 1999), and males were described to have earlier onset than females (Stromgren 1987). Once the psychotic symptoms emerge, the mental illness could persist for a lifetime with recurrent patterns, leading to severe impairment in social and occupational functioning, as well as massive societal costs and consequences. In particular, mortality rate is elevated among schizophrenia patients (Harris and Barraclough 1998), partially due to a high suicidal rate (Fenton et al. 1997). The society cost of the illness for one year in United States alone was estimated to be in excess of 62.7 billion USD (Wu et al. 2005).

The first treatments for the illness, discovered in the 1930s, were based on electroconvulsive shock (ECS) therapy (by Ugo Cerletti) (Cerletti 1954), insulin coma

therapy (by Manfred Sakel) (Sakel 1954) and prefrontal lobotomy (Jenkins et al. 1954). The latter two were later abandoned due to low efficacy. Drug treatments were developed from the 1950s, and were broadly defined into conventional neuroleptics (which in greek means "to clasp the neuron") (e.g. chlorpromazine); and the atypical antipsychotics (e.g. clozapine). Side effects from these drugs are common, including extra pyramidal symptoms (Kandel 2000) such as Parkinsonism, dystonia and tardive dyskinesia[1] (for conventional antipsychotics), and gain of weight and diabetes (for atypical antipsychotics). Current treatments mainly comprise a combination of medication, psychotherapy and social adjustment (Bustillo et al. 2001). ECS, one of the most controversial procedures in modern psychiatry, is still being used to attenuate symptoms of schizophrenia (Brindley 2008).

[1] Patients with parkinsonism fail to walk normally and usually develop tremor; Patients with dystonia have painful muscle spasms of the head, tongue, or neck; Patients with tardive dyskinesia, a chronic EPS, show features of slow, rhythmic and automatic movements.

1.5.2 Phenotypes and Diagnosis

1.5.2.1 Positive and Negative Symptoms

Both Kraepelin and Bleuler recognized that schizophrenia symptoms tend to cluster into distinct categories, and Bleuler was the first to describe schizophrenia symptoms as "positive" or "negative" (Sass 1989). "Positive symptoms" are symptoms of thoughts, emotion or perceptions that are beyond normal experience. They include hallucinations (false sensory perception, typically auditory but can also be visual, olfactory, tactile etc.), delusions (false beliefs based on incorrect inference of reality), as well as thought and speech disorganization. In contrast, "negative symptoms" are thoughts, emotion or perceptions common in normal people but absent in patients. They include social withdrawal, blunted effect (lack of emotional reactivity), alogia (poverty of speech), and avolition (lack of motivation) (Andreasen 1995; Gelder 1996). Another core feature of schizophrenia is cognitive deficits, specifically the impairment of memory, attention and executive function (Kandel 2000).

1.5.2.2 Endophenotypes

In addition to the aforementioned clinical phenotypes, the concept of endophenotype is also of relevance to the understanding of schizophrenia. Endophenotypes, as defined by Gottesman and Gould, are "measurable components unseen by the unaided eye along the pathway between disease and distal genotype" (Gottesman and Gould 2003). These phenotypes are quantitative and heritable, and in contrast to clinical phenotypes, they usually require the use of special processes or instruments to detect or measure, as they may not be readily observable (Owen et al. 2005a). In schizophrenia, examples of endophenotypes include auditory and sensory gating deficit (e.g. P50 prepulse inhibition, investigating acoustic startle reflex), oculomotor functioning (e.g. antisaccade

performance, concerning the rapid redirection of gaze to locations of interest), and cognitive phenotypes such as working memory and verbal memory (Braff et al. 2008).

1.5.2.3 Course of Illness

Schizophrenia is a chronic mental illness. The course of illness can be recognized as three phases (American Psychiatric Association. 1994): (a) the prodromal phase, usually occur in teenage years preceding onset, when patients exhibit mild phenotypes such as social isolation and movement disorder; (b) the acute phase characterized mainly by psychotic symptoms such as delusion and hallucination; and (c) the residual phase, similar to prodromal phase with less severe symptoms than during actue phase, sometimes negative features such as blunted effect are common. Over the duration of disease, patients typically enter numerous acute phases of psychotic episodes, intermittent with numerous residual phases with variable degree of remission. Scales such as the Brief Psychiatric Rating Scale, first published by Overall and Gorham in 1960s, have been used by clinicians to estimate the course of illness (Leucht et al. 2005).

1.5.2.4 Standardised Diagnostic Methods

Current diagnostic methods for schizophrenia evolved from the work of Kraepelin as well as the descriptions of first rank symptoms by Kurt Schneider in 1959 (Bertelsen 2002). The most influential diagnostic guidelines are DSM-IV (Diagnostic and Statistic Manual of Mental Disorder- 4th Edition) (American Psychiatric Association. 1994); and ICD-10 (Classification of Mental and Behavioural Disorders) from the World Health Organization (WHO 1992). The two diagnostic methods are partially concordant (Bertelsen 2002), and a comparison is presented in Table 1.1.

Table 1.1 Two most influential diagnostic guidelines for schizophrenia: the DSM-IV and ICD-10. *(table adapted from Gelder et al., 1996)*

DSM-IV	ICD-10
At least two of the symptoms in (a) must be present for a minimum of one month. Additionally, (b) must be present for six months, and criteria (c)-(e) must be fulfilled	Minimum of one clear symptoms, or at least two if less clear cut, as listed in (a)-(d), or at least two of the symptoms in (e)-(i) should be present for one month of more
(a) delusion, hallucinations, bizarre behavior and negative symptoms	**(a)** thought echo, thought insertion or withdrawal, and thought broadcasting
(b) occupational or social dysfunction	**(b)** delusion of control or passivity
(c) schizoaffective or mood disorder exclusion	**(c)** hallucinatory voices
(d) disturbance must not be due to medication or drug abuse	**(d)** persistent delusions
(e) if a patient has a pervasive development disorder, prominent delusions or hallucinations must be present for one month	**(e)** significant and consistent change of personal behavior
	(f) negative symptoms that are not due to depression or neuroleptic treatment
	(g) persistent hallucinations accompanied by deletions or by persistent over-valued ideas consistently occurring for weeks or months
	(h) incoherent or irrelvant speech
	(i) catatonic behavior

1.5.3 Aetiology and Neurobiology

Schizophrenia is a mental illness with a complex aetiology. A complete picture of the neurobiology of the disease is yet to be determined, but studies regarding the neurochemistry, neurodevelopment and neuropathology aspects of the illness have provided some insights into schizophrenia.

1.5.3.1 Dysfunction of Neurotransmitter Systems

The theories of neurochemical imbalances in schizophrenia originated from pharmacological studies. The dopamine hypothesis was first proposed to explain the aetiology of schizophrenia (van Rossum 1966). The hypothesis was based on two pharmacological observations: (1) First, schizophrenia symptoms improved in response to conventional antipsychotic medications, such as chlorpromazine, which worked through the dopamine receptors. It was established that such traditional neuroleptics worked by competitive blockade of the D2 subtype of dopamine receptor (Creese et al. 1976; Seeman et al. 1976). (2) Secondly, administrations of drugs such as amphetamines increase dopamine levels, concurrently exacerbate psychotic symptoms of schizophrenia (Abi-Dargham et al. 1998). Subsequently, it was suggested that the simple hyperactive dopamine theory cannot explain the complicated aetiology of schizophrenia, including the negative symptoms of the illness, and the depression of dopamine subtypes in the prefrontal cortex. The dopamine hypothesis was later reformulated, proposing increased dopaminergic transmission in the basal ganglia to underlie positive symptoms and psychosis, whereas dopamine deficits in the cortical area were associated with cognitive impairment (Owen et al. 2005b).

The new generation of schizophrenia drug treatment (atypical antipsychotics), however, generally had low affinity for dopamine receptors, yet they were effective in reducing

positive and negative symptoms (Kandel 2000). This implicated the involvement of another neurotransmission system, and led to an alternative hypothesis concerning glutamate transmission, in particular NMDA (N-methyl-D-aspartate) receptor signalling, the major excitatory glutamate transmitter pathway. The first evidence of the theory came from two drugs, phencyclidine (PCP, also known as “angel dust”) and ketamine (also known as “Special K”). Administration of either leads to both positive symptoms (such as hallucinations) and negative symptoms reminiscent of schizophrenia. Furthermore, the drugs were demonstrated to exacerbate psychotic symptoms in chronic schizophrenia patients (Krystal et al. 1994; Lahti et al. 1995). Both drugs act by blocking NMDA receptor subtypes. In contrast, NMDA-receptor agonists such as glycine were found to alleviate schizophrenia symptoms and provided promising clinical results in some studies when used in combination of other drug treatments (Goff et al. 1995; Evins et al. 2002). Finally, mice engineered with a hypomorphic allele (with reduced expression) of NMDA-receptor subunit *NR1* (glutamate receptor, ionotropic, N-methyl D-aspartate 1) were shown to display sensory gating deficits and behavioural phenotypes, including increased motor activity and deficits in social interaction, reminiscent of the human disease (Duncan et al. 2002; Duncan et al. 2004; Morris et al. 2004; Fradley et al. 2005). Similarly, glycine binding site mice mutants also displayed behavioural and cognitive phenotypes, for example hyperactivity, increased startle response and LTP deficits (Ballard et al. 2002).

The dopamine and glutamate hypothesis were both well-established with independent support and evidence. Moreover, the two neurochemical pathways may interact to result in pathogenesis of schizophrenia. One suggestion was that dopaminergic was secondary to glutamatergic transmission (Laruelle et al. 2003). Alternatively, they could act synergistically in the neurocircuit to result in disease phenotypes. In addition, a

number of other neurotransmission systems such as serontonin transmission and GABAergic cortical dysfunction have been implicated in schizophrenia pathogenesis (Carlsson et al. 1999).

1.5.3.2 Neurodevelopment and Neuropathology

Apart from the neurochemical imbalance models, a neurodevelopment theory was also proposed, suggesting schizophrenia could result from abnormal brain development, with pathology originated in the middle stage of intrauterine life (Murray and Lewis 1987; Dazzan and Murray 1999). One support for the neurodevelopment model was the observation that 'soft' neurological signs, such as involuntary movement and dyskinesias, could extend into early ages such as childhood or infancy (Murray and Lewis 1988). Other evidence came from neuropathology, which demonstrated abnormalities of the central nervous system during development. Changes in the cytoarchitecture of the brain, for example increase in ventriclular size in patients, have been identified (Harrison 1999; Harrison and Weinberger 2005). Advance in imaging techniques such as magnetic resonance imaging (MRI) and poisitron emission tomography (PET) scan, as well as improvement in neuroanatomic markers, further enhanced the discoveries in neuropathology. Although some of the findings in the field were inconsistent and the effects of changes might be small, reasonable evidence had fostered the view that such neuropathological abnormality could relate to schizophrenia (Harrison and Weinberger 2005; Owen et al. 2005b). Recent evidence from mouse models further suggested that adult neurogenesis, in particularly in the brain region of dentate gyrus, could be involved in the pathogenesis of schizophrenia (Duan et al. 2007; Yamasaki et al. 2008).

1.5.4 The Genetics of Schizophrenia

1.5.4.1 Evidence of Genetic Contribution

While the aetiology of schizophrenia remains ambiguous, there is consistent evidence for a strong genetic component and inherited factors in schizophrenia (McGuffin et al. 1995), with support from twin, adoption and family studies (Sullivan 2005). Heritability estimates were demonstrated to be ~0.80 (Cardno and Gottesman 2000; Sullivan et al. 2003). Twin studies demonstrated increased risk in monozygotic (~41-65% lifetime prevalence) versus dizygotic twins (up to 28%) (Cardno and Gottesman 2000). Adoption studies established a higher risk in the adoption cluster (offspring of one set of parents raised by unrelated strangers from early in life), and that adoptees have increased risk with regard to biological versus adoptive affected parents (Heston 1966; Kety et al. 1994). Finally, family studies showed a higher risk in first degree relatives (Sullivan 2005). Figure 1.7 summarizes some of these findings.

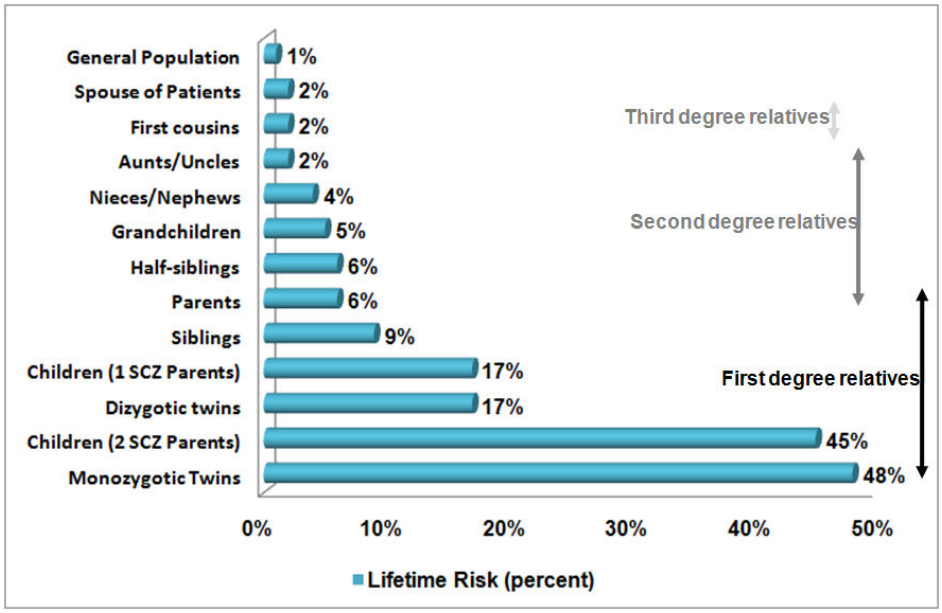


Figure 1.7 Lifetime risk of developing schizophrenia based on relationship. *(diagram adapted from Gottesman 1991)* (Gottesman 1991) The lifetime prevalence of schizophrenia in the general population is 1%. Lifetime risk increases as the degree of relatedness increases. (*SCZ*: schizophrenia)

1.5.4.2 A Search for Candidate Genes

Schizophrenia is a complex trait (Sullivan et al. 2003), probably involving a number of causative genes, with the complication of environmental and incidental factors (Norton et al. 2006). The evidence above suggested the existence of a strong genetic component, however the search for such genetic susceptibility factors has remained a challenge in molecular genetics. A description of the history of schizophrenia genetics would start with early linkage analysis, progressing to association studies of biological or positional candidates, and finally to recent genome-wide association studies (GWAS). These molecular genetic studies were also complemented by early identification of cytogenetic abnormalities, and recently whole-genome discovery and association of copy number variants in schizophrenia (to be discussed in section 1.6).

Standard linkage analysis used large pedigrees (multiplex families) to discover regions of the genome in linkage disequilibrium with the susceptibility locus/ disease allele. Parametric linkage analysis requires specification of the model of segregation, and has been useful in deciphering Mendelian traits. Non-parametric linkage analysis does not require knowledge on the mode of transmission, and was particularly useful for complex diseases such as schizophrenia (Burmeister et al. 2008). Early linkage results were, however, disappointing, probably due to inadequate sample sizes, and that schizophrenia presumably has a large number of susceptibility genes with small effect (Owen et al. 2005a). Meta-analysis only confirmed the inconsistencies of findings and further supported the view of multiple susceptibility loci (Badner and Gershon 2002; Lewis et al. 2003).

Association studies make use of large number of unrelated cases and controls, or family groupings such as sib pairs or trios (parent-child set) to identify genetic loci (Shelling and

Ferguson 2007). Association tests rely on polymorphic DNA markers, for example microsatellites, VNTR (variable number of tandem repeat) and SNP (single nucleotide polymorphism). The aim is to identify markers in linkage disequilibrium with disease variants by association tests. Such studies do not require specification of mode of transmission, can detect genes of small effect (Risch and Merikangas 1996), and improved the resolution of linkage analysis. Early association studies were mostly single-gene efforts, testing association of candidates from positional studies or from biological pathways associated with schizophrenia.

Although not all linkage and association studies in schizophrenia generated reproducible loci, some susceptibility genes or genetic region did receive multiple strands of evidence, and encompassed biologically plausible candidates. The candidates that received the most attention include (a) neuregulin 1 (*NRG1*), which plays a role in both neurodevelopment and glutamatergic synaptic transmission; (b) dysbindin (*DTNBP1*), part of the dystrophin protein complex and is involved in the uptake of glutamate into synaptic vesicles; (c) Disrupted in Schizophrenia 1 (*DISC1*), which plays a neurodevelopment role and was first identified by a translocation segregating in a schizophrenia family; (d) D-amino acid oxidase activator DAOA (G30/G72), which is involved in the oxidisation of d-serine, an endogenous modulator of NMDA receptors; and (e) regulator of G-protein signalling (*RGS4*), which modulates intracellular signalling for many G-protein-coupled receptors (Owen et al. 2005a; Owen et al. 2005b; Sullivan 2005; Burmeister et al. 2008) (Table 1.2).

Table 1.2 Schizophrenia candidate genes with evidence from linkage and association analysis. *(table modified from Burmeister et al., 2008; Sullivan et al., 2005)*

Gene	Gene Description	OMIM	Cytogenetic Band	Cyotegnetic Abnormality	Linkage Evidence	Association Study Support
NRG1	Neuregulin 1	142445	8p12	No	Yes	Yes
DTNBP1	Dystrobrevin binding protein 1	607145	6p22.3	No	Yes	Yes
DISC1	Disrupted in schizophrenia 1	605210	1q42.2	Yes	No	Yes
DAOA	D-amino acid oxidase and activator	607415	13q33.2	No	Yes	Yes
RGS4	Regulator of G-protein signaling 4	602516	1q23.3	No	Yes	Yes

Recent advance in molecular genetics included the realization of GWAS (Wollstein et al. 2007). These whole-genome association studies were made possible due to the increase in density and characterization of single nucleotide polymorphisms (SNPs) (Hinds et al. 2005; IHMC 2005), together with the development of genotyping chips.

Two schizophrenia GWAS with modest sample sizes (500–800 cases) first published in 2007 and early 2008 generated one candidate locus (pseudo-autosomal, near cytokine genes) of genome-wide significance (Lencz et al. 2007; Sullivan et al. 2008). The most recent GWA study, however, used a multistage association study design with an initial 479 samples and replication cohorts of 16,726 subjects, provided genome-wide significance at 3 candidate loci, and the strongest evidence at 2q32.1 near *ZNF804A* (a zing finger protein) (O'Donovan et al. 2008).

Finally, the latest SNP genotyping platforms used for GWAS also allowed identification of copy number variants, resulted in some promising findings of schizophrenia susceptibility loci (see section 1.6).

1.5.4.3 Genes and Environment

For full manifestation of the illness, the genetic susceptibility factors which have increasingly received evidence probably act in concert with non-genetic elements such as environmental, epigenetic or other stochastic factors. Evidence of environmental factors came from monozygotic twin studies, which suggested that although ~40-60% of the disease risks can be explained by genetic vulnerability, there is a considerable non-genetic contribution (Sullivan et al. 2003). Some environmental factors which demonstrated association with schizophrenia include obstetric complications, place and time of birth (e.g. urban areas, famine, winter), advanced paternal age, and exposure to infectious agents (e.g. influenza, rubella) (Sullivan 2005).

In summary, schizophrenia is a debilitating illness exhibiting tremendous genetic and phenotypic complexity. Like other complex diseases, schizophrenia is most likely to be multifactorial, with contributions from both genes and the environment. Pharmacological studies, brain imaging, and genetic research have offered us a glimpse of the disease aetiology, with more underlying factors waiting to be unveiled.

1.6 CNV in Schizophrenia and other Psychiatric Diseases

1.6.1 Early Studies on Chromosomal Abnormalities in Schizophrenia

With the involvement of CNVs being established for a number of complex diseases and human traits, it is not a major leap to hypothesize that CNVs could also account for the genetic basis of psychiatric disease, which in essence is a combination of behavioural, psychological and cognitive variations among individuals.

Structural variants have long been identified as mutations causing schizophrenia in familial studies. Several translocations, e.g. at *DISC1* (Disrupted in Schizophrenia 1) and *PDE4B* (Phosphodiesterase 4B), as well as the microdeletion on 22q11 in patients with DiGeorge/Velocardiofacial Syndrome, were all identified as high-penetrant schizophrenia variants through traditional cytogenetics and karyotyping (MacIntyre et al. 2003). These chromosomal abnormalities provided some classical examples endorsing the involvement of rare chromosomal rearrangements in the pathogenesis of schizophrenia.

1.6.1.1 DISC1 and Breakpoint Study in Schizophrenia

Disrupted in Schizophrenia 1 (*DISC1*) is a schizophrenia candidate gene with perhaps the strongest evidence so far, from human genetics, molecular and cell biology to insights from animal models. *DISC1* was first identified as a balanced reciprocal translocation t(1;11) (q42;q14) which cosegregated in a large Scottish family with schizophrenia and other psychiatric diseases (St Clair et al. 1990; Blackwood et al. 2001). The initial pedigree included 21 individuals with schizophrenia, bipolar disorder, recurrent depression and conduct disorder, of which 16 were identified with the 1q42 translocation, giving a highly significant linkage signal with LOD (logarithm of odds) score of >7.0. Two brain-expressed genes were revealed at the breakpoint: Disrupted in Schizophrenia 1 and 2 (*DISC1 and DISC2*) (Millar et al. 2000; Blackwood et al. 2001).

The 1q24 linkage result was replicated in a subsequent study in a Finnish cohort (Ekelund et al. 2001), and the involvement of *DISC1* in schizophrenia was later confirmed in a number of association studies (Hodgkinson et al. 2004; Kockelkorn et al. 2004). A number of mouse models (with missense mutations, truncated *DISC1* or inducible expression) have further provided evidence of *DISC1* being involved in behaviour and learning and memory (Koike et al. 2006; Clapcote et al. 2007; Pletnikov et al. 2008; Shen et al. 2008).

DISC1 plays important roles in neurodevelopment, cytoskeletal function and cAMP signalling (Chubb et al. 2008). The study of *DISC1* and its interacting proteins have provided us with important clues into disease pathogenesis (Brandon 2007). In particular, phosphodiesterase 4B (*PDE4B*), an interacting partner of *DISC1*, was also revealed as a disease-causing chromosomal translocation (Millar et al. 2005). Both *DISC1* and *PDE4B* are involved in cAMP signalling, an important pathway in neuronal signal network and the synapse (Millar et al. 2005; Bradshaw et al. 2008). *PDE4B* in flies (the dunce fly) was demonstrated to have a role in learning and memory, and was targeted by the antidepressant rolipram (Davis and Dauwalder 1991). These results strengthened the role of *DISC1* and its interactors as genetic predisposition factors of schizophrenia.

DISC1 translocation and the related examples suggest how disruption of genes through chromosomal rearrangement could reveal penetrating variants and biological pathways, a basis for further investigation in case-control cohorts and molecular genetics. Although gene disruption by balanced translocation does not involve copy number change, it is highly analogous to copy number variation, where the breakpoints of the rearrangement could lead to disruption of genes associated with behaviour and cognitive phenotypes.

1.6.1.1 22q11 Microdeletion and Schizophrenia

The 22q11 microdeletion syndrome (also known as DiGeorge syndrome, Velocardiofacial Syndrome, or CATCH22), is a complex disorder with patients displaying learning disability, cardiac defect, craniofacial abnormalities and palatal defect (Karayiorgou et al. 1995; Murphy 2002; Paylor and Lindsay 2006). Among patients there is an unusually high prevalence (~30%) of behavioral abnormalities, including schizophrenia, bipolar disorder, autism and other psychosis (Murphy et al. 1999; Murphy and Owen 2001; Fine et al. 2005). A number of linkage studies have also indicated involvement of this region in schizophrenia as well as bipolar disorder (Sklar 2002), with the first evidence coming from Lachman *et al.* (Lachman et al. 1997). The microdeletion was said to be one of the highest known risk factors for schizophrenia (Murphy and Owen 2001), second to family inheritance such as disease concordance among monozygotic twins.

The increased risk of schizophrenia results from hemizygosity of a 3 Mb region (or in some cases a smaller 1.5 Mb region) (Edelmann et al. 1999), and is expected to act through haploinsufficiency of one or more genes, or through unmasking deleterious polymorphism(s) in the region (Wilson et al. 2006a). To this end, a number of mouse models, including mice deleted with a 1.1 Mb region of the 22q11 locus generated from chromosome engineered deficiency, showed deficit in sensorimotor gating and cognitive impairment (Paylor et al. 2001; Paylor and Lindsay 2006).

The exact gene(s) causing the predisposition is unknown, and there are some functional candidates within the deleted region including catechol-O- methyltransferase (*COMT)*, proline dehydrogenase (*PRODH)* and the micro-RNA processing gene DiGeorge Syndrome Critical region 8 (*DGCR8)* (Bray 2008). *COMT* is one of the two principal

enzymes that degrade catecholamines (e.g. dopamine) (Meyer-Lindenberg et al. 2005). In 1996, Lachman *et al.* described a common polymorphic SNP leading to a valine-to-methionine codon substitution in *COMT* (Lachman et al. 1996), which was later demonstrated to alter dopamine in the pre-frontal cortex (Meyer-Lindenberg et al. 2005) and affect performance in working memory tasks in normal human subjects (Egan et al. 2001; Malhotra et al. 2002). Knockout mice of *Comt* also exhibited elevated dopamine levels in the prefrontal cortex and altered emotional and social behavior (Gogos et al. 1998).

The *PRODH/DGCR8* region represented another susceptibility locus within the 22q11 deletion (Liu et al. 2002). Mouse models with homozygous deletion of *PRODH* (Gogos et al. 1999) showed sensorimotor gating deficits, but with different effect. More recently, Stark and colleagues (Stark et al. 2008) generated another mouse model with a deletion of the segment syntenic with the 1.5 Mb deleted interval in human, removing a number of genes including *Comt*, *Prodh2* and *Dgcr8*. Transcriptional profiling of the mice revealed dysregulation of genes outside the deletion region, particularly those involved in synaptic transmission. The mouse model also revealed a rather unexpected role of *Dgcr8* and micro-RNA biogenesis in some of the sensorimotor phenotypes and abnormalities in neuronal morphology (Stark et al. 2008).

Finally, the breakpoints of the deletion were also of interest both in terms of deletion mechanism and gene involvement. The deletion was thought to cluster to several known homologous low copy repeats or segmental duplications, mediating the recurrent rearrangement of the 3 Mb or 1.5 Mb regions (Edelmann et al. 1999; Emanuel and Shaikh 2001; Shaikh et al. 2001; Sharp et al. 2005). Recent studies from Urban *et al.* used Nimblegen oligonucleotide array CGH and high-resolution PEM to detect

breakpoints with much higher accuracy. They revealed differences in patients whose chromosomal abnormalities had been indistinguishable using more conventional cytogenetic methods (Urban et al. 2006). These differences (of up to 14 genes and 200 kb either side of the breakpoint) may play a role in the variability of phenotypes in patients of hitherto identical breakpoints and deletion region.

In short, 22q11 microdeletion demonstrates a model of highly complex gene interaction and genotype-phenotype relationship, resulting in a variety of behavioral phenotypes including schizophrenia. Research into the region has also provided important insights into subsequently identified recurrent loci (ISC 2008; Walsh et al. 2008), many of which are mapped to mutational hotspots of segmental duplication in the genome similar to the 22q11 locus.

1.6.2 Large Scale CNV Screen in Schizophrenia Patients

Chromosomal abnormalities in schizophrenia patients provided the first evidence of copy number variation altering complex behavioural traits and maintaining disease-associated genotypes in the population. Recent genome-wide CNV studies provided emerging evidence of additional rare causative copy number variants in schizophrenia.

1.6.1.1 Summary of CNV Findings

Since 2006, a number of pilot whole-genome schizophrenia copy number screens have been published (Moon et al. 2006; Wilson et al. 2006a). Interpretation of the early results remained difficult, since the CNV techniques were at their infancy and inconsistent results were reported (Sutrala et al. 2007). A later CNV screen targeting 15 candidate genes for schizophrenia, including neuregulin (*NRG1*) and *DISC1*, also revealed no fruitful results (Sutrala et al. 2008).

The year 2008 was a turning point in terms of schizophrenia genetics, with promising results from a number of large scale whole-genome schizophrenia association studies including a number of reports on copy number variations (ISC 2008; Kirov et al. 2008; Rujescu et al. 2008; Stefansson et al. 2008; Vrijenhoek et al. 2008; Walsh et al. 2008; Xu et al. 2008).

The report from Walsh *et al.* was one of the first studies revealing the genomic architecture of schizophrenia based on CNV screening (Walsh et al. 2008). They screened a case cohort of 150 patients, and a second cohort of 83 childhood-onset schizophrenia cases, using both ROMA and SNP arrays. A number of rare variants were revealed; with breakpoints perturbing genes involved a number of key pathways during neurodevelopment. Of particular significance, Walsh *et al.* showed a higher proportion of

rare variants in cases compared to controls, suggesting rare CNVs are schizophrenia-predisposition factors.

Another two reports described results of large-scale genome-wide screens on thousands of schizophrenia patients using SNP arrays (ISC 2008; Stefansson et al. 2008). Collections of patient and control DNA samples were pooled from various clinical and research groups around the world. The first study by the International Schizophrenia Consortium (ISC) reported their CNV analysis on >3000 Caucasian patient samples with ethnically matched controls based on Affymetrix SNP arrays (ISC 2008). The second study by Steffanson *et al.* reported CNV findings on samples mainly from the SGENE consortium in addition to a number of participating groups (herein refer as SGENE+), using Illumina arrays as CNV screening platforms (Stefansson et al. 2008). The two studies combined revealed three novel schizophrenia-associated recurrent CNV loci. The role of CNV mutational burden was also confirmed in the ISC study.

A fourth study came from Xu *et al.* addressing the role of *de novo* copy number variations in schizophrenia (Xu et al. 2008). By screening 200 trios with an affected child (152 sporadic cases and 48 familial cases) and a similar number of non-affected control trios using Affymetrix SNP arrays, Xu and colleagues discovered a higher burden of *de novo* copy number mutations in sporadic disease cases.

And finally reports from Kirov *et al.* and Vrijenhoek *et al.* revealed a number of rare variants with relevance to synaptic development (Kirov et al. 2008; Vrijenhoek et al. 2008). Of particular interest was the gene Neurexin1 (*NRXN1*), which was identified as a CNV locus in both studies, and was later confirmed as a schizophrenia-associated recurrent CNV (Rujescu et al. 2008).

1.6.2.1 Identification of Large Recurrent Schizophrenia Loci

Among the most promising outcome from the various CNV screens was the identification of at least three recurrent loci of statistically significant association with schizophrenia, as seen in the two large-scale case-control studies from ISC and SGENE+ (ISC 2008; Stefansson et al. 2008). The three loci were at chromosome 1q21.1, 15q13.3 and 15q11.2, with the former two replicated in both studies (Figure 1.8). The regions at 1q21 and 15q13 were extremely rare events, occurring in ~0.1-0.3% in cases, and about ten times less frequent in controls (according to SGENE+). Estimated odds ratios were high (>10 in SGENE+ and 6 to 18 in ISC). The region at 15q11 was relatively more frequent among the three, but still at a rare 0.55% in cases (and 5 times less frequent in controls), with odds ratio at 2.7. Table 1.3 summarizes the details of the three schizophrenia-associated CNV loci.

Of the genes affected in the three regions, a number of highly plausible schizophrenia candidates are particularly worth mentioning. At 1q21, the gene connexin-50 (*GJA8)* encodes a gap junction subunit and was previously reported as associated with schizophrenia in case-control and family studies (Ni et al. 2007). This CNV region has also been shown as associated with schizophrenia by linkage analysis (Stefansson et al. 2008). The 15q11.2 CNV partially overlaps with the Prader-Willi Syndrome locus (Christian et al. 1995) and maps to the breakpoints as identified in some of the Prader-Willi patients (Murthy et al. 2007). Within the ~500 kb region there is the cytoplasmic FMR1 interacting protein 1 (*CYFIP1*) gene*,* which encodes a protein that interacts with both the fragile X mental retardation protein FMRP and the translation initiation factor eIF4E (Napoli et al. 2008). The CYFIP1/FMRP interaction regulated mRNA translation in neuronal dendrites, demonstrating its role in synaptic plasticity and brain development. It was noted that both Prader-Willi Syndrome patients and fragile X patients frequently

have autistic features (Rogers et al. 2001; Dimitropoulos and Schultz 2007). Finally at the third locus 15q13.3, the α7 nicotinic receptor gene *CHRNA7* is a schizophrenia candidate gene (Freedman et al. 1997), and was shown to interact with neuregulin1-erbB signalling (Hancock et al. 2008).

A noteworthy characteristic of the three recurrent loci was that they were all flanked by LCRs, which, analogous to the genomic disorders as discussed before, reflects the underlying architecture of the genome (ISC 2008). NAHR, for example, could be mediated by these LCRs, generating recurrent deletions and reciprocal duplications (Lupski 2006). Consistent with this, reciprocal duplications have been observed for these recurrent loci (ISC 2008). The mutational rate for this type of genomic rearrangement hotspot is estimated to be 10^{-6} to 10^{-4}, many folds higher than for SNP (10^{-8}) (Lupski 2007). This may partially explain the recurrent feature of schizophrenia at a similar rate in all populations, and the maintenance of schizophrenia as a disease in the population even though the disease confers reduced fecundity (Stefansson et al. 2008). Analyzing the 3 deletions in the Icelandic population (part of SGENE Consortium) with accurate genealogy provided further evidence for negative selection in these 3 deletions (St Clair 2008).

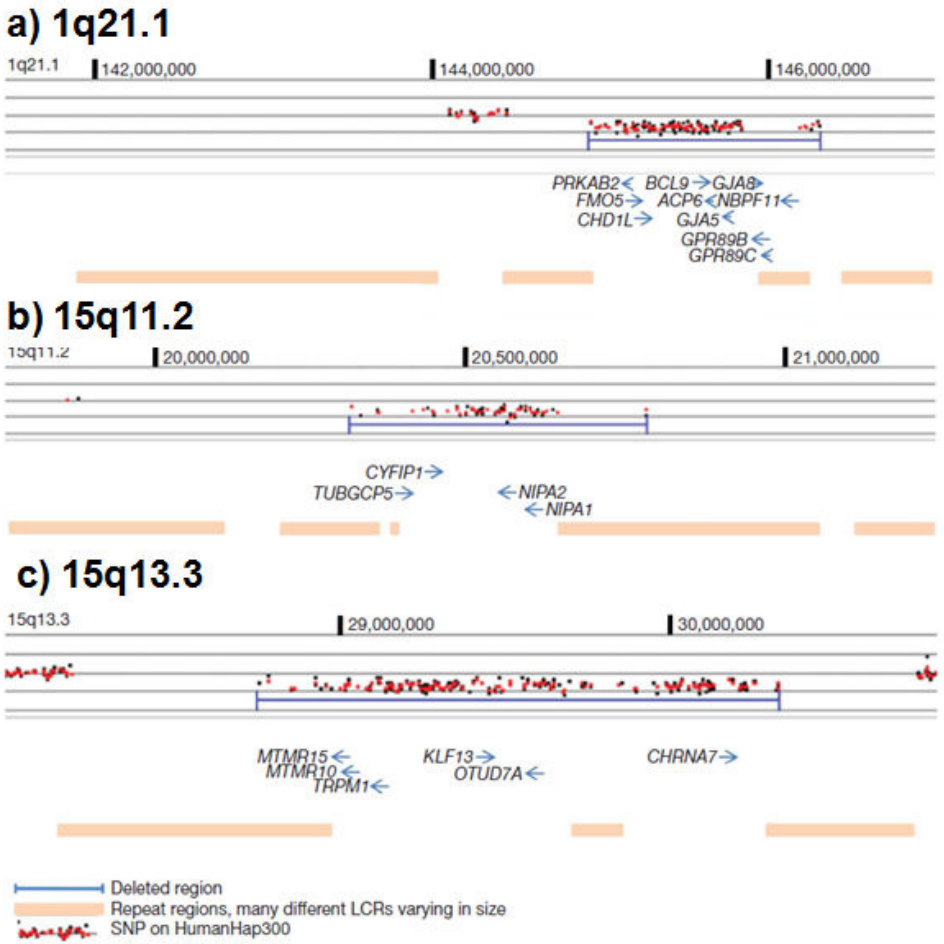


Figure 1.8 Genomic loci of the three novel recurrent deletions associated with schizophrenia (Diagram adopted from Stefansson *et al.*) a) 1q21.1 region; b) 15q11.2 region; c) 15q13.3 region

Table 1.3 Three novel recurrent deletions associated with schizophrenia.

				1q21.1	15q11.2	15q13.3
Estimated Chr Coordinates				chr1:142-146.3 (Mb)	chr15:20.31–20.78 (Mb)	chr15:28 – 31 (Mb)
Estimated Size				~1.3 Mb (small)/ ~3.5 Mb(large)	470kb	~3 Mb
Prevalence	SGENE+	Case	(n=4718)	0.23%	0.55%	0.17%
		Control	(n=41199)	0.02%	0.19%	0.02%
	ISC	Case	(n= 3391)	0.2949%	NA	0.2654%
		Control	(n= 3181)	0.0314%	NA	0.0000%
Odds Ratio (p-value)	SGENE+			14.83 *(p=0.024)*	2.73 *(p=0.007)*	11.54 *(p=0.040)*
	ISC			6.6 *(p=0.023)*	NA	17.9 *(p=0.0023)*
Affected Genes				*SEC22B, NOTCH2NL, HFE2, TXNIP, POLR3GL, LIX1L, RBM8A, PEX11B, ITGA10, ANKRD35, PIAS3, POLR3C, ZNF364,CD160, PDZK1, NBPF11, FAM108A3, PRKAB2, FMO5, CHD1L, BCL9, ACP6, GJA5, GJA8, GPR89A, NM_207400, NBPF15*	*TUBGCP5, CYFIP1, NIPA2, NIPA1, BC044583, WHDC1L1*	*CHRFAM7A, MTMR10, TRPM1, has-mir-211, KLF13, OTUD7A, CHRNA7*

SGENE+: Report from Stefansson et al. 2008; ISC: Report from ISC, 2008

1.6.2.2 Increased Mutation Burden of CNV in Schizophrenia Patients

Another major finding from the CNV screens was the excess of mutational burden as conferred by rare CNVs in cases versus control. This provides important insight into the genetic model of schizophrenia.

The effect of mutational burden by CNVs was evaluated by Walsh *et al.*, the ISC and Xu *et al.* (ISC 2008; Walsh et al. 2008; Xu et al. 2008). All three studies detected a large number of rare variants in the cases, and demonstrated the causal role of these rare variants in increasing disease risks. In particular, Walsh *et al.* reported a three-fold increase of the proportion of rare, genic variants (those that delete or duplicate genes) compared to total CNVs in cases (15%) versus control (5%). The effect was magnified when only childhood-onset cases were taken into account (20%), reflecting a possible increased genetic burden of childhood psychiatric disorders (Walsh et al. 2008).

The ISC study also demonstrated a higher CNV burden of rare CNV (as defined as CNV <1% frequency in all samples screened). The extent of the CNV burden was smaller compared to the report from Walsh *et al.* (from 1.15 to 1.4 times increase). The signal was intensified (to 1.45 to 1.67) when only the extremely rare CNV events – the singletons occurring once in all samples screened - were considered (ISC 2008).

Xu *et al.* added another level of genetic complexity by studying affected trios (sets of parents and affected child) and demonstrated the role of *de novo* rare CNVs in disease. Their results revealed an eight-times increased rate of *de novo* CNVs in sporadic cases compared to controls. The effect was much smaller in familial cases, showing a distinct difference in the genetic determinations of sporadic versus familial cases (Xu et al. 2008).

All these studies came to the conclusion that rare variants (in particular the rare genic CNVs) were collectively significant risk factors for schizophrenia. In sporadic cases in particular, *de novo* CNVs were significantly involved in disease predisposition.

1.6.2.3 Rare Variants Converging into Neurodevelopmental Pathways

Although the role of rare CNVs in schizophrenia pathogenesis has been collectively demonstrated, the bulk of the published schizophrenia CNV regions are composed of unique events, with minimal overlap between studies, apart from the three large recurrent loci. It is therefore extremely difficult to pinpoint which CNV(s) or gene(s) is/are schizophrenia risk factors. As a method to decipher the molecular basis of the disease, one could look for convergence of molecular components into biological pathways. This is possible with the increasing amount of accumulated CNV data on schizophrenia.

One common feature that may start to emerge from these rare events is the involvement of proteins related to synaptic development and functions (Walsh et al. 2008), in particular those that contribute to neuregulin and neurexin signalling in the synapse (Figure 1.9). The strongest evidence comes from the 2p16.3 locus with the gene neurexin1 (*NRXN1*), one of the very few recurrent schizophrenia loci that has been replicated so far in multiple studies. The 2p16.3 locus has been previously shown as associated with autism (Feng et al. 2006; Szatmari et al. 2007; Kim et al. 2008a; Marshall et al. 2008; Zahir et al. 2008). Evidence suggesting its role in schizophrenia first came from Kirov *et al.* (Kirov et al. 2008) who screened 93 patients using array CGH and identified a segregating *NRXN1* CNV in a pair of affected schizophrenia siblings and their asymptomatic mother. Walsh *et al.* also detected a neurexin CNV in a pair of identical twins diagnosed with childhood-onset (COS) schizophrenia, among a cohort of 83 COS patients (Walsh et al. 2008). Subsequently, Vrijenhoek *et al.* identified one CNV

in the region by screening an initial cohort of 54 patients by array CGH, and revealed another 4 cases when screening this locus in more patients (Vrijenhoek et al. 2008). Furthermore, the ISC detected a number of neurexin CNVs with different breakpoints in cases (as well as in controls) (ISC 2008).

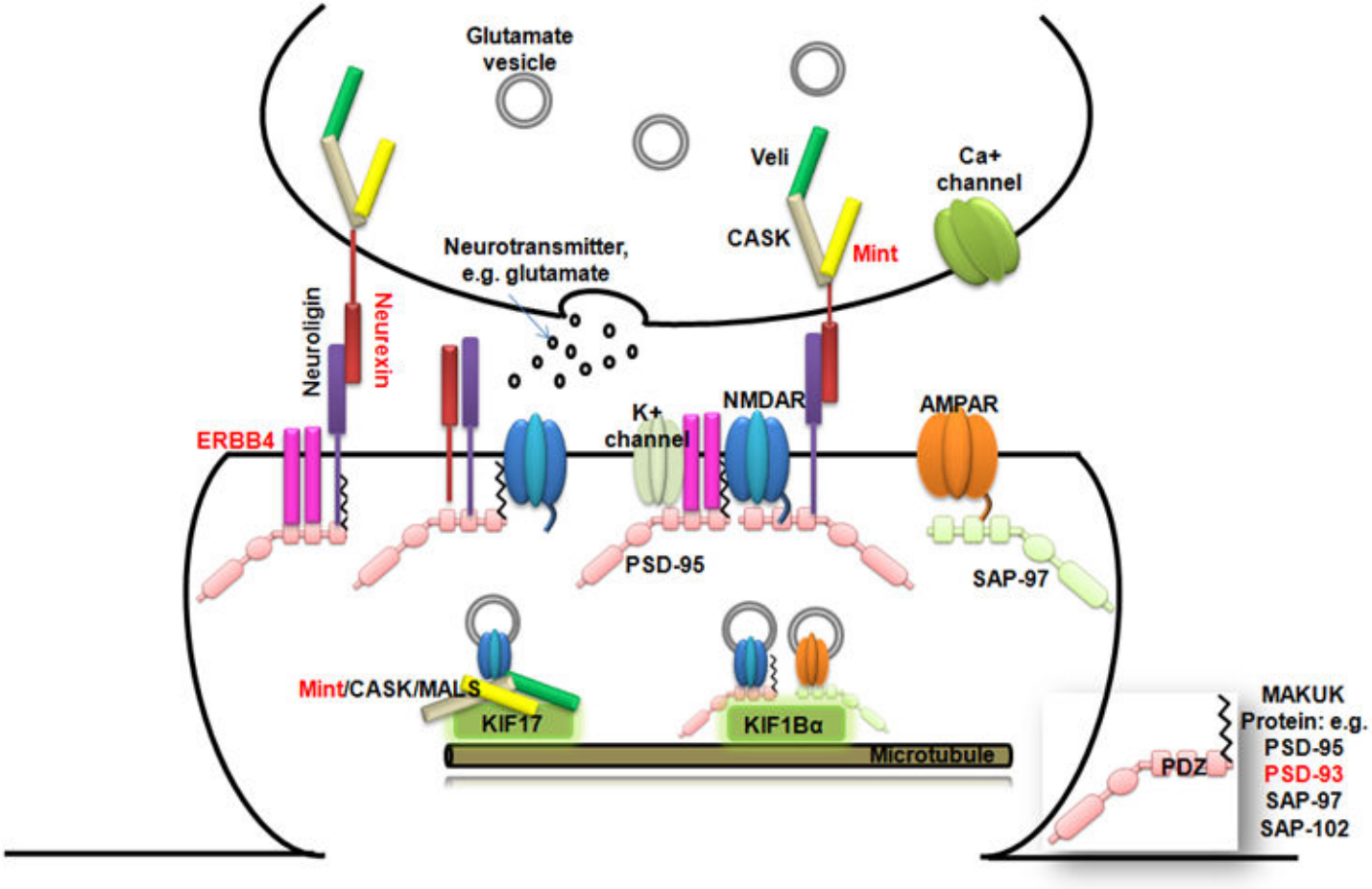


Figure 1.9 Schematic diagram of the synapse displaying known involvement of synaptic proteins in schizophrenia CNV loci (Scannevin and Huganir 2000; Kim and Sheng 2004; Keith and El-Husseini 2008).
Red: proteins identified as schizophrenia CNV loci. (legend to be continued on next page)

*ERBB4, v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4; Mint, amyloid beta A4 precursor protein-binding; CASK, calcium/calmodulin-dependent serine protein; MALS, lin-7 homolog; K+ channel, potassium channel; NMDAR, NMDA (*N*-methyl-D-aspartate) receptor; AMPAR, AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor; Ca+ channel,calcium channel; KIF17 kinesin family member 17; KIF1Bα, kinesin family member 1B α; PSD-95, postsynaptic density protein 95; PSD-93, postsynaptic density protein 93; SAP-97, synapse-associated protein 97; SAP-102 synapse-associated protein 10*

Apart from the above genome-wide studies, Rejuscu *et al.* recently targeted 3 neurexin genes (*NXRN1*, *NXRN2* and *NXRN3*) using a candidate CNV approach to look for structural rearrangement in ~3000 patients and >30,000 controls (Rujescu et al. 2008). They identified 61 *NRXN1* deletions (one *de novo*) and 5 duplications throughout the locus. By restricting the analysis to CNVs that disrupt exons, the authors revealed a ten-fold increase in mutational burden, with 0.17% in cases harbouring such CNVs compared to 0.020% in controls (odds ratio ~10). This recurrent locus distinguishes itself from the three aforementioned large recurrent deletions in two ways: first the breakpoints of the neurexin CNVs were not flanked by LCRs. There was no common breakpoint (consistent with previous findings), and the CNVs vary in size (18 kb to 420 kb). Long range phasing analysis on the *de novo* case provided evidence against NAHR as a rearrangement-generating mechanism (Rujescu et al. 2008). Secondly, rather than deleting or duplicating whole gene(s) as described previously, CNVs in this locus most likely acted by disrupting the gene *NRXN1*. It was concluded that NRXN1 deletions (in particular the ones that affect exons) confer risk of schizophrenia (Rujescu et al. 2008).

Neurexin 1 belongs to a large family of proteins that act as neuronal cell-surface receptors. Neurexin is concentrated at the pre-synaptic membrane, where it functions as a neuroligin receptor (Reissner et al. 2008). Interestingly, neuroligins were shown to be sufficient to trigger presynaptic differentiation through neurexin (Dean et al. 2003), and neurexins could in turn trigger postsynaptic differentiation (Craig and Kang 2007). This transynaptic neurexin-neuroligin complex is important for excitatory glutamatergic and inhibitory GABAergic synapses in the mammalian brain (Craig and Kang 2007), playing a central role in synapse formation and neurotransmission.

Of direct relevance to the *NXRN1* CNV was a schizophrenia study identifying 3 patients carrying genomic deletions of the *CNTNAP2* gene, whilst no CNV was detected in 512 controls (Friedman et al. 2008). *CNTNAP2* encodes Caspr2 contactin-associated protein 2), a member of the neurexin superfamily (Poliak et al. 1999). Furthermore, the same study which first detected the *NXRN1* disruption in schizophrenia revealed in another patient a CNV involving *APBA2* (*Mint2*) (Kirov et al. 2008), a neuronal adaptor protein that binds neurexins as part of a *CASK*-containing protein complex. It is likely that the complex functions in neurotransmitter synaptic vesicle docking/fusion through recruitment of the vesicle fusion protein Munc-18 to the sites of exocytosis (Biederer and Sudhof 2000).

The study by Walsh *et al.* further revealed a number of CNVs disrupting genes involved in synaptic transmission or neurodevelopment (Walsh et al. 2008). Of particular interest is the neuregulin (*NRG*) signalling pathway, which has diverse roles from neuronal migration, axon guidance, glial cell development, axon myelination, neurotransmitter receptor expression and synapse formation (Mei and Xiong 2008). These important functions position neuregulin as a key signalling protein for synaptic plasticity and neuronal survival.

Genes involved in the *NRG* pathway and disrupted by CNVs included two interacting partners *ERBB4* and *MAGI2* (Walsh et al. 2008). *ERBB4* is a type I transmembrane tyrosine kinase receptor for neuregulin-1 (*NRG1*). It was detected as a deletion in a patient. The neuregulin-erbB complex was suggested to interact with PDZ-containing proteins at neuronal synapses (Garcia et al. 2000). The complex interacts with *MAGI2* at neuronal synapses, detected as duplicated in another patient (Walsh et al. 2008). In addition, *ERBB4* was suggested to be recruited by *PSD95* (a PDZ-containg protein) to

the neurexin-neuroligin complex for synapse formation (Lin et al. 2000), providing a link between the CNVs discussed in this section so far. These evidences provide a glimpse into the complexity of the neuronal network, suggesting how the apparent genetic heterogeneity revealed by the schizophrenia CNV screens could be reconciled.

Lastly, although the three novel recurrent loci further suggested heterogeneity, they also confirmed the convergence of CNV loci to candidate neuronal and synaptic pathways. The α7 nicotinic acetylcholine receptor *CHRNA7A* at 15q13.3 was recently shown to be targeted by neuregulin1-erbB signalling to axons for surface expression in sensory neurons (Hancock et al. 2008). The *CYBIP1* protein deleted at 15q11.2 - together with its interactor, the fragile X mental retardation protein FMRP - regulates mRNA translation at the synapse (Napoli et al. 2008). FMRP was shown to regulate mRNA stability for PSD95 (Todd et al. 2003; Zalfa et al. 2007). Furthermore, in a study comparing the gene expression profiles of three lymphoblastoid cell lines - from (a) an autistic patient with the 15q11-13 duplication, (b) a Fragile X mental retardation patient with autism and (c) a normal control - Neuregulin-2 (*NRG2*) was detected as one of the 68 dsyregulated genes common in both patient cell lines (Nishimura et al. 2007).

In conclusion, investigating copy number variations as schizophrenia genetic risk factors has elucidated not only the genomic architecture of the disorder, but also the biology of the disease. Further screening for schizophrenia CNVs, and integrating them into biological pathways, will offer better understanding of the genetics of schizophrenia as well as other psychiatric diseases.

1.7 Scope of Thesis

The aim of this dissertation is to identify copy number variations which could play a role in schizophrenia pathogenesis. The thesis consists of three main sections:

In the first section (Chapter 3), I aim to assess the role of copy number variations in familial cases of schizophrenia with other psychiatric disorders. Using whole-genome tiling path (WGTP) array CGH, I will look for segregation of CNVs with diseases in 3 families, each with a schizophrenia proband and at least another close affected family member. Rare, family-specific variants that are not identified in normal controls will be further investigated as putative disease-causing variants. I will also use a candidate gene approach to study CNV in an extended pedigree with a high resolution oligonucleotide array, and to look for segregation with disease.

The second section (Chapter 4 and 5) will describe a population-based CNV screen on 91 schizophrenia patients and 92 matched controls using the WGTP platform. Following recent success of similar CNV studies, and their implication of rare variants in disease aetiology (ISC 2008; Stefansson et al. 2008; Walsh et al. 2008), I will focus on the study of rare, schizophrenia-specific CNVs and their role in psychiatric phenotypes. Replication of previous findings will be reported, as well as novel putative disease CNV loci in our dataset, with an emphasis on the assessment of gene content. In addition, I will examine the role of common copy number variations, an approach that has been less explored in previous schizophrenia CNV studies. I will attempt two methods to assess the contribution of frequent CNVs in disease. Candidates generated from these approaches will be tested in an extended cohort using target-specific PCR-based assays.

The last section (Chapter 6) will comprise of a CNV screen on variants overlapping with a set of genes playing crucial roles in behavior and cognition- the NMDA (N-methyl-D-aspartate) receptor complex. This complex consists of synaptic molecules for learning and memory functions. Components of the complex are also associated with various psychiatric disorders including schizophrenia (Grant et al. 2005). I will assess the contribution of CNVs in these genes in normal HapMap individuals. One particular genomic locus, the N-ethylmaleimide-sensitive factor at 17q21, will targeted for further investigation.

Chapter 2
Materials and Methods

2.1 Array Comparative Genome Hybridization (Array CGH)

2.1.1 Patient and Control DNA Samples

The schizophrenia patient cohort consists of 100 Scottish patients, who were recruited through the inpatient and outpatient services of hospitals in South East Scotland. All patients met lifetime DSM-IV criteria for schizophrenia, based on consensus between two trained psychiatrists. All patients gave written consent for their DNA to be used in genetic studies. DNA samples were extracted from blood and were provided by Professor Douglas H Blackwood, the Royal Edinburgh Hospital.

A matched control cohort consists of 100 DNA samples of Scottish people with normal phenotype. The participants were recruited from surviving members of the Scottish Mental Survey of 1921. These participants resided in the Lothian area of Edinburgh, Scotland in 1932, when they were at the age of 11 years old, and their cognitive functions were assessed using the Scottish Mental survey. The participants were recruited between 1999 and 2001 at the age of 79 years, and their cognitive function assessed again. DNA samples were extracted from the blood of the patients at the mean age of 79 years, and were provided by Professor Ian J. Deary, University of Edinburgh.

Each array experiment was performed with a test DNA sample (schizophrenic patient or Lothian Birth control sample) hybridized against the standard reference DNA (*NA10851*) extracted from a normal HapMap lymphoblastoid cell line (*GM10851*; 46,XY) obtained from Coriell Institute for Medical Research.

2.1.2 DNA Labelling

The construction and experimental use of whole genome human tiling path BAC array at the Sanger Institute has been described previously (Fiegler et al. 2006; Fiegler et al. 2007). For each CGH experiment, the test DNA sample was labelled in Cy3-dCTP and the reference (*NA10851*) DNA in Cy5-dCTP using BioPrime labelling kit (BioPrime DNA Labelling Kit, Invitrogen Ltd). To avoid dye bias in labelling, a dye-reversal hybridisation (test DNA labelled in Cy5- dCTP and reference DNA in Cy3-dCTP) was performed for each test DNA. The labelling reaction consists of 150ng of DNA and 60 µL of 2.5x random primer solution, made up with nuclease-free water to 130.5 µL. After denaturing the DNA for 10 min at 100°C, 15 µL of 10x dNTP (deoxynucleoside triphosphate) mix (1 mM dCTP, 2 mM dATP, 2 mM dGTP, and 2 mM dTTP in TE buffer), 1.5 µL of 1 mM Cy5-dCTP or Cy3-dCTP (Amersham, GE Healthcare), and 3 µL of Klenow fragment were added on ice to a final reaction volume of 150 µL. The sample mixture was then incubated at 37°C overnight, and the random priming reaction terminated by adding 15 µL of stop buffer supplied in the kit.

Following DNA labelling, unincorporated nucleotides were removed using Microcon YM-30 Filter Devices (Millipore). 150 µL of the labelled test sample, combined with 150 µL labelled reference sample, was applied to a Microcon column placed in a 1.5 mL microfuge tub and spun at 12,000 rcf (relative centrifugal force) for 5 min. The flow-through containing unincorporated dyes was discarded. Labelled DNA was retained in the filter and was washed with 300 µL HPLC water and spun as before, with flow-through discarded. 100 µL of HPLC water was added to the sample and the filter column was placed upside-down to spin in a fresh tube at 2000 rpm (revolutions per minute) for 2 min to collect DNA sample. 3 µL of each sample was analysed on 2.5% agarose gel

electrophoresis and 1.5 µL of each sample was analysed in a NanoDrop™ ND-1000 or ND-8000 UV-VIS Spectrophotometer to check for successful labelling and amplification. The amplified sample should run as a smear, with the bulk fragments below 500 bp (Figure 2.1).

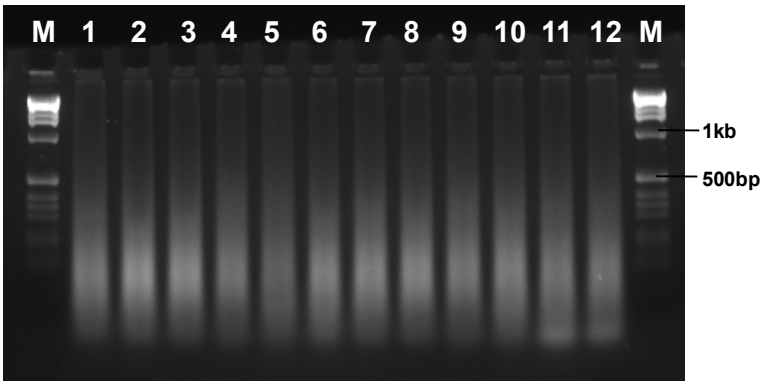


Figure 2.1 Amplified DNA samples after BioPrime Labelling Procedure. 3 µL of each labelled, amplified DNA sample was run on a 2.5% agarose gel stained with ethidium bromide. Most the DNA fragments are below 500bp.

2.1.3 Sample Precipitation and Preparation

For each hybridisation, the combined Cy3 and Cy5 labelled DNA was precipitated with 135 µL of human Cot-1 DNA (1µg/µL, Roche Diagnostics) and 35 µL 3M NaAc (pH 5.2) in ice-cold 100% ethanol (EtOH) at -20oC overnight. The use of excess Cot-1 DNA suppresses repeat sequences that are present in both the test and reference DNA. In separate tubes, 70 µL herring sperm DNA (10 mg/mL, Sigma Aldrich) was simultaneously precipitated using 7 µL 3M NaAc and 100% EtOH as pre-hybridisation. Before hybridisation, samples were spun down and DNA pellets were washed in 500 µL 80% EtOH, then dried and resuspended in 120 µL hybridisation buffer (50 % formamide, 5-7.5 % dextran sulphate, 0.1 % Tween 20, 2x SSC, 10 mM Tris pH 7.4, 10mM Cysteamine). All samples were then denatured for 10 min at 72°C.

2.1.4 Array Hybridisation

Hybridizations were carried out on a Tecan HS Pro Hybridization Station (Tecan Group Ltd.). One microarray slide was fitted into each 63 x 20-mm hybridisation chamber. Slides were subjected to a short wash (0.5 min, PBS/Tween 20/2 mM cysteamine). Prior to sample hybridisation, 100 µL of the herring sperm mixture (containing herring sperm DNA) was injected to each hybridisation chamber for a 45-min pre-hybridisation at 37°C with medium agitation frequency. Subsequently, 100 µL of the hybridisation mixture (containing fluorescent samples) were injected and the samples were hybridised for 21 h at 37°C with medium agitation frequency. Slides were then subjected to a series of stringent washes with PBS/Tween 20/2 mM cysteamine (wash time 0.5 min, soak time 0.5 min, 15 cycles at 37°C), 0.1x SSC (wash time 1.0 min, soak time 2.0 min, 5 cycles at 54°C), PBS/Tween 20/2 mM cysteamine (wash time 0.5 min, soak time 0.5 min, 10 cycles at 23°C), and HPLC water (wash time 0.5, soak time 0.0, 1 cycle at 23°C) before drying for 2.5 min using nitrogen gas. Slides were removed for scanning and the Tecan stations were cleaned with HPLC water using the final system drying program. All hybridisation chambers and plugs were cleaned thoroughly using a cotton bud soaked in HPLC water, followed by a repeat cleaning step with 80% EtOH then left air dried.

2.1.5 Image Acquisition and Data Analysis

Arrays were scanned using an Agilent laser scanner (Agilent Technologies) at a 5 µm resolution setting. Initial data analysis was performed using the BlueFuse software (BlueGnome Ltd). Array spots were defined by the automatic grid feature of the software, manually adjusted where necessary. After correction of local background value, fluorescence intensities of all spots were extracted with a user-defined protocol. Spots with low signal intensities ("amplitude" < 100 in both channels) or inconsistent

fluorescence patterns ("confidence" < 0.5 or "quality" = 0) were excluded. Each spot was then normalised among the array blocks (sub-arrays).

Subsequent analysis was performed with a custom perl script Dyeswapper2.1 (Dimitrios Kalaitzopoulos and Richard Redon, Sanger Institute). The script merges the dye reversal experiments, and determines a central distribution for the log2 signal ratio of all clones, which for a normal individual (with majority of clones in the diploid state) should be a value close to zero. Outlier clones that fall outside a threshold of this central distribution were subsequently identified as copy number variations.

A variability measure, termed SDe, is also determined as an estimation of the standard deviation of the experiment. It is calculated based on the 68.2th percentile (equivalent to one standard deviation from the mean) of the absolute values for all combined log2 ratios for each chromosome (chromosome SDe) and for the whole genome (Global SDe). Dye-reversal experiments for each test DNA were fused and the signal intensity ratios of each clone in the two hybridizations were averaged into a combined ratio (cR). Clone replicates differ >8 SDe were excluded from further analysis. Dye-swap experiments were accepted for subsequent CNV calling only if the following criteria were fulfilled: (1) Global SDe < 0.06; (2) Global clone exclusion rate < 10%; (3) Clone exclusion rate per individual chromosome < 20%.

Copy number changes for each DNA profile were identified using the CNVFinder algorithm (Fiegler et al. 2006) which was based on SDe thresholds. A score (Si) was assigned to the combined ratios of each individual clones in an experiment for CNV calling.

If cR ≥6x SDe or cR ≤ – 6x SDe, then Si = 1 or Si = – 1, respectively.

If cR ≥4x SDe or cR ≤ – 4x SDe, then Si = 0.5 or Si = – 0.5, respectively.

If cR ≥3x SDe or cR ≤ – 3x SDe, then Si = 0.25 or Si = – 0.25, respectively.

If cR ≥1x SDe or cR ≤ – 1x SDe, then Si = 0.1 or Si = – 0.1, respectively.

A stringent 6 x / 4 x SDe ratio threshold was applied: a locus is considered deleted or duplicated where the hybridisation ratio of the corresponding clone exceeds the value of (+/-) 6 x SDe, or if two or more consecutive clones exceed the value of (+/-) 4 x SDe. The 3 x SDe and 1 x SDe thresholds were used to extend CNV regions and to merge clones into a CNV event, for neighbouring clones in the same CNV region with reduced signal response.

The normalized log2ratio for each clone was then plotted against its position along the genome, and the list of CNVs for each dye-swapped experiment was generated.

2.2 Agilent Oligo Custom-Designed Array CGH

2.2.1 Custom Array Designs

Custom arrays were designed with Agilent's eArray microarray design tool. Multiplexed array formats of 4 x 44K were chosen. Probe lengths were optimized to 48-60 bp.

Two custom designs were made (with Richard Redon, Sanger Institute): the first targets 1.5 Mb at chromosome 7p12, with 34036 probes tiling across the region of the gene ATP binding cassette sub-family A (*ABCA13*) where a deletion CNV was identified. The second design targets 1.5Mb at chromosome 9q34, with 34761 probes tiling across a CNV downstream of the gene olfactomedin1 (*OLFM1*). Both designs include 2118 Agilent control probes plus around 9000 probes spaced evenly across all 24 chromosomes.

2.2.2 DNA Labelling

DNA labelling protocol was similar to that for WGTP tiling path array but in half volume reactions and with slight modification to the DNA clean-up steps. Briefly, 150 ng of DNA was added to 30 µL of 2.5x random primer solution (BioPrime Labeling Kit, Invitrogen), and made up with nuclease-free water to 65.25 µL. After denaturing the DNA for 10 min at 100°C, 7.5 µL of 10x dNTP mix (1 mM dCTP, 2 mM dATP, 2 mM dGTP, and 2 mM dTTP in TE buffer), 0.75 µL of 1 mM Cy5-dCTP or Cy3-dCTP (Amersham, GE Healthcare), and 1.5 µL of Klenow fragment (BioPrime Labeling Kit, Invitrogen) were added on ice to a final reaction volume of 75 µL. The reaction was incubated at 37^{o}C overnight over activated charcoal and stopped by 7.5 µL of stop buffer (BioPrimer Labeling Kit, Invitrogen).

Following DNA labelling, unincorporated nucleotides were removed using Microcon YM-30 Filter Devices (Millipore). 405 µL of 1x TE (Promega) was added to each labelled sample, loaded to a Microcon filter placed in a 1.5 mL microfuge tube. Samples were spun at 8,000 rcf for 10 min, followed by a wash with 480 µL 1x TE and spun at 8,000 rcf for 13 min. The filter was inverted and placed into a fresh 1.5 mL microfuge tube and spun for 1 minute. The volume of each eluate was measured and recorded. If the sample volume exceeded 21 µL, the sample would be returned to the filter to be spun again in the inverted position until all samples were < 21 µL. An appropriate volume of 1xTE was then added to bring each sample to 21 µL. 1.5 µL of each sample was analysed in a NanoDrop^{TM} ND-1000 or ND-8000 UV-VIS Spectrophotometer to determine yield and specific activity of the labelling and amplification reaction. Labelled DNA was either combined for the next step of hybridization preparation, or stored overnight at -20^{o}C in the dark for hybridisation on the next day.

2.2.3 Preparation of Labelled Genomic DNA

The appropriate Cy5 and Cy3 samples were combined for a mixture of 39 µL in a new 2 mL microfuge tube. For hybridisation preparation, the following reagents were added: 5 µL Cot-1 DNA (1.0 mg/mL, Roche Diagnostics), 11 µL Agilent 10x Blocking Agent and 55 µL Agilent 2x Hybridization Buffer, bringing the mixture to a final hybridization volume of 110 µL. The samples were mixed by pipetting. Samples were incubated in a 95^{o}C heat block for 3 min, followed by a 30-min incubation in a 37^{o}C heat block.

2.2.4 Array Hybridisation

A clean 4x44K gasket slide was loaded onto the Agilent SureHyb chamber base with the gasket label facing up. 100 µL of the hybridization mixture was slowly dispensed onto the gasket well in a “drag and dispense” manner. All four gasket wells were loaded. The

microarray slide was then placed onto the gasket slide with the “active” side (with the Agilent labelled barcode) facing down. To finish the chamber assembly set-up, the cover was placed onto the slides with the clamp slide across the assembly, followed by hand-tightening of the clamp. The assembled chamber was rotated to wet the slides and to assess the mobility of the bubbles. Stationary bubbles were removed by tapping the assembly on a hard surface. For hybridisation, the assembled slide chamber was left for 24 hours on a rotator rack in a 65 °C oven at a rotating speed of 20 rpm.

2.2.5 Slide Washing

After hybridisation, the chamber was disassembled by loosening the clamp, sliding the clamp off the assembly and removing the chamber cover. The array-gasket sandwich was removed from the chamber base and was submerged into a slide-staining dish (dish #1) which was filled with Oligo array CGH Wash Buffer 1. The sandwich was prised open using a pair of forceps such that the gasket slide would slip away from the microarray slide. The microarray slide was slotted into a slide rack placed in a second stain-dish (dish #2) filled with Oligo array CGH Wash Buffer 1. This step was repeated for up to 4 additional slides from the hybridisation.

The microarray slide(s) were incubated in dish #2 for 5 minutes at room temperature with magnetic stirring. The slide(s) were then transferred to the third slide-staining dish (dish #3), which was filled with Oligo array CGH Wash Buffer 2 pre-warmed to 37^{o}C on a hot plate with magnetic stirrer. The slides were incubated for 1 minute. The slides were then slowly removed from the slide rack. This removal step would take 5-10 seconds and the slides would be dried during the process.

2.2.6 Image Acquisition and Data Analysis

Slides were scanned at 5 μm resolution using an Agilent Scanner with Agilent Scanner Control software v7.0. The image was analyzed using Agilent's Feature Extraction (FE) software V9.5. A protocol which incorporated a custom data filtering and normalization method (Dr Richard Redon & Tomas Fitzgerald, Sanger Institute) was applied for data analysis.

2.3 Polymerase chain reaction (PCR)

All oligonucleotides for use as PCR primers were designed using the web-based program Primer3 (Rozen and Skaletsky 2000) with default parameters unless otherwise stated. Oligonucleotide sequences are listed in appendix 1.

2.3.1 PCR genotyping of the 3p26 Deletion near *CHL1*

General PCR amplification was performed in a 25 µL reaction containing approximately 50-200 ng genomic template DNA, 1.25 U *AmpliTaq* DNA polymerase (Applied Biosystem), 125 µM each dNTP, 5 µM each primer and 1x amplification buffer II with $MgCl_2$ (Applied Biosystem). The cycling protocol is shown in table 2.1.

Table 2.1 Cycling protocol for general PCR amplification A generic PCR cycling protocol consists of an initial denaturation of 94°C for 60 sec, followed by 30 cycles of 94°C for 30 sec (denaturation), 55°C for 30 sec (annealing), and 72°C for 60sec/kb of expected PCR products (extension).

Temperature (°C)	Time (s)	Cycle number
94	60	1
94	30	30
55	30	
72	60/kb	
72	120	1

PCR products were resolved by agarose gel electrophoresis to confirm the presence of the product at expected size.

A deletion at 3p26 5' upstream of the gene Close Homolog of L1 (*CHL1*) was PCR genotyped. Amplification was performed in 25 µL reaction containing approximately 50 ng genomic template DNA and PCR reagent mix as mentioned above. To distinguish the deletion allele from the wild-type allele, PCR primers tiling across the CNV region were designed to validate the deletion polymorphism and to sequence the CNV breakpoints (Figure 2.2). A set of one forward primer (CHL1-fwd) and two reverse primers (CHL1-fwd-5 + wt-rev-2 & del-rev-1) were chosen as the genotyping primers, which generated a ~400bp band corresponding to the wild type allele, and a ~700 bp band corresponding to the deletion allele. Each DNA sample was amplified using the same forward primer and both of the reverse primers. PCR products for each sample were pooled and run on a large 2.5% agarose electrophoresis gel at 200V for 80 min. Figure 2.2 shows the gel electrophoresis analysis for this PCR assay.

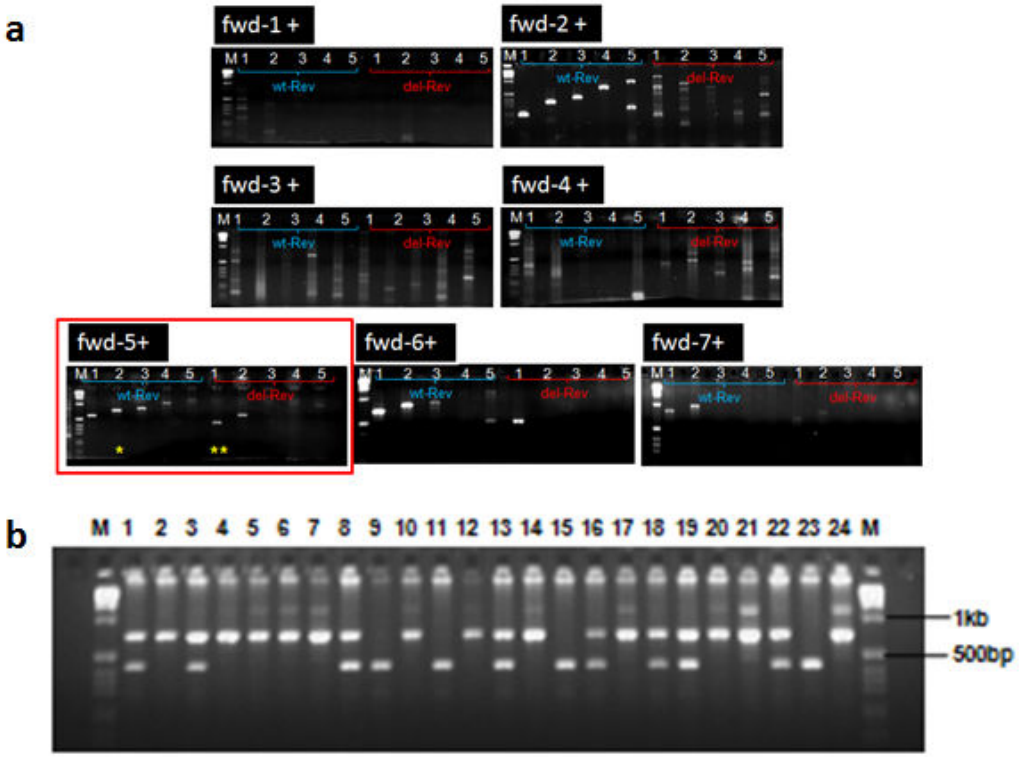


Figure 2.2 PCR genotyping of a deletion at 3p26 5' upstream of Close Homolog of L1 (*CHL1*). a) PCR results from a series of forward primers used in combination of two series of reverse primers, one for the wild-type allele (wt-Rev) and another for the deletion allele (del-Rev). **b)** The selected set of genotyping primer: Under gel electrophoresis analysis, the ~700bp band indicates the presence of the wild-type allele whereas the ~400bp indicates the presence of the deletion allele. A single band at ~700bp or at ~400bp represents homozygous wild-type and homozygous deletion DNA sample respectively, whereas a sample heterozygous for the deletion will show both bands.

2.3.2 Long Range Polymerase Chain Reaction (LR-PCR)

Oligonucleotides for use as long-range PCR primers were designed using Primer3 (Rozen and Skaletsky 2000) with the following parameters: primer melting temperature (Tm) ranges from 58°C to 63°C (optimal 61°C) and primer length ranges from 28 to 32 bp (optimal 30bp).

Long range PCR was performed using Elongase® Enzyme Mix (Invitrogen). PCR amplification was performed in 50µl reaction containing approximately 100 ng of genomic template DNA in a mixture containing 200 µM dNTP, 10 µM each of forward and reverse primer, 1 µl of enzyme mix and the two supplied 5x Buffers combined to give a final Mg2+ concentration of 1.8 mM. The cycling protocol is shown in table 2.2:

Table 2.2 Cycling protocol for long-range PCR amplification Long-range PCR cycling protocol consists of an initial denaturation of 94°C for 30 sec, followed by 35 cycles ^ of 94°C for 30 sec (denaturation), 61°C for 30 sec (annealing), and 68°C for 10 min (extension).

Temperature (°C)	Time	Cycle number
94	30 s	1
94	30 s	
61	30 s	35
68	10 min	
72	2 min	1

PCR products were resolved by 0.8-1% agarose gel electrophoresis to confirm the presence of the product at expected size (Figure 2.3). Selected PCR products were sent for sequencing.

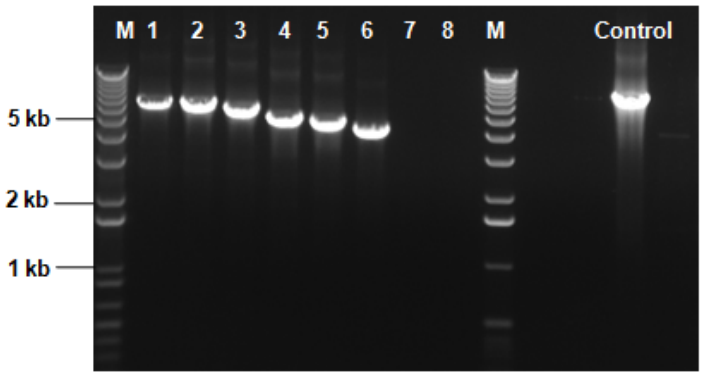


Figure 2.3 Long Range PCR gel electrophoresis analysis of a tandem duplication at Chr1p36. Lane 1-8: Long range PCR results from combinations of the same forward primer with different reverse primers tiling across the breakpoint.

2.3.3 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

2.3.3.1 SYBR Green Method

Oligonucleotides for use as PCR primers in SYBR Green qPCR assays were designed using Primer3 (Rozen and Skaletsky 2000) with parameters as follows: amplicons were designed to be 80 to 150 bp; primer melting temperature (Tm) ranged from 57°C to 63°C (optimal 60°C); primer length from 18 to 27 bp (optimal 20bp); and primer GC content from 20 to 80%. For internal control, a primer set for the gene Beta-2-Microglobulin (β2M) was used as control primer.

SYBR qPCR amplifications were performed in 25 µL reactions containing 10ng of genomic DNA, 12.5 µL of 2x qPCR MasterMix Plus SYBR® green I (Eurogentec) reaction buffer (with HotGoldStar *Taq* polymerase) and 300 nM of forward and reverse primers. For each target region or primer, reactions were performed in triplicate with the test DNA sample(s), and separate triplicate reactions were performed using a reference genomic DNA *NA10851* or *NA15510* (Coriell Institute for Medical Research) as a calibrator. Test primer(s) and control primer for each DNA sample were run on the same plate.

2.3.3.2 Taqman Method with MGB Probes

Oligonucleotides for use as PCR primers and probes in Taqman qPCR assays were designed using the software PrimerExpress® v2.0 (Applied Biosystem) with default settings for MGB probe and primer design. The best score MGB probes and primers with probes ranging from 15-22 nucleotides (nt) were selected.

Taqman qPCR amplifications were performed as duplex reactions with assays for the test and control in the same well. The reaction mixture for a 25 µL reaction was as follows:

Table 2.3 Reagent mixture for each Taqman quantitative PCR assay.

Reagent	Final Concentration/Amount
Control primer (Sigma Aldrich)	900nM
Test primer (Sigma Aldrich)	300nM
5'VIC labeled MGB control probe (Applied Biosystem)	100nM
5'FAM labeled MGB test probe (Applied Biosystem)	100nM
Taqman Universal qPCR Master Mix (Applied Biosystem)	1x
Additional AmipliTaq Polymerase (Applied Biosystem)	2.4 unit
Additional dNTP	50 µM
DNA	~10ng

2.3.3.3 Thermal Cycler and Reaction Condition

Reactions for SYBR or Taqman assays were run in Optical 96-well Reaction Plates (Applied Biosystem) using an ABI Prism 7900 Thermal Cycler with the standard thermal cycling mode as shown in Table 2.4. At the end of the PCR reactions, samples were

subjected to a dissociation curve analysis to confirm specificity of amplification. Figure 2.4 shows examples of a) a dissociation curve showing multiple dissociation peaks (which are discarded for further experiments), and b) a dissociation curve giving a discrete dissociation peak (which are kept for further experiments and analysis),

Table 2.4 Cycling protocol for Quantitative PCR amplification. Quantitative PCR cycling protocol consists of an initial heat activation step (for Hotstart Polymerase in the mastermix) of 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec (denaturation) and 60°C for 60 sec (annealing and extension).

Temperature (^{o}C)	Time	Cycle number
95	10 min	1
95	15 s	40
60	60 s	

2.3.3.4 Standard Curve Generation

For each set of primers, a standard curve was prepared using 2-fold serial dilutions of a reference genomic DNA *NA10851* or *NA15510* from Coriell Institute for Medical Research. Dilutions were made for at least 5 concentrations from 50ng to 3.125ng to completely cover the range of expected DNA copy number changes. Each concentration was done in quadruplicate. Standard curve for the duplex reactions were also performed for the Taqman assays. The slope of the standard curves and qPCR efficiency for each primer was determined. Primer sets with standard curve slope values ranging from -3.0 to -3.9 were accepted. Figure 2.5 shows an example of a typical standard curve.

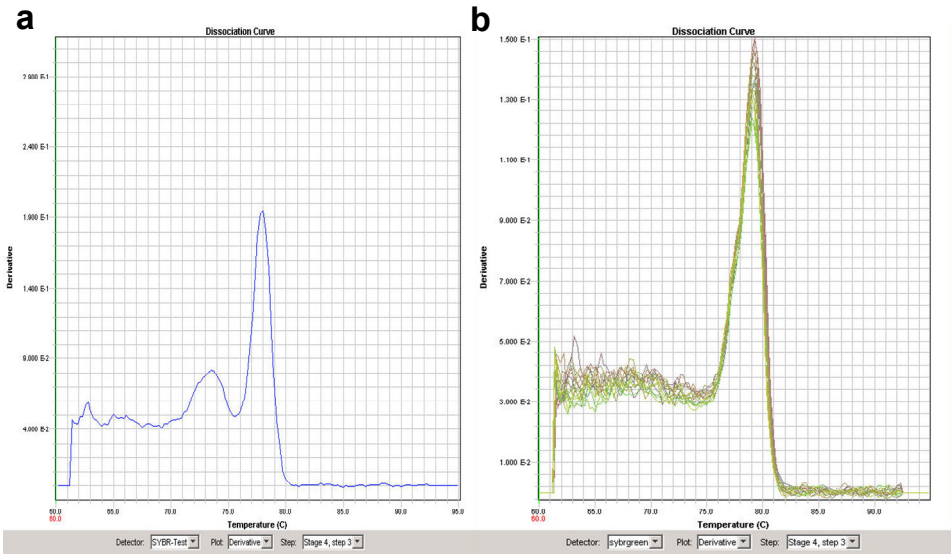


Figure 2.4 Dissociation curves analysis for different primer sets. (a) Example of a primer set giving multiple dissociation peaks. **(b)** Example of a dissociation curve giving a discrete dissociation peak.

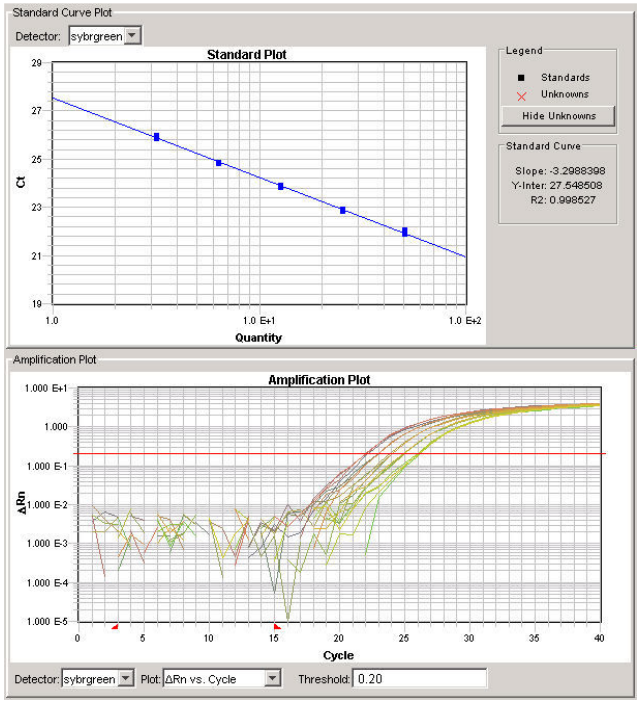


Figure 2.5 A typical standard curve of the control primer. (a) A standard plot of Ct against DNA quantity (in $\log_{10}$ scale) in five 2x dilution from 50 ng to 3.125 ng. **(b)** The amplification plot of the 5 DNA dilutions.

2.3.3.5 DNA Quantification and Data Analysis

Quantification of DNA was based on increased fluorescence in each reaction, expressed in terms of threshold cycle (Ct) values generated by the thermal cycler. Data was exported using the associated SDS 2.2 software in text format. The relative copy number fold change of each target genomic region was normalized to the calibrator DNA. This Test-versus-Calibrator fold change was calculated using the formula below. Standard deviation for each fold-change analysis was also calculated using the following formula (Pfaffl 2001):

$$FoldChangeRatio = \frac{\left(E_{test}\right)^{\Delta Ct(|control-affected|)}}{\left(E_{reference}\right)^{\Delta Ct(|control-affected|)}}$$

where E is a measure of the efficiency or the amplification of the primer. E is calculated by:

$$E = 10^{-1/slope}$$

2.4 Fluorescence In-Situ Hybridisation (FISH)

2.4.1 Growing Cell lines

Lymphoblastoid cell lines were obtained from the Coriell Institute of Medical Research. Cell lines were grown in RPMI 1640 medium (Invitrogen), supplemented with 20% fetal calf serum (Invitrogen) and 2mM L-glutamine/ 100 U/mL penicillin/ 100 µg/mL streptomycin (Sigma). Cultures were incubated at 37ºC and 5% CO2.

2.4.2 Preparation of Extended Chromatin Fibre Slides

Approximately 2-3 ml of healthy cell suspension of each cell line was centrifuged at 1200 rpm for 5 min. Cell pellets were washed twice with PBS and re-suspended in PBS to a final concentration of approximately 2-3 x 10^6 cells/mL. Using a pipette, 10-12 µL of the cell suspension was spread onto the upper part of a polylysine-coated slide (Sigma). Slides were air-dried and fitted into a Perspex Candenza block assembly (made in-house) one at a time, clamped in a nearly vertical position in a bent metal rack. By applying solution to the gap at the top of the Candenza assembly, the slide was washed with 150 µL lysis solution (5 parts 70mM NaOH, 2 parts absolute ethanol) followed by 150 µL of 96% ethanol. The slide was allowed to drain until the meniscus stopped falling (approximately 30s). The Candenza clamps were removed and the slide was carefully lifted away from the Candenza assembly with extended chromatin fibres attached to the slide surface.

Fibre slides were air-dried and fixed with acetic acid (3:1 96% ethanol/acetic acid fixative) for 5 minutes, then dehydrated in an ethanol series (70%, 90%, 100%). Finally, slides were pepsin treated at 37ºC for 5 minutes by incubating in the fixing solution (0.01M HCL with 0.01% pepsin (Sigma)), and were dehydrated again through an ethanol series. Air-dried slides were then stored in a sealed box at –20ºC.

2.4.3 Preparation of Fosmid Clone Insert DNA

Fosmid clones corresponding to DNA sequences within or near the copy number variation regions were ordered from the Sanger Institute. Details of the fosmid clones used in the study are listed in Appendix A. Each fosmid clone was inoculated in 10 mL LB media with 12.5 µg/mL chloramphenicol and was cultured at 37ºC overnight. On the next day, the 10 mL culture was pelleted at 2,000g for 10 minutes. Genomic DNA was isolated from each clone using Phase Prep BAC DNA Kit (Sigma Aldrich) with modification to manufacturer's protocol. 300 µL Resuspension Solution was used to resuspend the cells, followed by the addition 300 µL Lysis Solution. The tubes were then inverted 5 times to allow mixing and incubated for 5 min at room temperature. 300 µL chilled Neutralisation Solution was added and the tubes inverted 8 times before incubation on ice for 5min. The tubes were centrifuged at 15,000g for 20 min at 4°C and the clear supernatant was transferred to a fresh 1.5 mL microfuge tube. 540 µL isopropanol was added to the clear supernatant, and the tubes were centrifuged at 15,000g for 20 min at 4°C to precipitate the DNA. The pellet was washed with 2 mL 70% ethanol and air dried briefly.

The pellet was then resuspended in 540 µL Elution Solution, followed by the addition of 1 µL 1:10 RNase cocktail and a 10-min incubation at 60°C to digest residual RNA (ribonucleic acid). 40 µL Sodium Acetate Buffer Solution (3M, pH 7.0) was added to the tubes to adjust the salt concentration. To remove endotoxin and other impurities, 100 µL of Endotoxin Removal Solution was added followed by a 5-min incubation on ice. The tubes were then warmed at 37°C for 5min prior to centrifugation at 16,000g for 3min. The clear upper phase containing the DNA was transferred to a clean tube and the lower blue phase containing endotoxins and other impurities was discarded. This endotoxin

removal step was repeated once more. 540 µL DNA Precipitation Solution was then added with 1 µL pellet paint to precipitate the DNA again by centrifugation at 21,000g for 20 min at 4°C. The pellets were washed with 150 µL 70% EtOH, air dried and resuspended in 20 µL water. The DNA concentration and quality was assessed using a NanoDrop^{TM} ND-1000 or ND-8000 UV-VIS Spectrophotometer.

2.4.4 Amplification and Labeling of DNA probes

2.4.4.1 GenomePlex® Whole Genome Amplification

10ng of the isolated fosmid DNA (at concentration of 1 ng/µL) was WGA amplified using the GenomePlex® Complete Whole Genome Amplification Kit (Sigma) following manufacturer's instructions. Briefly, 10 ng (1 ng/µL) DNA was fragmented with 1 µL 10x Fragmentation Buffer at 95°C for 4 min and immediately cooled on ice. The fragmented DNA was then used to create a library following the addition of 2 µL 1x Library Preparation Buffer and 1 µL Library Stabilisation Solution. The samples were heated at 95°C for 2 min and cooled on ice prior to the addition of 1 µL Library Preparation Enzyme. Samples were incubated at 16°C for 20 min, 24°C for 20 min, 37°C for 20 min and 75°C for 5 min. For amplification of the library, 47.5 µL nuclease-free water, 7.5 µL 10x Amplification Mastermix and 5 µL WGA DNA Polymerase were added and the samples subjected to a PCR cycling of 95°C for 3min followed by 17 cycles of 94°C for 15 sec and 65°C for 5 min.

2.4.4.2 DNA Fluorescent Labelling

Amplified DNA from the clones was labelled with either Biotin-18-dUTP or Digoxigenin-11-dUTP (Roche Diagnostics) in a PCR reaction containing 1 µL of the amplified DNA from WGA, 60 µM of either Biotin or Digoxigenin labelled dNTP, 40 µM dTTP, mM MgCl2, and 2.5 U BioTaq polymerase (Bioline) in a 25 µL reaction volume. The thermal

conditions were 3 min at 94°C, followed by 17 cycles of 15s at 94 °C and 5min at 65 °C. The fragment sizes of the labelled product were assessed by gel electrophoresis running 1 µL on a 1% agarose gel with ethidium bromide.

2.4.4.3 Probe Fragmentation

To reduce the fragment sizes prior to hybridisation to the fibre slides, 4 µL 1 µg/µL DNAse I (Sigma) was added to the remaining 24 µL labelled DNA and incubated at 15°C for approximately 1hr 30 min. The fragment sizes of each sample were reassessed on 1% agarose gel electrophoresis stained with ethidium bromide, while the remaining samples were kept on ice. Optimal fragment sizes for FISH were considered to be 50-500 bp in length (Figure 2.6). DNase I digestion was resumed if necessary.

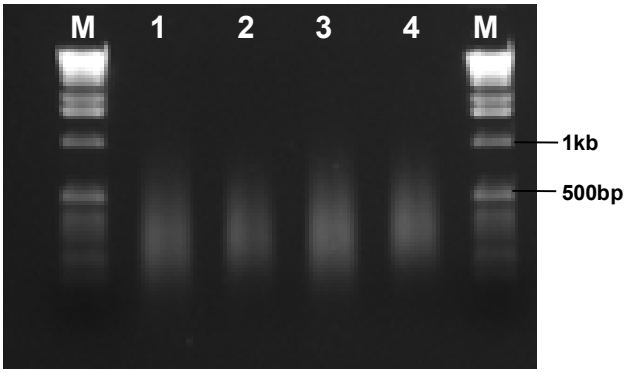


Figure 2.6 Labelled DNA probes for Fluorescent In Situ Hybridisation. Fosmid probes were labelled with either Biotin or Digoxygenin dUTP. Probes were digested by DNase I to reduce DNA fragment sizes to approximately 50-500 bp for optimal hybridisation results. Digested DNA fragment sizes were checked using a 1% agarose gel stained with ethidium bromide.

2.4.5 Immunofluorescence and Image Acquisition

FISH was carried out as follows: Before hybridization, slides were denatured in 70% formamide/ 2x SSC for 2 min at 70°C. Slides were then quenched in 70% ice-cold ethanol, dehydrated through an ethanol series and air-dried.

For each hybridization, approximately 100 ng of each biotin- or digoxigenin-labelled probe was combined with 1 µL Cot1 DNA (1 µg/µL, Roche Diagnostics) and 10 µL hybridization buffer (50% deionised formamide, 2x SSC, 10% dextran sulphate, 0.1% Tween 20, 10mM Tris pH7.4). Labelled probe mixtures were denatured at 65°C for 10 min, followed by a pre-annealing incubation at 37°C for 30 min to 3 hours. Denatured probe mixtures were added to the slides, covered with 22 x 36 mm plastic coverslips, sealed with rubber cement and incubated at 37°C overnight for hybridization. On the next day, rubber cement were removed and slides were incubated twice in 2xSSC (43°C, 5 min), once in 50% formamide for 5 min at 43°C, and once again in 2xSSC (43°C, 5 min). Slides were then transferred to 4xST (4 x SSC, 0.05% Tween) for 5 min at room temperature.

Hybridised probes were detected by two layers of antibodies. Biotin-labeled probes were detected by avidin-FITC (Vector Laboratory), followed by biotinated antiavidin D as secondary antibody (Vector Laboratory). Digoxigenin-labeled probes were visualized by monoclonal mouse anti-dig antibody (Sigma) followed by Texas Red-conjugated goat anti-mouse IgG (Invitrogen, Carlsbad, CA) as secondary antibody. Slides were incubated sequentially in each antibody layer (200 µL under Nescofilm®) for 20 min at 37°C in the dark. In between the two layers of antibody detection, slides were washed in 4xST (5 min, room temp) three times.

After probe detection, each slide was mounted with DAPI (4′,6-diamidino-2-phenylindole) (Vector Labs, Orton Southgate, UK) under a 22 x 50 mm plastic coverslip. Edges of the coverslips were sealed with nail varnish. Slides were stored at 4°C.

Images were captured and processed with the SmartCapture software (Digital Scientific, Cambridge, UK).

Chapter 3
Familial CNV Study
in Schizophrenia

3.1 Whole-Genome CNV Screen in Families

3.1.1 Whole-genome Array CGH Screen in Three Familial Cases

The familial nature of schizophrenia is well-established. While acknowledging the presence of sporadic forms of the disorder, one strategy to understand schizophrenia genetics is to focus on familial cases, by assessing inheritance patterns of mutations within individual families and their segregation with disease. We applied this approach to study transmission of CNVs in families with schizophrenia and other neuropsychiatric conditions. Using the WGTP BAC array as a CNV detection platform, we screened 30 schizophrenia probands, and selected 3 familial cases for screening of CNVs predisposing to psychiatric diseases. In each case the proband was diagnosed with schizophrenia under DSM-IV criteria, and either of the parents was affected, together with at least one affected close relative in the family. The three pedigrees F-29, F-41 and F-192 are shown in Figure 3.1.

Each DNA sample was hybridized twice in dye swap experiments. The algorithm CNVFinder (Fiegler et al. 2006) was applied to detect copy number variants in affected members for each family. To detect potential disease-causing variants we applied a two-stage filtering: (a) to investigate CNVs segregating with disease, only BAC clones detected as CNVs in all affected members within each family were included in our analysis. This resulted in 13, 54 and 55 CNV clones in F-29, F-41 and F-192 respectively; and (b) to investigate rare variants not present in healthy controls, BAC clones frequently deleted/duplicated in the HapMap WGTP control dataset (Redon et al. 2006) (defined as CNVs in >10% of HapMap samples) were considered as benign copy number polymorphisms and therefore removed from analysis.

The analysis returned two CNV regions, the first was a duplication involving two clones (*Chr1tp-23A3* & *Chr1tp-1C7*) at chromosome 1p36 in pedigree F-29, and the second a deletion involving a single clone (*Chr8tp-17G4*) at chromosome 8q22 in pedigree F-41. Quantitative PCR confirmed the 1p36 duplication (Figure 3.2), but did not validate the 8q22 deletion in F-41. Further analysis revealed that log2ratio variations reported by the clone *Chr8tp-17G4* was most likely an experimental artefact.

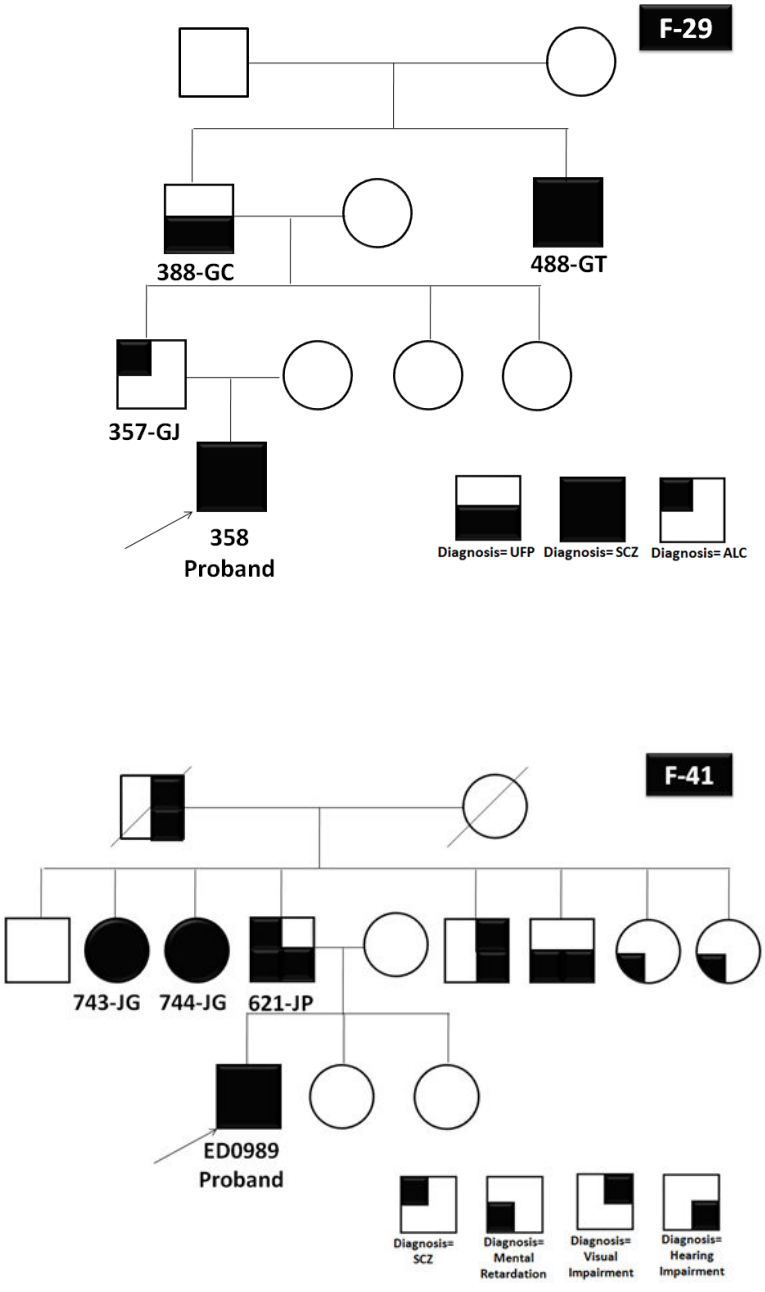


Figure 3.1 (to be continued)

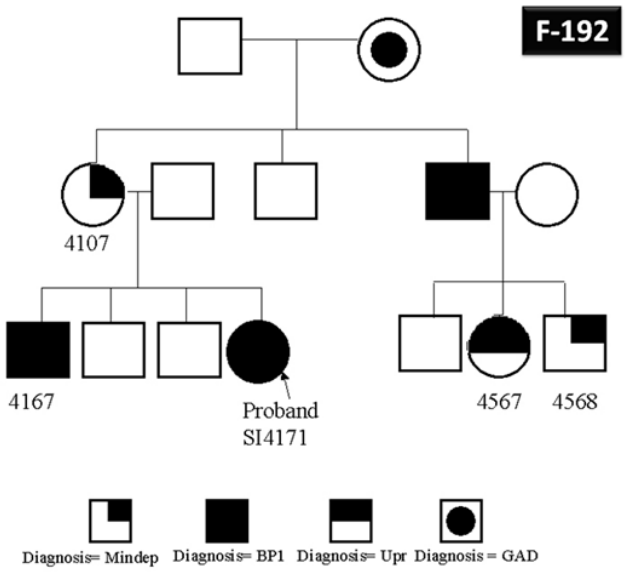


Figure 3.1 Three families analysed by whole-genome CNV screen.
(a) Pedigree F-29; **(b)** Pedigree F-41; **(c)** Pedigree F-192
UFP, unspecified functional psychosis'; SCZ. Schizophrenia; ALC, Alcoholism; Mindep, minor depression; BP1, bipolar affective disorder I; UPR, unipolar depression; GAD, General Anxiety Disorder

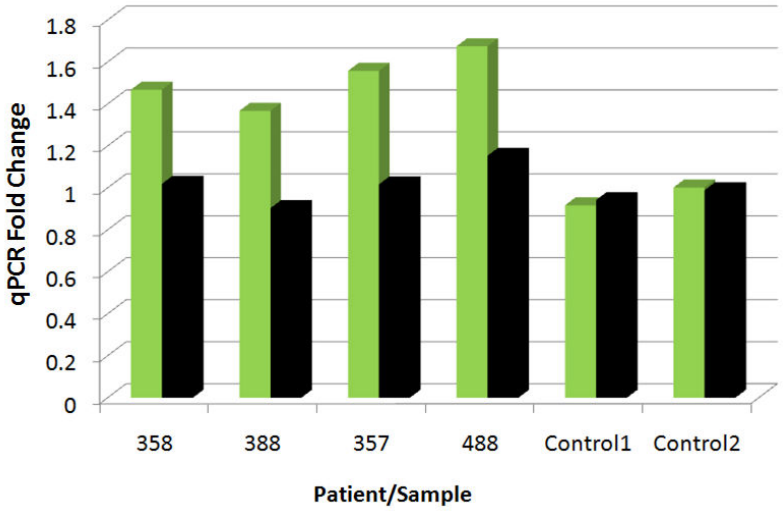


Figure 3.2 qPCR validation of duplication 1p36 in Pedigree F-29. Quantitative PCR using primers located at the 5' end of *SPSB1* within the duplication region (green) validated the duplication in all 4 patients (358, 388, 357, 488). The second set of primers (black) was at the 3' invariant end of *SPSB1*. Duplication was confirmed as detected by the 5' *SPSB1* primer.

3.1.2 Characterization of the Rare Familial Duplication at 1p36.22

At WGTP resolution, the duplication at 1p36.22 in Pedigree F-29 was located at chr1:9189407-9385015 (~200 kb in size). The CNV clones overlapped with two genes, namely hexose-6-phosphate dehydrogenase (*H6PD*) and splA/ryanodine receptor domain and SOCS box containing 1 (*SPSB1*). *H6PD* encodes an enzyme involved in glucose metabolism, and could play a role in unfolded protein response pathways leading to skeletal myopathy (Lavery et al. 2008).

SPSB1 (also known as SSB-1) possesses a SPRY domain (with repeats in splA and RyR) which binds to MET, a receptor protein-tyrosine kinase. It was shown that *SPSB1* could interact with MET to enhance Ras/ERK signalling (Wang et al. 2005). It has been demonstrated - by the use of a mouse model - that ERK (extracellular signal-regulated kinase) signalling plays an important role in cognition. Downstream of neurotransmitters and neuronal receptors, this signalling pathway plays a key function in transducing neuronal signals of downstream transcription machinery, regulating gene expression, and leading to synaptic changes (Brambilla 2003). For example, overexpression of the *Fyn* kinase in mice results in depletion of phosphorylated ERK and other key neuronal signalling components and causes impaired spatial memory in the mouse model of Alzheimer's Disease (Chin et al. 2005). SynGap, a Ras GTPase-activating protein with close association to NMDA receptor, by regulating ERK signalling, can lead to changes in synaptic activity including long term potentiation (LTP) (Komiyama et al. 2002).

We further characterized the structure of this rare duplication identified in psychiatric disease carriers, by performing fluorescence in situ hybridization on extended chromatin fibres (Fiber-FISH) prepared from a patient cell line. Two overlapping probes were hybridized, the first located entirely at the 5' duplicated region of *SPSB1* and the second

one covering both the 5' duplicated and the 3' invariant region of *SPSB1*. The CNV was demonstrated to be a tandem duplication in the same orientation (Figure 3.3).

To map the duplication breakpoints, we used a conventional long-range PCR method with primers tiling across 10 kb either side of the estimated breakpoints to PCR amplify the breakpoint junction. PCR fragments were subsequently sequenced, and the CNV was delineated as a duplicated segment of 86 kb in length at chr1:9240833-9326667 (Figure 3.4). The 3' breakpoint falls on the first large intron of the 5' end of *SPSB1*, within a SINE element (*AluY* repeat). The 5' breakpoint falls on the second last intron of the 3' end of *H6PD*, and is located near another *AluY* repeat. The region is also enriched with simple tandem repeats. These SINE and simple elements may act as chromosomal fragile sites for chromosomal rearrangements. For example, *Alu* repeats have been suggested to predispose to genome instability (Stankiewicz and Lupski 2002).

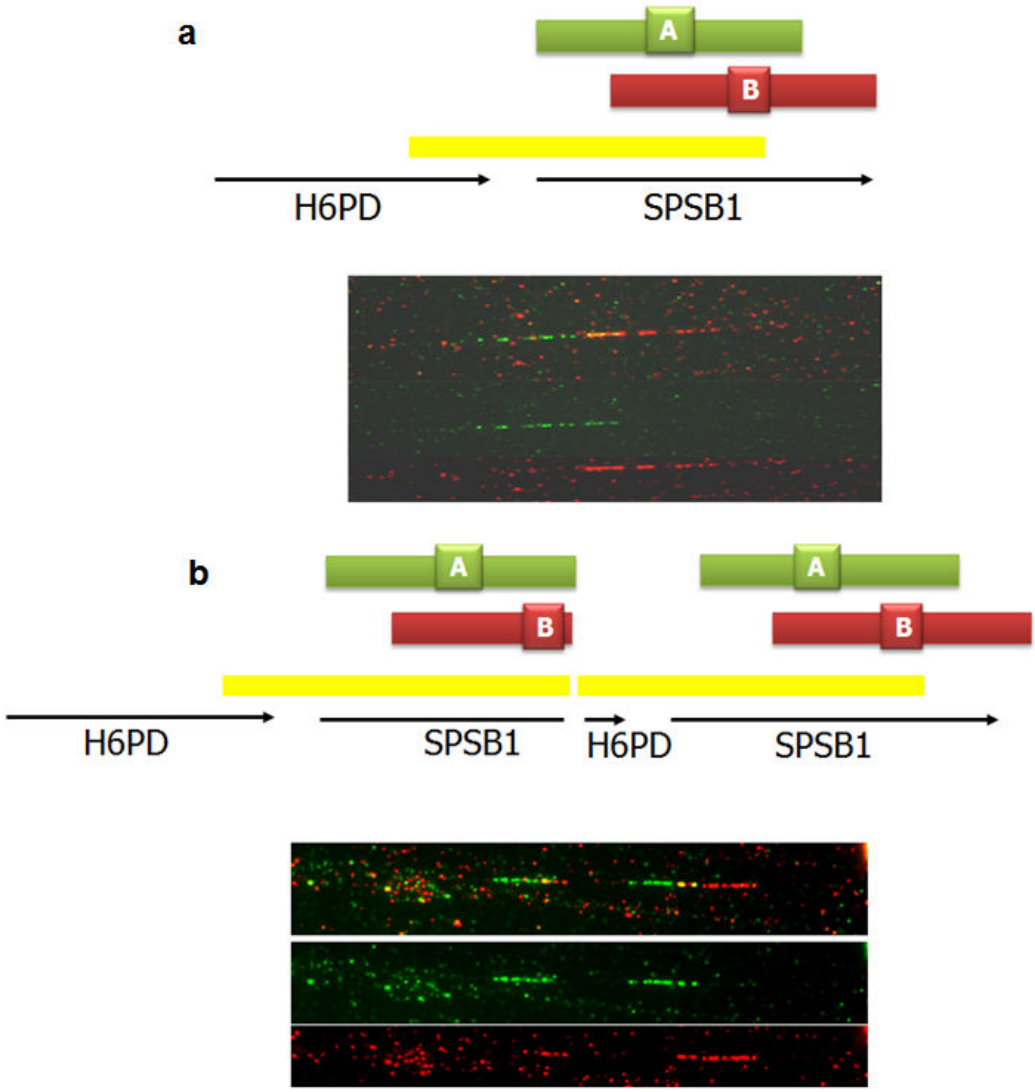


Figure 3.3 Delineation of 1p36.22 duplication structure by Fiber-FISH. Fiber-FISH was performed with two overlapping probes: Probe A (green), located at the 5' duplicated region of *SPSB1* and Probe B (red), partially overlapped with the 3' invariant region of *SPSB1*. FISH results for **a)** a normal control cell line **b)** a patient cell line (358) showing duplication. (yellow block: estimated duplication region)

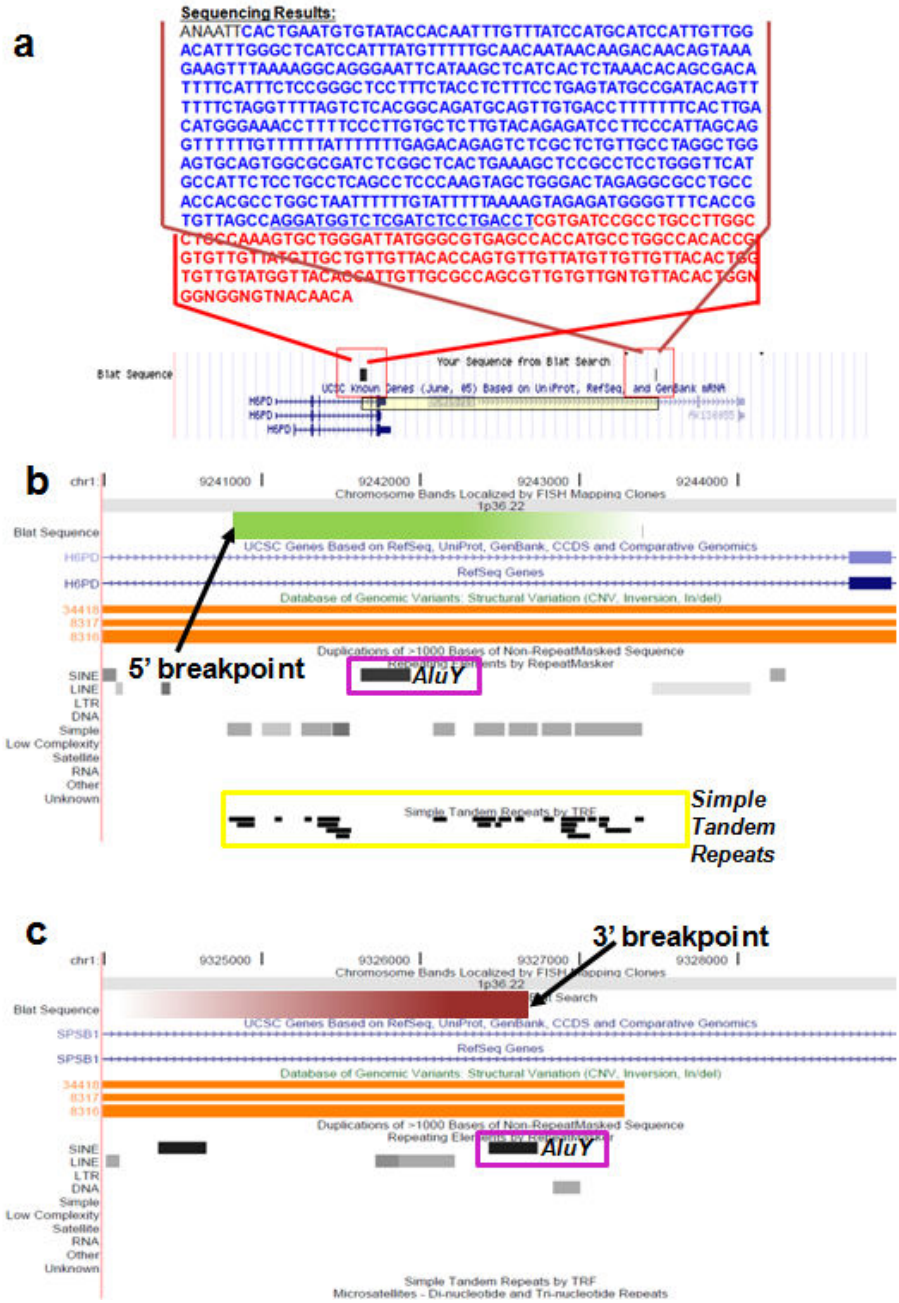


Figure 3.4 Sequencing breakpoints of 1p36.22 revealed repeat structures. a) CNV breakpoint junction was amplified using long range PCR. The amplified fragment was sequenced for breakpoint delineation. **b)** 5' breakpoint was mapped to chr1: 9240833 bp, the second last intron of *H6PD*. A block of SINE element (*AluY* repeat) was identified near the breakpoint, as well as a number of simple tandem repeats. **c)** 3' breakpoint was mapped to chr1: 9326667 bp, intron 1 of *SPSB1*. The breakpoint was located at block of *AluY* repeat sequences.

3.1.3 Known Genomic Rearrangements Near the 1p36.22 duplication

At the time we discovered the 1p36 duplication (in 2006) there was no reported chromosomal rearrangements at the region. The duplication was proposed to be a rare, private variant, potentially a disease predisposition factor within the affected family. As array CGH and other CNV detection techniques matured, more normal variants were detected and deposited into public databases, for example the Database of Genomic Variant (DGV). At the time of writing (in 2008), there are two reports in DGV of copy number variants in the 1p36.22 region (Pinto et al. 2007; Zogopoulos et al. 2007). Zogopoulos *et al.* reported 2 duplications in a control cohort of 1190 samples from Canada. Pinto et al. reported 4 duplications in a cohort of 776 controls (506 Germans & 270 HapMap samples). The 1p36.22 duplication is therefore recurrent in the population and is not unique to affected family members of F-29. We also noticed that the duplication as detected in these samples all have slightly different breakpoints, falling on different repeat sequences, even in the same study (Pinto et al. 2007). This example implicated potential genetic heterogeneity of a recurrent CNV at the sequence level. Alternatively, we cannot rule out that limited resolution of various detection methods compared to PCR sequencing could lead to imprecision in reported CNV breakpoints.

Table 3.1 Known genomic rearrangement at 1p36.22. Data and information gathered for the 1p36.22 region from the Database of Genomic Variant (DGV), compared with results from our current study.

Study	CNV type	# of samples	chr coodinates	method/platform	cohort info
Zogopoulos et al. (2007)	gain	2	chr1:9255290-9316595	Affymetrix 500K and 100K	1190 control samples (Ontario, Canada)
Pinto et al. (2007)	gain	1	chr1:9250000-9311805	Affymetrix 500K	776 control samples (506 Germans, 270 HapMap)
Pinto et al. (2007)	gain	3	chr1:9238434-9327280	Affymetrix 500K	776 control samples (506 Germans, 270 HapMap)
Redon et al. (2006)	NA	0	NA	WGTP BAC array	270 HapMap
Current Study	gain	4	chr1:9240833-9326667	WGTP BAC array and sequencing	3 schizophrenia families and 92 LBC controls

3.2 Deletion at *ABCA13* in an Extended Family with Schizophrenia

3.2.1 Evidence of Functional Mutations of *ABCA13* at 7p12.3

In collaboration with the Psychiatric Genetics group in University of Edinburgh, we investigated whether functional mutations of the ATP Binding Cassette Gene 13 (ABCA13) at chromosome 7p12.13 region could be associated with schizophrenia. A previous cytogenetic study had identified a translocation breakpoint at this locus in a patient with schizophrenia. Moreover, subsequent mutation screening within the ABCA13 gene detected an excess of rare, truncation SNP mutations in patients compared to control (personal communication, Professor Douglas Blackwood).

ABCA13 belongs to the ATP Binding Cassette (ABC) family of integral membrane proteins which regulates the transport of a wide variety of substrates across biological membranes (Dean et al. 2001). ABCA13 is an unusually large gene, spanning 450 kb across chromosome 7p12.13. It has a modular structure with two elements: a hydrophobic, transmembrane structure at the N-terminus, followed by a cytosolic domain which contains the ATP-binding domain (Prades et al. 2002). Although the potential function of ABCA13 in the brain or in central nervous system is not well known, a number of ABC transporters have been implicated in neurological diseases, due to their function in brain lipoprotein transport and their involvement in metabolism and homeostasis (Kim et al. 2008). ABCA1, for instance, was shown to be involved in the metabolism of cholesterol, phospholipids and apolipoprotein in the central nervous system (Hirsch-Reinshagen et al. 2004). Cholesterol homoeostasis at the neuronal membrane plays an important role in the processing of amyloid precursor protein (APP) and the regulation of beta-amyloid peptide levels. Several independent studies have demonstrated a role for ABC transporters in APP processing and their involvement in

Alzheimer's disease (Koldamova et al. 2003; Wahrle et al. 2005; Chan et al. 2008; Uehara et al. 2008).

In addition, ABC transporters are active drug efflux transporters across bio-membranes, and some members (especially those in the ABCB subfamily) are responsible for drug transport at the blood-brain barrier (Loscher and Potschka 2005). In line with this, polymorphisms and mutations in genes belonging to the ABC family have been implicated in drug metabolism (Cascorbi 2006; Hermann and Bassetti 2007): for instance, some ABCB1 genotypes were linked with responses to anti-depressant drug treatment (Takao et al. 2006; Uhr et al. 2008).

3.2.2 Oligo Array CNV Screen in an Extended Pedigree

To identify copy number changes in the region, a custom oligo array (Agilent Technologies) was designed with 34,036 probes covering 1.5 Mb of sequence spanning ABCA13 at chromosome 7p12. Validation experiments demonstrated the reliability and sensitivity of this custom array (see Appendix B).

Array CGH was then performed on DNA samples from 5 patients with possible CNV at ABCA13 (Family 10, 18, 178, 340 and a sporadic case) (Table 3.2). A deletion of approximately 11 kb in size was detected in Patient 4398 (with schizophrenia) from Family 340. The deletion is located in intron 53-54 of ABCA13 at chr7:48,552-48,563 kb (Figure 3.5). We subsequently performed CNV screening on custom arrays for all DNA samples from Family 340, presenting with varied psychiatric phenotypes (Figure 3.6). Out of 11 family members with DNA samples available, 6 have psychiatric disorders ranging from schizophrenia, depression (with variable degree) and anxiety disorder. 5 out of these 6 affected members carry the deletion.

In the first generation, the ABCA13 CNV was present in Sample 5688. Clinical details and DNA sample was not available from her deceased partner. The deletion was inherited by all members in the second generation, three of whom (4398 (proband), 5671 & 5725) had severe psychiatric disorders.

In the third generation, the deletion was present in two siblings (5687 & 5690) who had single episode of depression. However, another sibling with depression (5693) did not inherit the CNV. It is also noted that individual 5713, indicated as a normal individual in the pedigree, had a presumed single episode of depression which did not pass DSM-IV criteria for depression.

Table 3.2 Schizophrenia patients with DNA analysed on custom arrays.

	Schizophrenia Patient	Family	CNV
Proband 1	7808	178	no
Proband 2	7804	18	no
Proband 3	3341	10	no
Proband 4	4398	340	yes
Proband 5	7812	sporadic case	no

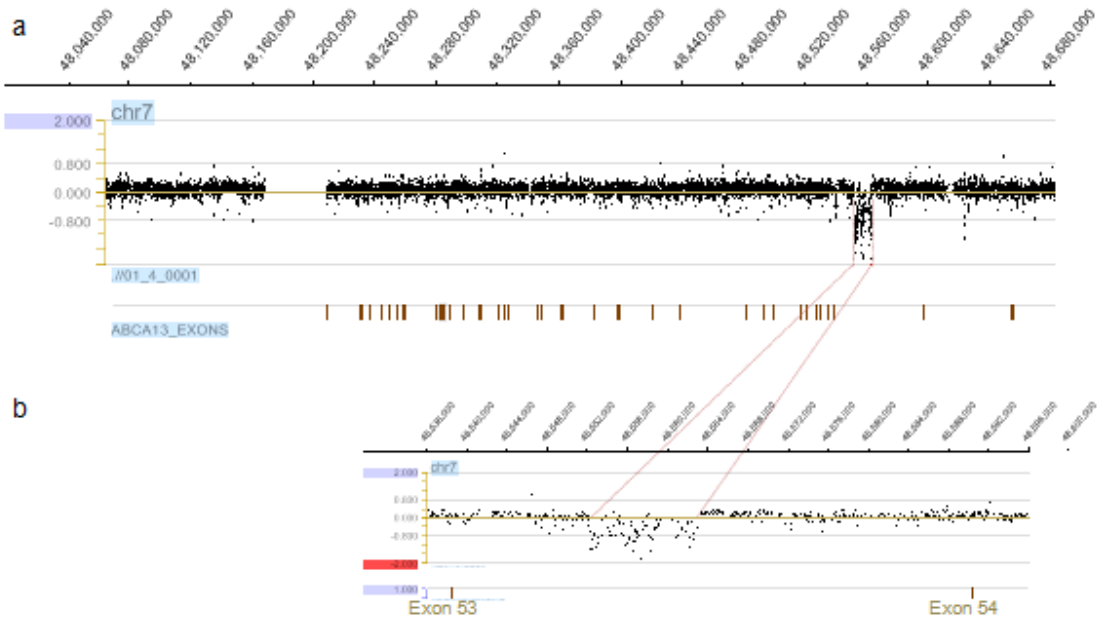
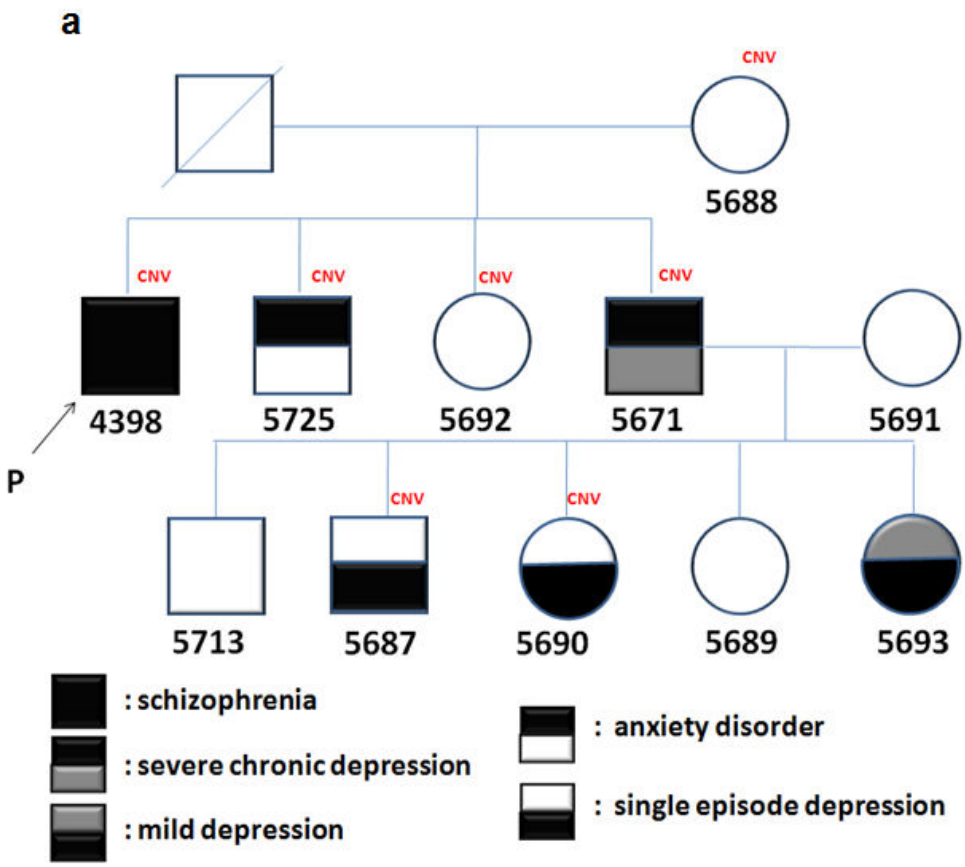


Figure 3.5 Array CGH detection of the 7p12 deletion for Patient 4398. Custom oligo array CGH was performed on 4398, the proband from Family 340 (against reference HapMap NA10851) **a)** Log2ratio profile at chr7:480,400,00-486,800,00. A deletion was detected at ~485.5 Mb on chromosome 7 (vertical bars in brown depicts the genomic locations of the ABCA13 exons). **b)** Detailed view of the deletion, located at ~48,552-48,563 kb, between exons 53 & 54 of *ABCA13*. The deletion size is ~11 kb.



(Figure 3.6 to be continued)

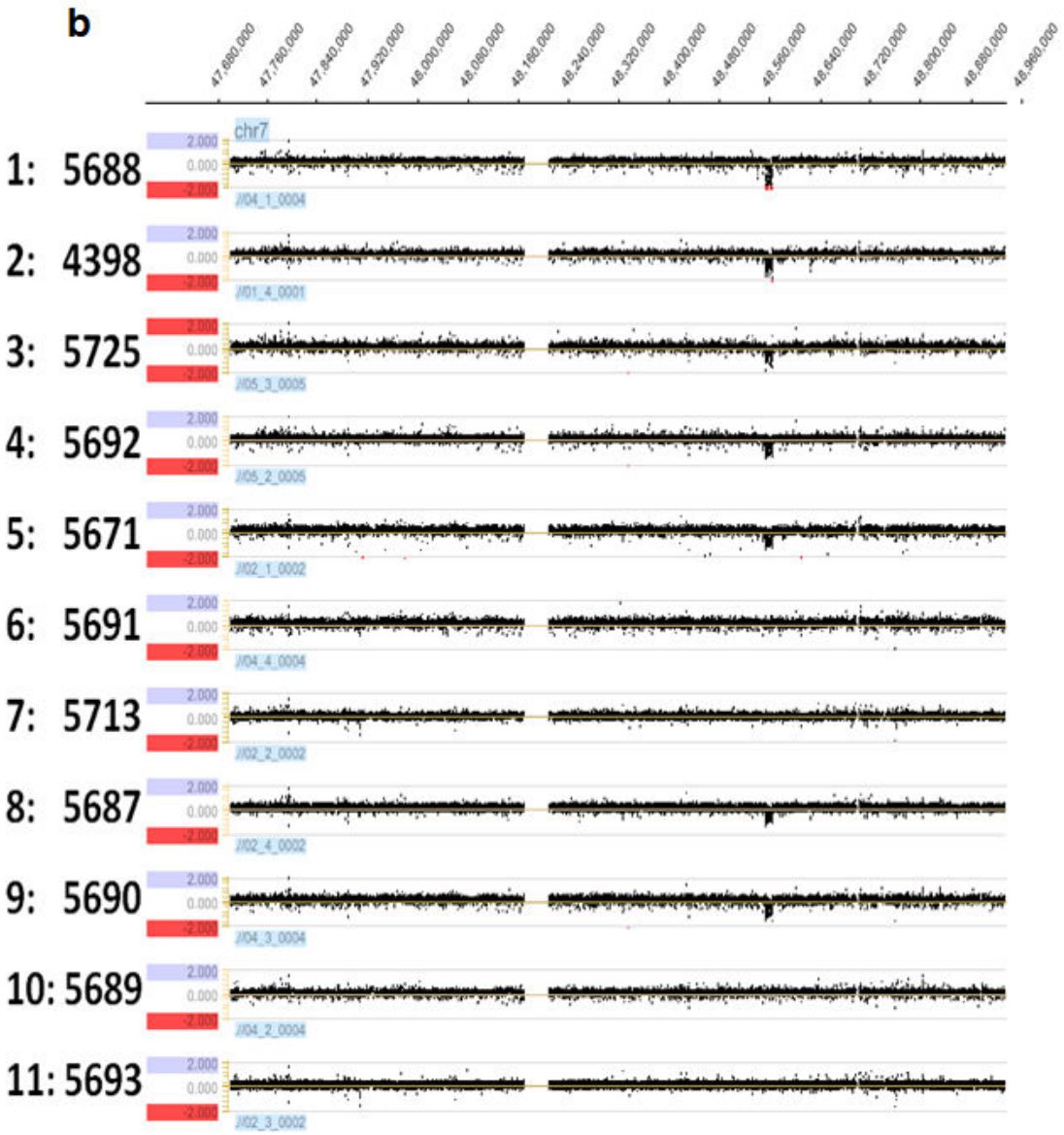


Figure 3.6 Array CGH profiles for all available members in Family 340. a) Pedigree of family 340. Individuals detected with the *ABCA13* deletion are marked "CNV" (in red). **b)** Oligo array CGH was performed on both affected and non-affected members of the family (against reference HapMap NA10851). *ABCA13* genomic region at chr7:47,680,000-48,960,000. A deletion at ~485.5 Mb was detected in samples 5688, 4398, 5725, 5692, 5671, 5687 and 5690.

3.3 Chapter Summary and Discussion

In this chapter, we described a CNV investigation in familial cases of schizophrenia and related neuropsychiatric conditions. As a first approach, we applied a whole-genome WGTP array CGH platform for 3 families. We detected a candidate CNV at 1p36 disrupting the gene *H6PD* and *SPSB1* in all affected members within F-29, but not among 270 normal HapMap individuals. The variant was further characterized by Fiber-FISH, long range PCR and breakpoint sequencing.

The second approach was focused on one candidate gene - *ABCA13* - on chromosome 7p12, a region previously linked to schizophrenia by cytogenetic studies (personal communication, Prof. Douglas Blackwood). With a high-resolution custom array we screened for cryptic CNVs not detected previously by other cytogenetic techniques. A ~11 kb deletion in intron 53-54 of the *ABCA13* was detected in one of the 5 families. The deletion segregated with the disease in all affected members except an individual with mild depression (5693), who could be explained as a phenocopy. On the other hand, a normal sibling (5692) was detected to inherit the deletion.

The inheritance of *ABCA13* CNV revealed the complexity in studying schizophrenia and related psychiatric disorders. Genetic heterogeneity and the complexity of the phenotype, characteristics of many human behavioural traits including schizophrenia, have challenged the study of psychiatric genetics for decades (Cantor and Geschwind 2008). In our study, we faced complicated genotype-phenotype correlation due to a number of reasons: First diagnostic criteria could only serve as guidelines in the complicated spectrum of psychiatric diseases. Overlapping clinical phenotypes in individuals exist, blurring the boundaries of normal to psychiatric patients of variable degrees. An individual in F-340, for example, was classified normal but in fact had single episode of

depression. Secondly, disease-causing variants may have incomplete penetrance and variable expressivity. This will lead to promising but incomplete segregation, as seen in our study and in several cases of well-established structural variations. The schizophrenia candidate gene *DISC1*, for example, was originally identified as an almost but not complete segregating translocation, in an extended Scottish pedigree with schizophrenia and multiple psychiatric conditions (St Clair et al. 1990; Blackwood et al. 2001).

Despite these complications, precedent studies focusing on causative factors in rare, inherited forms of the disease have shown this to be an effective strategy. *DISC1* was one such example. Another relevant example comes from the history of Alzheimer's disease genetics. Rare familial studies of Alzheimer's disease in the past were instrumental in the identification of the amyloid precursor protein (*APP*). These studies in turn provided important insights for investigation of the more common, sporadic form of disease. Furthermore, recent CNV studies in Alzheimer's, again focusing on autosomal dominant forms, identified the duplication of the *APP* locus as a predisposition factor in 8% of familial cases screened (Delabar et al. 1987; Rovelet-Lecrux et al. 2006; Sleegers et al. 2006).

Analogous to our familial CNV study in schizophrenia, the above approaches demonstrates how inherited forms of diseases could simplify the study of genetically complex neurological disorders. Studying schizophrenia as Mendelian disease trait has potentially two important advantages. First, schizophrenia probably comprise of some rare, virtually unique mutations (e.g. CNVs) as disease-causing variants (Abrahams and Geschwind 2008), only identifiable through extremely large case-control cohorts in population-based association. To circumvent the requirement for such large sample

sizes, looking for disease transmission in inherited forms could be more efficient and cost-effective.

Secondly, CNVs identified in rare familial cases could generate further candidates of disease association. CNVs could disrupt genes, for example, while the same gene might harbor more common variants with modest disease risk, and could be identified through targeted sequencing efforts. In addition, defining interaction partners and biological pathways of CNV gene candidates, one may discover novel disease candidates. For instance, *PDE4B* (Phosphodiesterase 4B) (Millar) and *NUDEL* (NudE-like) (Ozeki et al. 2003), both interacting factors of *DISC1*, were identified as schizophrenia candidates following the discovery of *DISC1* translocation.

Nevertheless, caution has to be taken in familial studies of such a complex psychiatric disease. First, genetic factors underlying schizophrenia and other psychiatric conditions may exhibit pleiotropic effects, with some CNVs demonstrating incomplete penetrance or variable expressivity. This will complicate segregation analysis in families, as in our case of *ABCA13* deletion. Secondly, our understanding of CNVs is still rudimentary. Evolving knowledge and views on this type of variant is likely to influence clinical decision on what is benign and what is causative. The 1p36 duplication we described, for example, suggested that some genomic loci may be under-ascertained for CNVs in the past or at present. Variants defined as “rare” or “unique” may in fact occur in the general population, although screening in large sample sizes is required for detection. This also alludes to the need for public resources and database for improved CNV entries, which will be discussed in Chapter 7.

The familial studies described in this chapter are limited in scale. Future work on larger-scale family-based schizophrenia studies, based on more pedigrees (as described for

the Alzheimer's study), or with larger kindred (such as the one used to identify the *DISC1* translocation) will be highly informative. Alternatively, families or trios may also be used to address the *de novo* nature of CNVs in schizophrenia. All these will be complemented with population-based association studies, another approach we employed which will be described in the next Chapter.

Chapter 4
Population-Based CNV Study in Schizophrenia

4.1. Experimental Design and Array Data Quality Control

4.1.1 Case-Control CNV Screen Experimental Design

In this chapter we will present a population based case-control CNV study design to detect schizophrenia-related CNVs. The CNV screen was mainly based on a core CNV dataset generated by hybridizations of 91 schizophrenia samples (SCZ) and 92 ethnically matched Lothian Birth Control Cohort samples (LBC) on WGTP arrays. The CNV data was then analysed for rare copy number variants specific to the SCZ cohort, and frequent copy number variants for disease association study.

A subset of the WGTP CNV data was validated by independent platforms, including quantitative PCR and high resolution oligonucleotide array (Nimblegen, Inc.). Furthermore, 39 patients in the SCZ cohort were hybridized on the Affymetrix SNP array 6.0 platform, and CNV data was included for validation of the WGTP dataset (source: University of Edinburgh, as part of International Schizophrenia Consortium).

WGTP array CGH data was analyzed using the algorithm CNVFinder (Fiegler et al. 2006) (discuss in Chapter 2). Affymetrix SNP array data was analyzed using the Birdseye package (Korn et al. 2008) and PLINK (Purcell et al. 2007). The ascertainment scheme for the Affymetrix CNV data emphasized rare, large CNV events and therefore represents a more conservative calling algorithm than the WGTP data. The SNP platform also provided higher-resolution data and facilitated the mapping of CNV breakpoints with higher precision.

Affymetrix SNP array data on a further 206 SCZ samples (from the same DNA source as our initial cohort) was used to investigate recurrent variants in SCZ which are more likely to be disease-relevant.

Figure 4.1 shows a summary of the case-control DNA samples and the CNV detection platforms used in the current study.

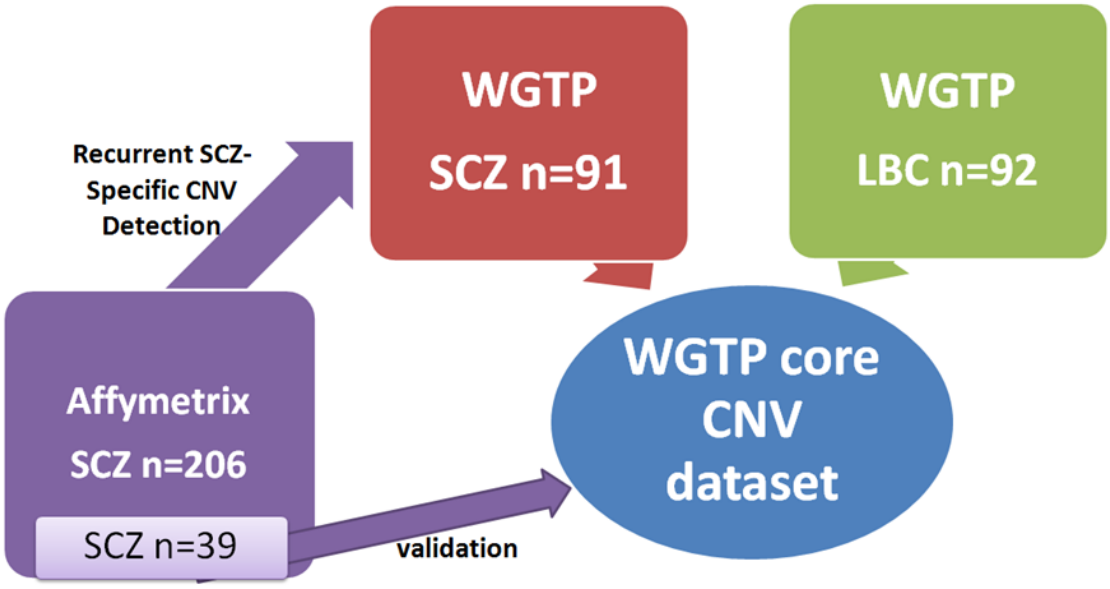


Figure 4.1 Case and control cohorts and CNV detection platforms used in our study

4.1.2 WGTP Array Data Quality Control

BAC array based comparative genomic hybridization has been routinely used for the detection of copy number variations in clinical samples. Normalization and filtering of the array data is critical to enhance the discovery of true biological signals, and to remove experimental artefacts for accurate and reliable results. We performed initial normalization on the WGTP data as previously described (and as documented in section 2.1) (Fiegler et al. 2006). Additional normalization and filtering steps, including correction for GC content and filtering clones of artefacts, were summarized in Figure 4.2, with details included in Appendix C.

In addition to manual inspection of profile qualities, all hybridization results were monitored with two statistical indicators for quality control. The first is global SDe, an estimate of standard deviation of all log2ratio signals in the genome for a particular sample profile. The second is the clone retention rate after fusion of dye swap experimental data. Experiments were accepted for further analysis only if global SDe<0.06, global clone retention rate >90% and clone exclusion rate per individual chromosome <80%. Figure 4.3 presents the quality control indicators for all SCZ and LBC hybridizations included in the CNV analysis.

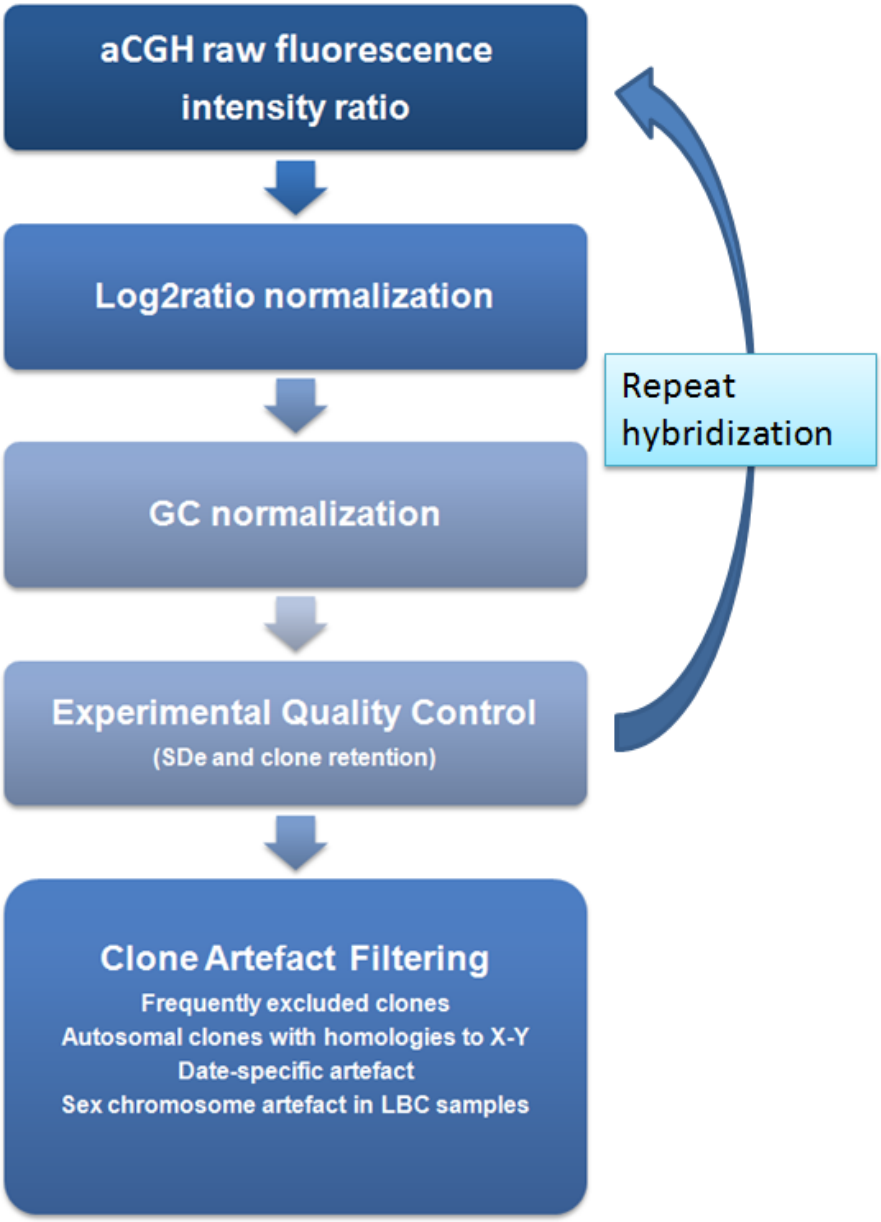


Figure 4.2 Normalization and filtering steps applied to WGTP hybridization data before CNV analysis.

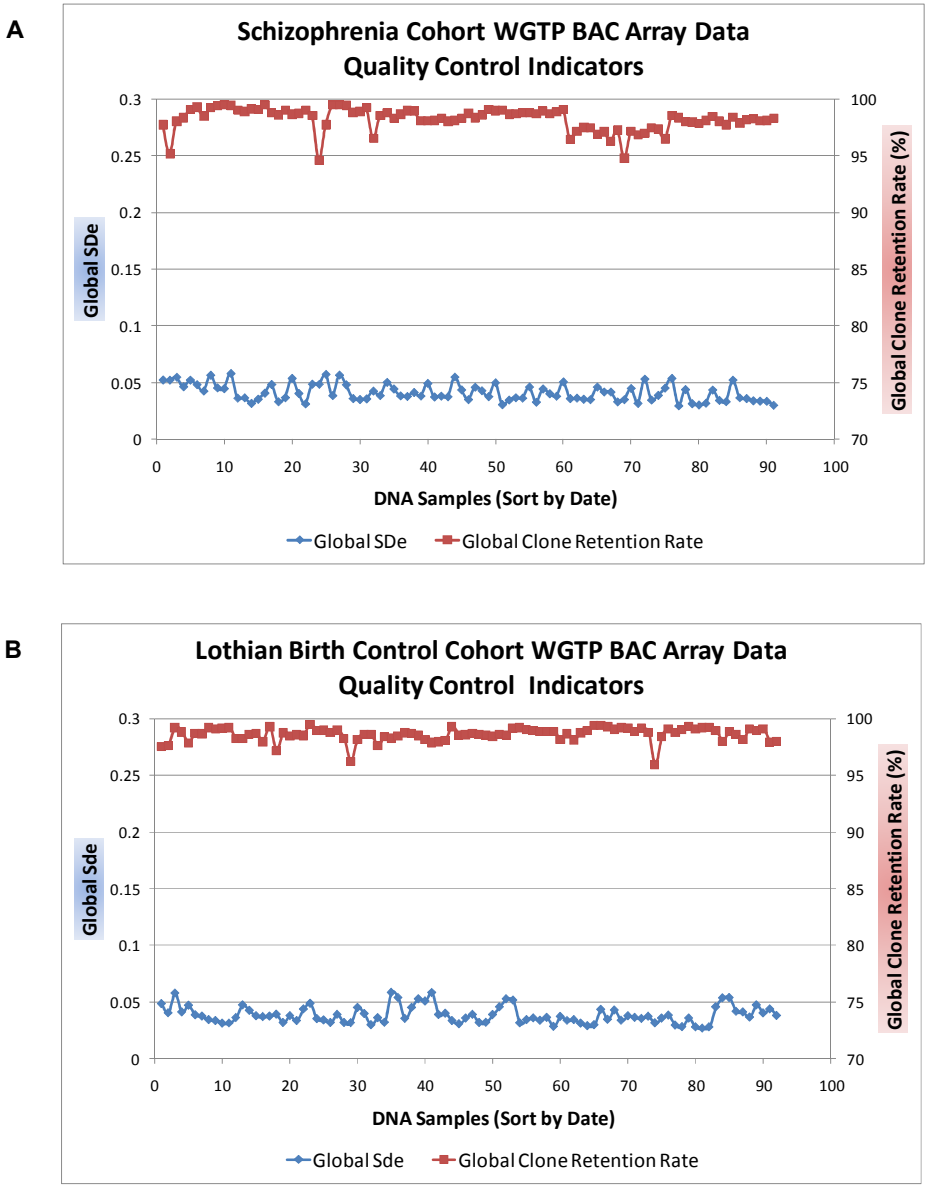


Figure 4.3 WGTP array quality control indicators. Global SDe (blue) and global clone retention rate (red) are plotted against DNA samples (sorted by date) for all samples included in analysis. **a)** SCZ (n=91); **b)** LBC (n=92)
(Global SDe cut-off point was <0.06. Global clone retention rate cut-off point was >90%)

4.2 Copy Number Variation Detection on the WGTP array

4.2.1 Distribution of CNVs and CNV Regions in Case and Control Cohorts

The core copy number variation dataset in the study was generated from DNA samples hybridizing against the WGTP array platform. On average we detected 66 CNVs per SCZ sample (case) and 72 CNVs per LBC sample (control). A list of per-patient CNV information (number and types of CNVs) is given in Appendix D-1.

The total number of CNV calls was 6,036 in the cases (3,650 gains, 2,386 losses) and 6,666 in the controls (3,973 gains 2,689 losses) (Figure 4.4a). CNV size distributions were similar in the two cohorts (Figure 4.5), with the same median length of 237.7 kb. CNV sizes were estimated from the chromosomal coordinates of affected BAC clones. Since average BAC clone size was 164 kb, the reported CNV size may be an overestimation of the true CNV coordinates (McCarroll et al. 2008; Perry et al. 2008).

In each sample cohort, individual CNVs from for multiple individuals may overlap at the same genomic location. To generate sets of non-overlapping CNV genomic locations, we grouped for each cohort all overlapping CNVs into copy number variation regions (CNVRs). The total number of CNVRs was 539 in SCZ (249 gain only; 159 loss only; 131 gain or loss) and 583 in LBC (286 gain only, 127 loss only, 170 gain or loss) (Figure 4.4 b&c).

CNVRs were classified as either “rare” (occurring in only one sample) or “recurrent” (occurring in multiple samples) with respect to the cohort. There were 224 rare and 315 recurrent CNVs in the SCZ cohort, compared with 237 rare and 346 recurrent ones in LBC. The proportion of rare and recurrent CNVs in the cases and controls (Figure 4.4b) showed no statistically significant difference. Distributions of gene content were also

similar for cases and controls (Figure 4.6), with majority of CNVRs contain one or more genes in both SCZ (355 out of 539 CNVRs; 66%) and LBC (390 out of 583; 67%).

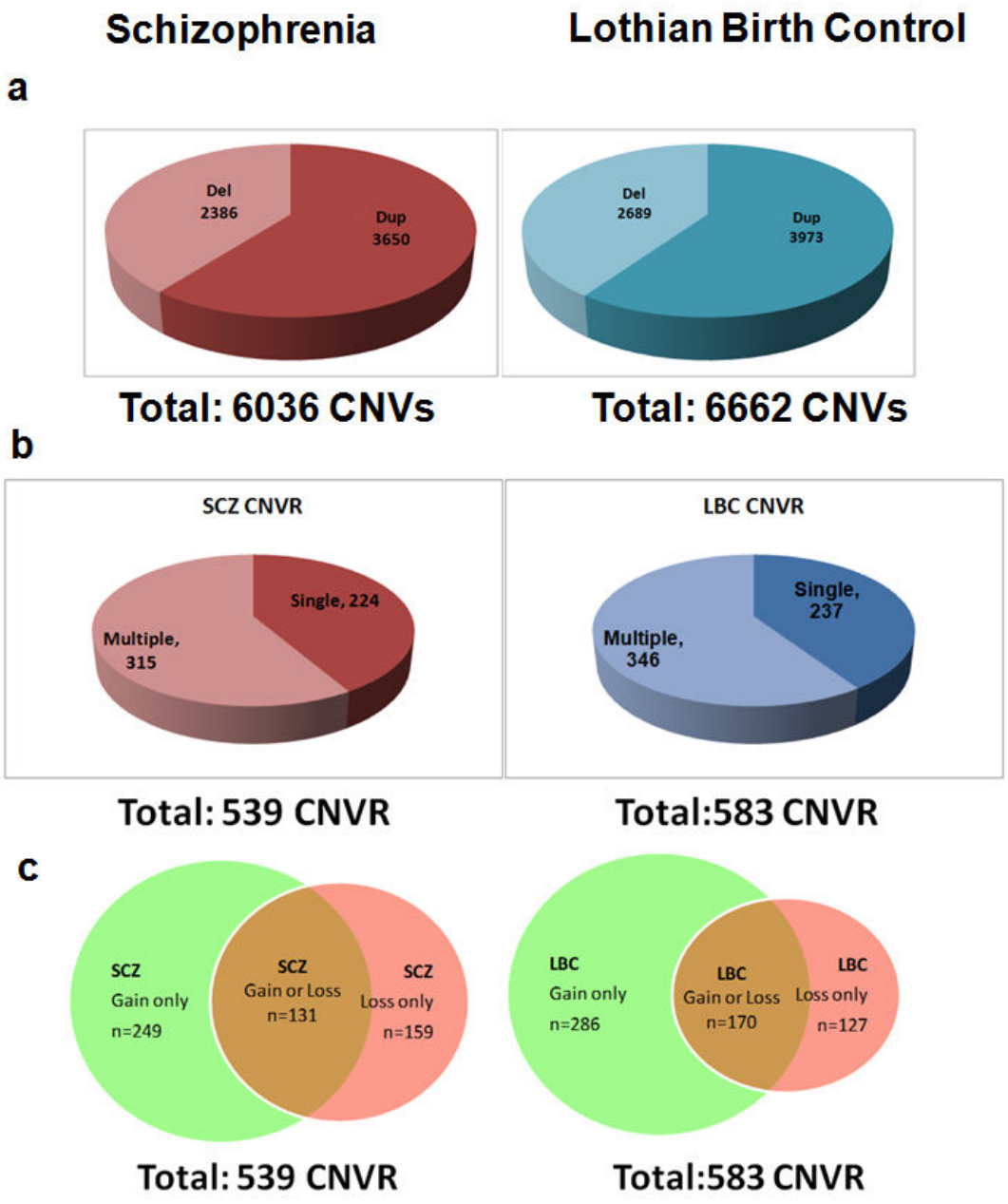


Figure 4.4: Frequency and types of CNVs and CNVRs in SCZ and LBC. a) Number of CNVs (gains/losses) in SCZ and LBC **b)** Number of CNVRs (rare/recurrent) in SCZ and LBC **c)** Number of CNVRs (gains/losses) in SCZ and LBC

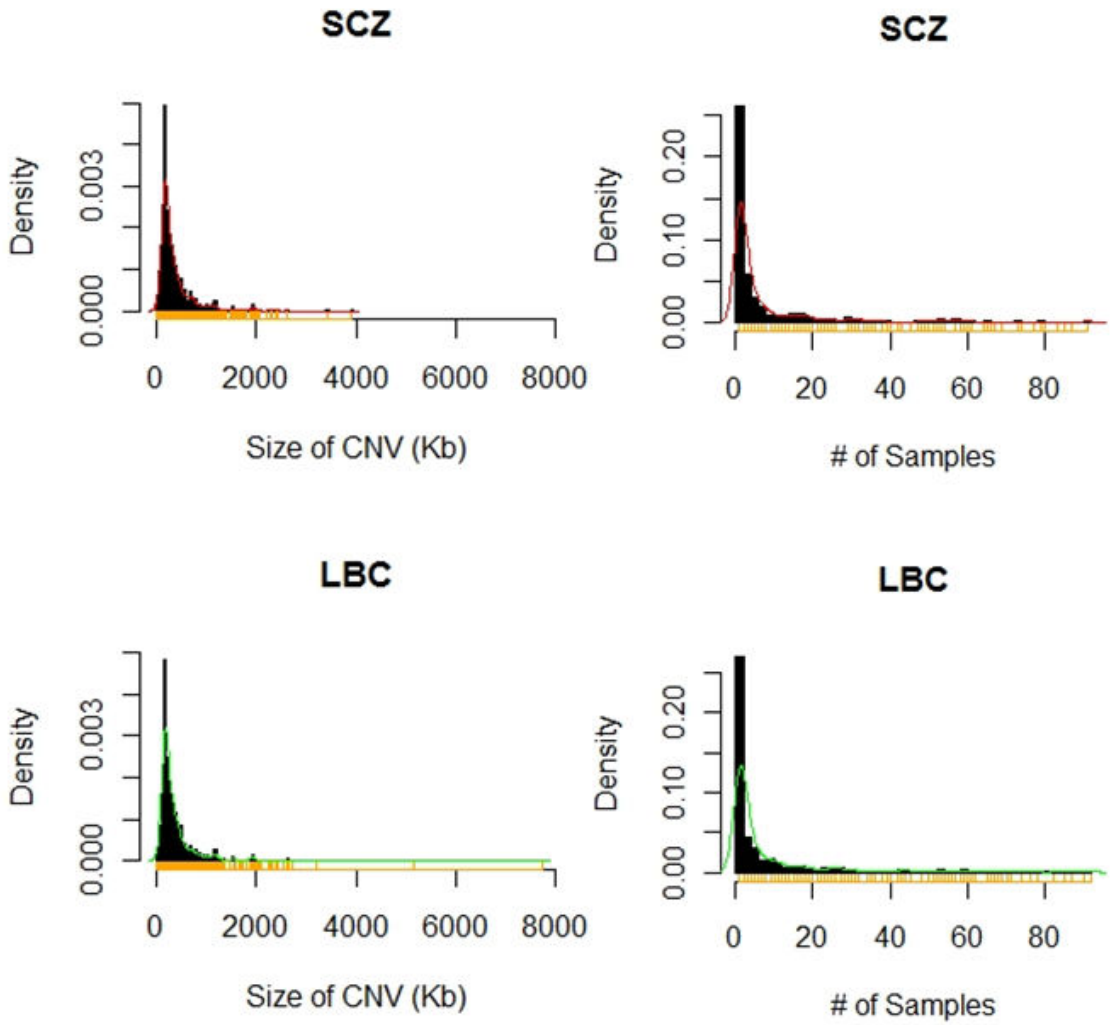


Figure 4.5 Size and frequency distributions of CNVRs in Schizophrenia (SCZ) and Lothian Birth Control (LBC) Cohorts detected using the WGTP platform.
***Top left*:** CNV size distribution in SCZ
***Top right*:** CNVR frequency distribution in SCZ
***Bottom left*:** CNV size distribution in LBC
***Bottom right*:** CNVR frequency distribution in LBC

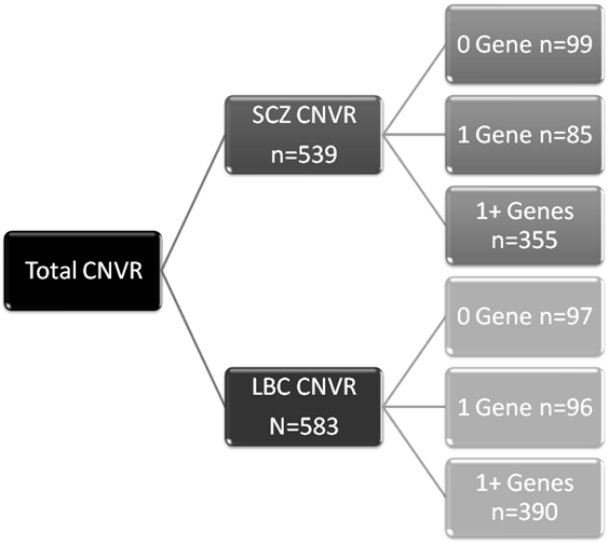


Figure 4.6 CNVR gene content in the SCZ and LBC cohorts.

4.2.2 Bias of CNV Discovery Rate in SCZ Versus LBC

Although CNVs/CNVRs in our cases and control data showed comparable distributions in terms of CNV size, frequency and gene contents, we detected more CNVs/CNVRs in the control cohort than in patients with schizophrenia (10.4% more CNVs in total, 9.10% more CNVs per sample). This bias in CNV discovery rate was mainly due to differences in DNA quality, leading to variability in the quality of array CGH profiles in the two cohorts (with control DNA samples of better quality on average). Lower DNA quality increases the fluorescence intensity log2ratio variance (SDe) and consequently decreases the sensitivity of CNV detection by CNVFinder.

Figure 4.7 shows a negative correlation between number of CNVs detected and estimated standard deviation SDe. The regression (logarithmic) lines fitted on the SCZ and LBC correlation data are comparable (Fig 4.7a). However, because the SDe distribution for LBC samples shows a slight shift towards lower values (mean SDe in:

SCZ = 0.41; LBC= 0.038) (Fig 4.7b), we detected on average a higher CNV number in the control cohort.

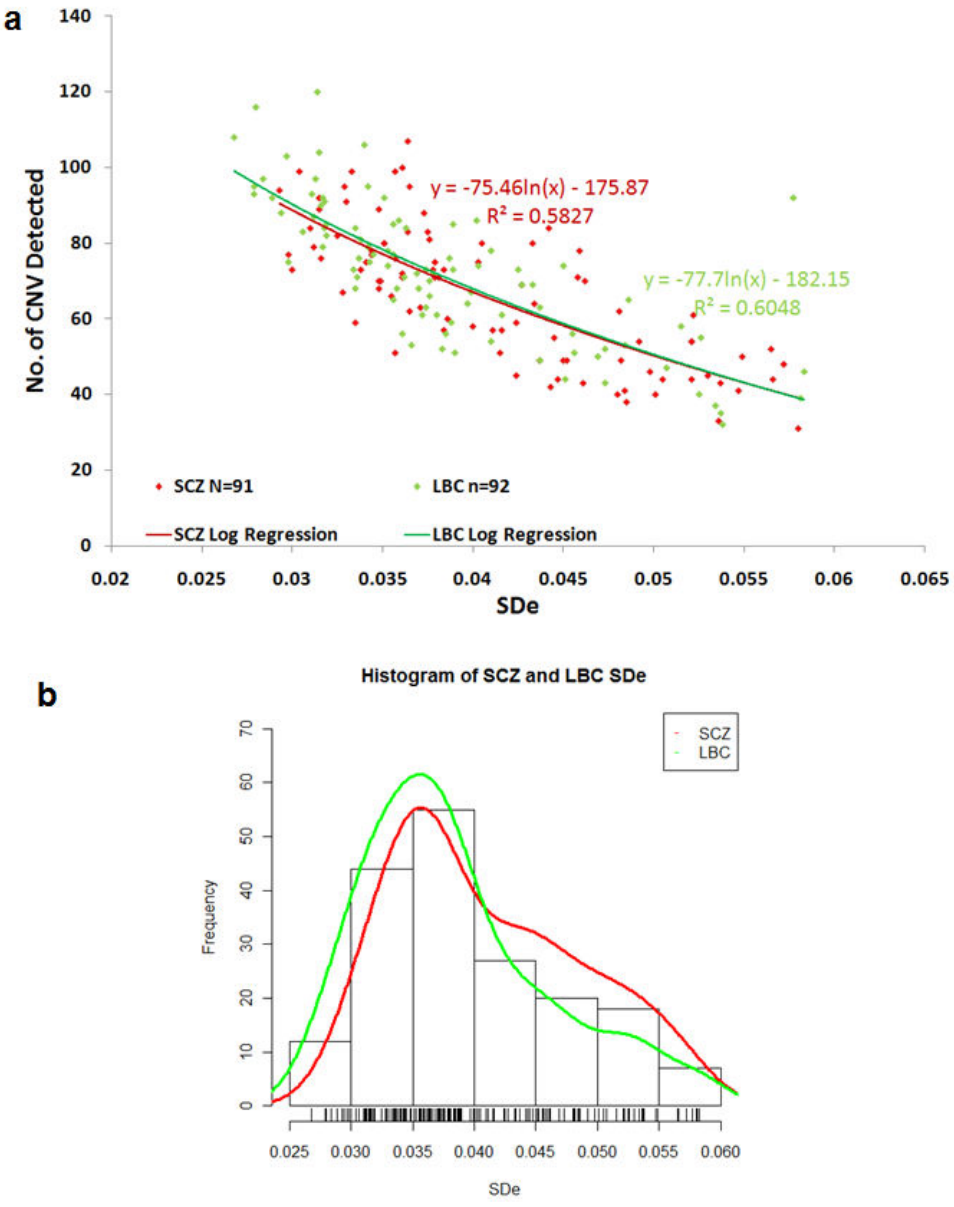


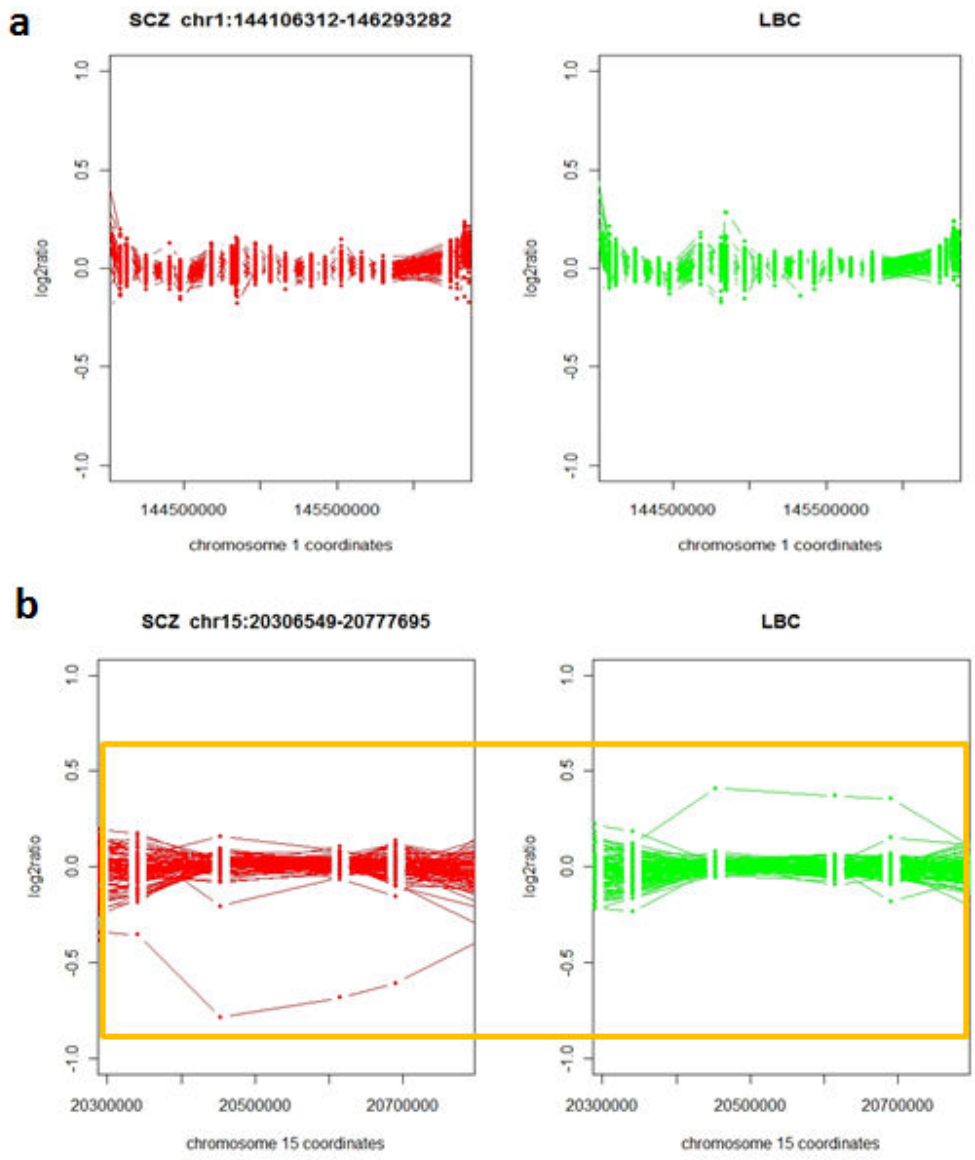
Figure 4.7 Correlation of CNV discovery rate with data quality in SCZ and LBC. a) Number of CNVs detected in each sample against profile variability (SDe). Regression fit showed negative correlation for both SCZ (red) and LBC (green) data. **b)** SDe distribution for SCZ and LBC experiments. Mean SDe is 0.041 in SCZ and 0.038 in LBC, indicating an overall higher quality of the LBC data.

4.3 Comparing WGTP Data with Known Schizophrenia CNV Regions

Novel findings of large-scale genome-wide schizophrenia CNV screen (ISC 2008; Stefansson et al. 2008) revealed 3 recurrent CNV schizophrenia loci at 1q21.1, 15q11.2 and 15q13.3. These 3 recurrent deletions, together with the previously identified 22q11 deletion, represented 4 CNV regions that demonstrated statistical significant association with schizophrenia. We therefore screened the WGTP CNV data (SCZ and LBC) against these 4 genomic loci.

Figure 4.8 illustrates the SCZ (red) and LBC (green) WGTP data at these 4 loci. Of note, we detected the same recurrent deletion at 15q11.2 as Steffanson *et al.* (Stefansson et al. 2008), as shown in one sample in the SCZ cohort. The region was also duplicated in one LBC sample (Figure 4.8b). Given the estimated frequency of 0.55% and 0.19% in cases and controls (Stefansson et al. 2008), we expected the deletion to occur in 0 or 1 sample in our cases, whilst being absent in the controls, which was consistent with our finding. This also confirmed the ability of our platform to detect disease-associated rearrangement of similar size and frequency. As described in section 1.6, the recurrent deletion was flanked by LCRs (Figure 4.9). Within the deletion was a candidate gene *CYFIP1* (cytoplasmic FMR1 interacting protein 1 isoform), which encoded a protein involved in the regulation of translation in neurons, with important functions in synaptic plasticity and brain development (Napoli et al. 2008) (see section 1.6).

At the other three known regions, 1q21.1, 15q13.3 and 22q11, no large deletion was identified in our SCZ cohort. Smaller reciprocal duplications were detected at 1q21 and 22q11, although these have been reported in control cohorts at equal or higher frequency than in corresponding case cohorts (ISC, 2008), and therefore are likely to be benign.



(Figure 4.8 to be continued)

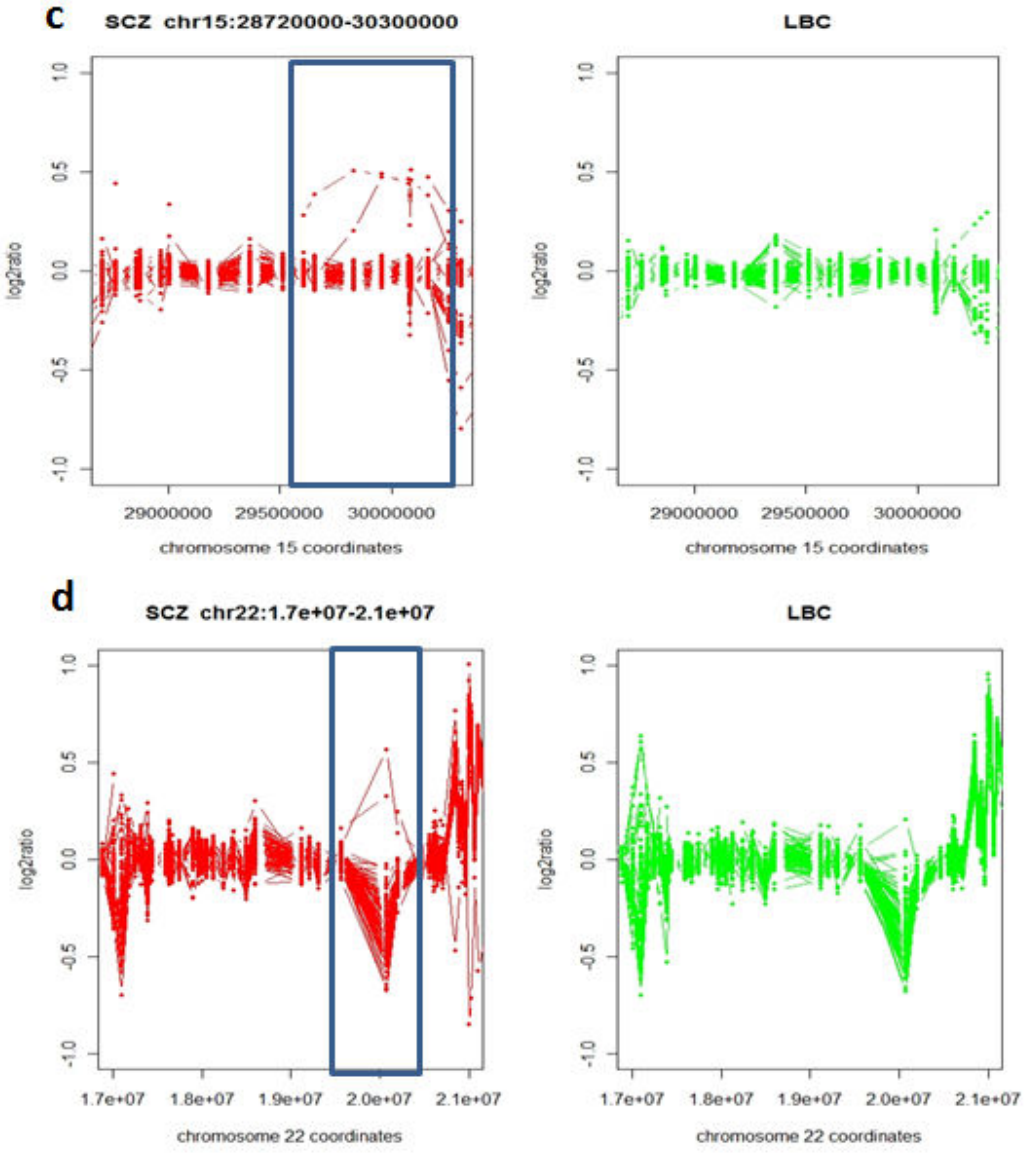


Figure 4.8 WGTP data compared with 4 known schizophrenia CNV loci. SCZ data was presented in red and LBC in green **a)** 1q21.1: no large rearrangement detected **b)** 15q11.2: a large deletion was detected in SCZ, and a reciprocal duplication in LBC (orange box) **c)** 15q13.3: two smaller duplications were detected in SCZ (blue box); **d)** 22q11: two smaller duplications were detected in SCZ (blue box)

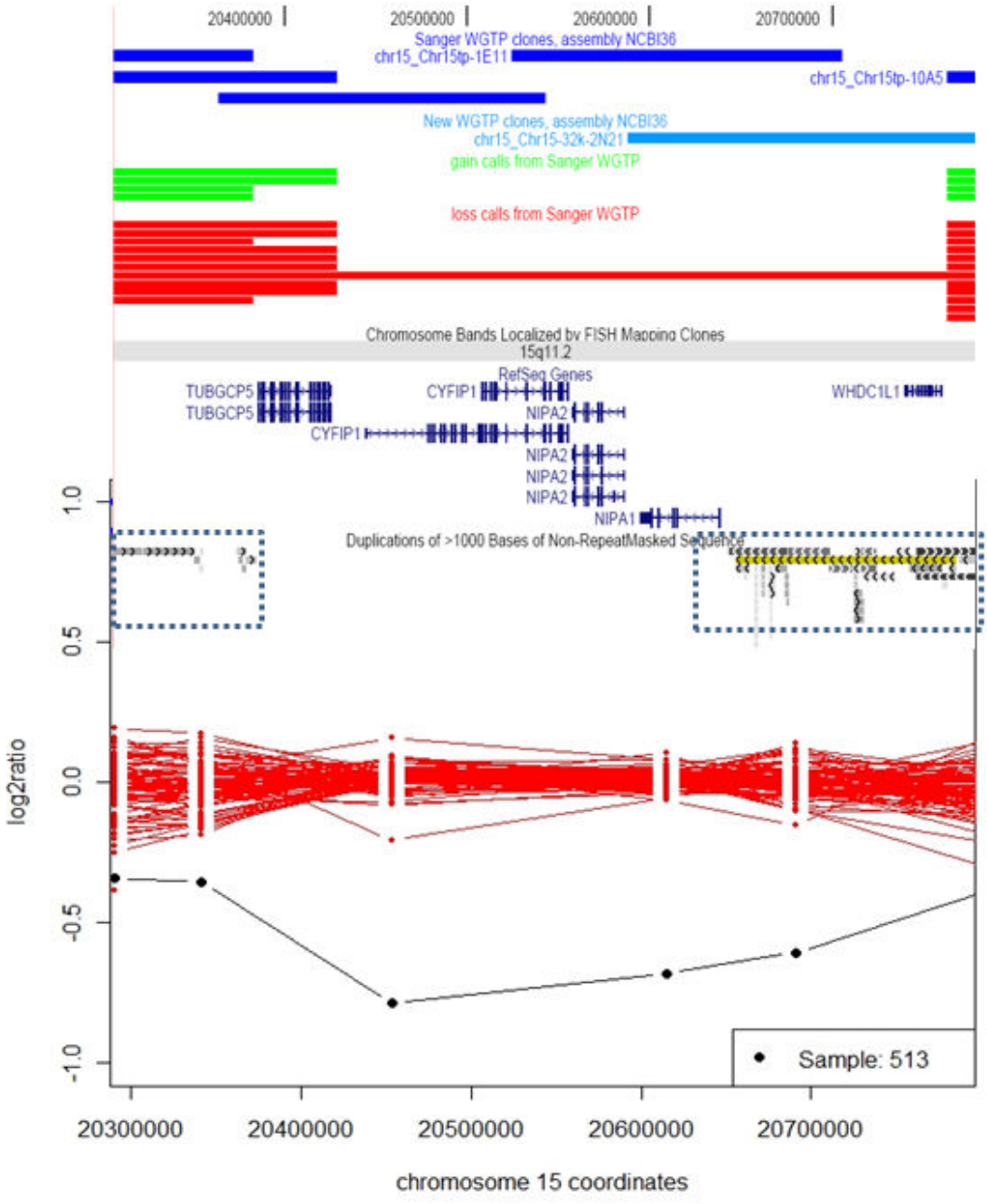


Figure 4.9 WGTP data detected a deletion at 15q11.2 in one patient. The deletion spans gene *CYFIP1*. Regions of low copy repeats (LCR) are highlighted in blue boxes.

4.4 Rare Variants Specific to the Schizophrenia Cohort

With evidence that rare variants significantly associated with schizophrenia were replicated in our WGTP dataset, we concentrated our first approach on rare CNVs that were only detected in schizophrenia samples. This approach was also reinforced by recent CNV findings detecting a higher occurrence of rare, and in some cases *de novo* CNVs in cases compared to controls (ISC 2008; Walsh et al. 2008; Xu et al. 2008). The International Schizophrenia Consortium, for instance, reported a 1.448-fold increase of extremely rare CNVs (singletons in 3000+ cases) in cases versus control. Another report from Walsh *et al.*, suggested the increase of rare genic CNVs was as high as 3-fold in cases versus control. Extrapolating from these data, one could hypothesize that 1/3 to 2/3 of the rare case-specific CNVs would be disease-causing variants.

4.4.1 Rare Variant Detection Using Consecutive Clone Calling Criteria

In our rare variant analysis we included only CNVs involving more than one consecutive clone. This is a common practise in CNV studies to increase data stringency. The corresponding CNVs[2] were grouped into 324 CNVRs in cases and 360 CNVRs in controls.

Less than 1/3 of the CNVRs were cohort-specific (113 in cases and 141 in controls), indicating that majority of variants detected were recurrent copy number polymorphisms in the Scottish population (existing in both Scottish cases and controls).

The 113 SCZ cohort-specific CNVs are listed in Appendix D-2.

[2] This set of CNVs includes a total of 3,980 CNVs in the cases and 4,444 CNVs in the controls.

4.4.2 Validation of SCZ-Specific Rare Variants

A reliable detection platform is crucial in CNV discovery. The WGTP array CGH platform available to our study has previously been tested extensively in terms of accuracy and sensitivity for CNV discovery (Fiegler et al. 2006). We performed additional validation experiments specific to our dataset through qPCR, Nimblegen oligonucleotide array and with Affymetrix SNP array data provided by the ISC.

Validations were performed for two purposes: 1) to confirm the presence of a subset of CNV calls as a measure of platform detection accuracy; and 2) to delineate CNV boundaries and assess the involvement of candidate genes. 29 regions were validated by a combination of these methods (Table 4.1). Details of validation experiments and validation rates were provided in Appendix E.

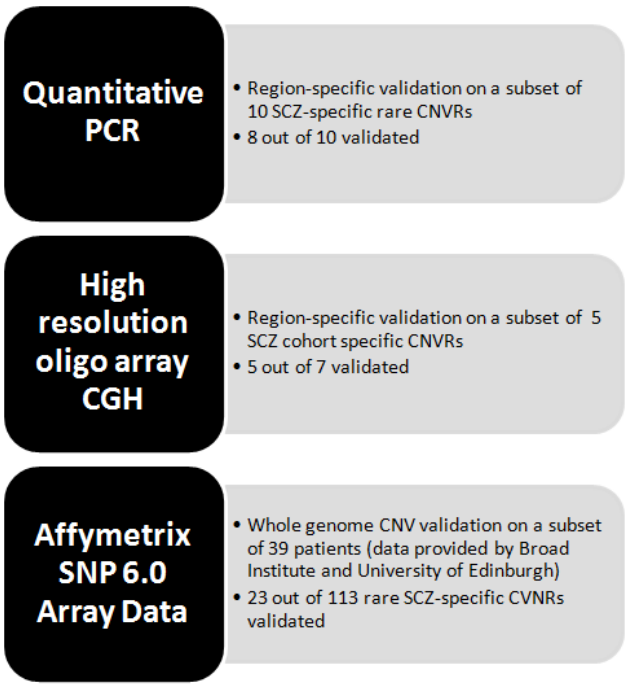


Figure 4.10 WGTP CNV dataset validation strategies.

Table 4.1 Validated SCZ-specific rare variants. 29 regions were validated by a combination of qPCR, Nimblegen oligonucleotide array and Affymetrix SNP array. (***Red***: *DNA samples used in validation.* ***Black***: *genes deleted/duplicated as refined by higher resolution CNV detection platform*)

WGTP Chr coordinates	Genes	Refined Chr Coordinates by higher resolution array	Samples	Validation			CNV Type	
				qPCR	Nimb legen	Affy	gain/ loss	rare/ recurrent
chr5:110212743-110577198	WDR36, TSLP	N/A	**850**	√			loss	rare
chr6:22212251-22607763	PRL	N/A	**5324**	√			loss	rare
chr7:93658283-94064936	SGCE, CASD1, COL1A2	N/A	**4203**	√			loss	rare
chr8:9670799-9988807	TNKS, MSRA	chr8:9830000-9870000	**3789**	√	√		loss	rare
chr9:137112070-137464520	C9orf62, OLFM1	chr9:137280000-137450000	**323**, 4179, 5758	√	√		gain	recurrent
chr3:86998287-87362355	**VGLL3**, CHMP2B	chr3:86898287-87462355	**5386**		√		gain	rare
chr1:68325265-68897278	**DEPDC1,GPR177,RPE65**	chr1:68334011-68755246	**3766**			√	gain	rare
chr1:71346767-71893785	**NEGR1**	chr1:71379226-71687039	**3766**	√		√	gain	rare
chr2:127333465-127814482	**BIN1,CYP27C1**,ERCC3,MAP3K2	chr2:127182184-127688195	**3975**			√	gain	rare
chr3:143326111-143612526	**GK5,TFDP2,XRN1**	chr3:143303643-143554513	**4710**			√	gain	rare
chr3:189546564-189761158	**LPP**	chr3:189564871-189682182	**3945**, 7183			√	loss	recurrent
chr4:11668861-11990799	N/A	chr4:11838410-11948682	**4748**			√	loss	rare
chr4:28258973-28570510	N/A	chr4:28415433-28541944	**3945**			√	loss	rare
chr5:25051483-25333181	N/A	chr5:25121394-25405536	**1295**			√	gain	rare
chr6:1994540-2304998	**GMDS**	chr6:2001510-2293508	**3815**			√	gain	rare
chr6:97469364-98475130	**C6orf167,KLHL32**	chr6:97426034-98410421	**1085**	√		√	loss	rare
chr6:165840858-166673790	**PDE10A,PRR18,SFT2D1,T**	chr6:165893867-166666551	**5758**			√	gain	rare
chr7:17795285-18873121	**HDAC9,SNX13**	chr7:17674099-18634682	**5541**	√		√	loss	rare
chr7:88344723-88962788	**ZNF804B**	chr7:88292092-88731204	**1295**			√	loss	rare
chr8:13581606-13896999		chr8:13684044-13829803	**5541**			√	loss	rare
chr11:31339430-31549584	**DCDC1,DPH4,ELP4,IMMP1L**	chr11:31312028-31554033	**1278**			√	gain	rare
chr12:17761729-18154145	**RERGL**	chr12:17776260-18038013	**3409**	√		√	loss	rare
chr13:112568274-112939870	**ATP11A,F1,F7,MCF2L,PCID2,PROZ**, CUL4A	chr13:112541419-112904310	**385**			√	gain	rare
chr13:113166604-113658050	GRTP1,ADPRHL1,**ATP4B,DCUN1D2,TFDP1,TM CO3**,FAM7B,GAS6,GRK1	chr13:113051168-113366491	**6638**			√	loss	rare
chr14:74945401-75396561	JDP2,**BATF,C14orf1,FLVCR2,TTLL5**	chr14:75046923-75379455	**4100**			√	gain	rare
chr15:99550338-100036184	CHSY1,,SNRPA1,**PCSK6**,TARSL2,TM2D3	chr15:99764132-99956872	**6638**, 385, 7294			√	gain	recurrent
chr20:14561834-14936485	**MACROD2**	chr20:14650902-14813485	**4716**			√	loss	rare
chr21:34576661-34852873	**C21orf51,KCNE1,KCNE2,RCAN1**	chr21:34644865-34827865	**899**			√	gain	rare
chr21:44384864-44825939	AIRE,C21orf2,C21orf33,DNMT3L,ICOSLG,PFKL, TRPM2,LRRC3,**C21orf29,C21orf9,KRTAP1-1,KRTAP1-2,KRTAP1-3,KRTAP1-4,KRTAP1-5**,KRTAP1-6, KRTAP1-7, KRTAP1-8,KRTAP1-9,KRTAP1-10	chr21:44730300-44890569	**5307**			√	gain	rare

4.4.3 Rare Variants in SCZ with Genes Involved in Psychiatric Disorders

One way to assess whether a variant is likely to be disease-causing is to examine its gene content. A query of the Genetic Association Database (GAD) provided by the National Institutes of Health ascertains 8 SCZ cohort-specific CNVRs to contain genes represented in the PSYCH (psychiatric-related) category. These genes are *AHR* (aryl hydrocarbon receptor), *TRPM2* (transient receptor potential cation channel, subfamily m, member 2), *DMPK* (dystrophia myotonica-protein kinase), *PIK3C3* (phosphoinositide-3-kinase, class 3), *GABRG3* (gamma-aminobutyric acid a receptor, gamma 3), *OXTR* (oxytocin receptor), *SGCE* (sarcoglycan, epsilon) and *KIF2A* (kinesin heavy chain member 2) (Table 4.2).

In particular, *PIK3C3*, a gene encoding the phosphoinositide-3-kinase, was detected as duplicated in patient 3857 (ED1013). The PI3K kinase is responsible for receptor-mediated signal transduction and intracellular trafficking. Neuregulin-1, a putative schizophrenia susceptibility gene, was suggested to regulate cell adhesion through a PI3K dependent pathway (Kanakry et al. 2007). *PIK3C3* itself was a putative schizophrenia and bipolar disorder candidate, supported by a number of genetic association studies (Stopkova et al. 2004; Duan et al. 2005; Saito et al. 2005; Tang et al. 2008).

SGCE, a gene detected as deleted in patient 4203, encodes transmembrane components of the dystrophin-glycoprotein complex. *SGCE* is involved in Myoclonus-Dystonia (Marechal et al. 2003; Valente et al. 2005; Misbahuddin et al. 2007), a movement disorder characterized by myoclonic jerks, dystonia and a variety of psychiatric symptoms including anxiety, depression and obsessive-compulsive disorder (Doheny et al. 2002). Furthermore, *SGCE* knockout mice were shown to have anxiety,

depression and altered dopamine levels, all phenotypes associated with schizophrenia (Yokoi et al. 2006).

Oxytocin Receptor (*OXTR*), another gene identified as duplicated in the SCZ cohort, encodes a neurohypophyseal hormone important in brain development (Insel et al. 1999). The role of oxytocin was highlighted in mammalian social behaviours (Israel et al. 2008). Abnormalities of the oxytocin system have been implicated in several neuropsychiatric disorders, including schizophrenia, autism , obsessive-compulsive disorder and post-traumatic stress disorder (Marazziti and Catena Dell'osso 2008). Pharmacologically, oxytocin has been proposed as an antipsychotic drug (Caldwell et al. 2008). Oct -/- mice demonstrated social deficit (Winslow and Insel 2002). Furthermore, mouse models suggested that the oxytocin system is involved in the response to phencyclidine (Lee et al. 2005; Caldwell et al. 2008), a drug when administrated lead to schizophrenia-like behavioural symptoms including hallucination. Schizophrenia endophenotypes such as prepulse inhibition deficits were augmented in the Oct knockout mice (Caldwell et al. 2008).

Table 4.2 SCZ-specific rare variants with genes associated with psychiatric disorders. Genes within CNVRs were compared to the Disease Association Database returning 8 CNV regions with genes related to psychiatric disorders (disease class: PSYCH).

CNVR	GENE	GENE DESCRIPTION	LOCATION	CYT BAND	DISEASE CLASS: PSYCH
3815-gain-14	OXTR	oxytocin receptor	chr3: 8686094-8880344	3p25	attention deficit disorder conduct disorder oppositional defiant disorder, autism
5386-loss-27	KIF2A	kinesin heavy chain member 2	chr5: 61438895-61723574	5q12-q13	schizophrenia
5324-gain-28	AHR	aryl hydrocarbon receptor	chr7: 17320942-17609921	7p15	dementia
4203-loss-28	SGCE	sarcoglycan, epsilon	chr7: 93658283-94064936	7q21-q22	Tourette syndrome; obsessive compulsive disorder, dystonia
5660-gain-61	GABRG3	gamma-aminobutyric acid (gaba) a receptor, gamma 3	chr15: 25060215-25335769	15q12	alcohol dependence, autism
ED1013-gain-73	PIK3C3	phosphoinositide-3-kinase, class 3	chr18: 37767209-38078089	18q12.3	schizophrenia, bipolar disorder
4100-loss-63	DMPK	dystrophia myotonica-protein kinase	chr19: 50553125-51160225	19q13.3	myotonic dystrophy
5307-gain-54	TRPM2	transient receptor potential cation channel, subfamily m, member 2	chr21: 44384865-44825939	21q22.3	bipolar disorder

In the rare variants specific to the SCZ cohort, we also detected several CNVs contain brain- or neuronal-related genes not currently associate with the Genetic Association Database. These CNV loci overlapped with genes *NEGR1, PDE10A, HDAC9 and RCAN1* (Figure 4.11).

The neuronal growth factor, *NEGR1*, was duplicated in one patient. It is brain-expressed and was suggested to be involved in cell adhesion (Figure 4.11a). *PDE10A*, a phosphodiesterase involved in corticostriatal signalling, was duplicated in another patient. *PDE10A* plays a role in cognition, locomotion and behavioural phenotypes in mice (Hebb et al. 2008), and *PDE10* inhibitor has been studied as a novel drug treatment for the cognitive symptoms of schizophrenia (Menniti et al. 2007; Schmidt et al. 2008) (Figure 4.11b).

HDAC9, a histone deacetylase which couples neuron-induced electrical activity to muscle cells (Mejat et al. 2005), was disrupted at the deletion breakpoint of a patient (Figure 4.11c). *RCAN1*, a calcineurin inhibitor involved in calcium-mediated signalling, was disrupted at the duplication breakpoint of another patient (Figure 4.11d). These two variants were further demonstrated to be recurrent in an extended schizophrenia cohort and will be discussed in section 4.4.

Given our modest sample size, we were not able to determine a statistically valid association of any of these rare regions with disease. Rather, our experimental design was aimed at cataloguing SCZ-specific CNVs as a collective dataset with candidate genes of disease relevance. Based on previous studies proposing a large proportion of case-specific rare variants being involved in disease, and the evidence of biologically plausible candidates, this set of genes represents a set of good candidates for future

functional investigation. Accruing CNV data from the literature and collaborative efforts, our dataset can also be extended to confirm disease-causing recurrent CNV loci, as in the case of *CYFIP1* deletion at 15q11.2.

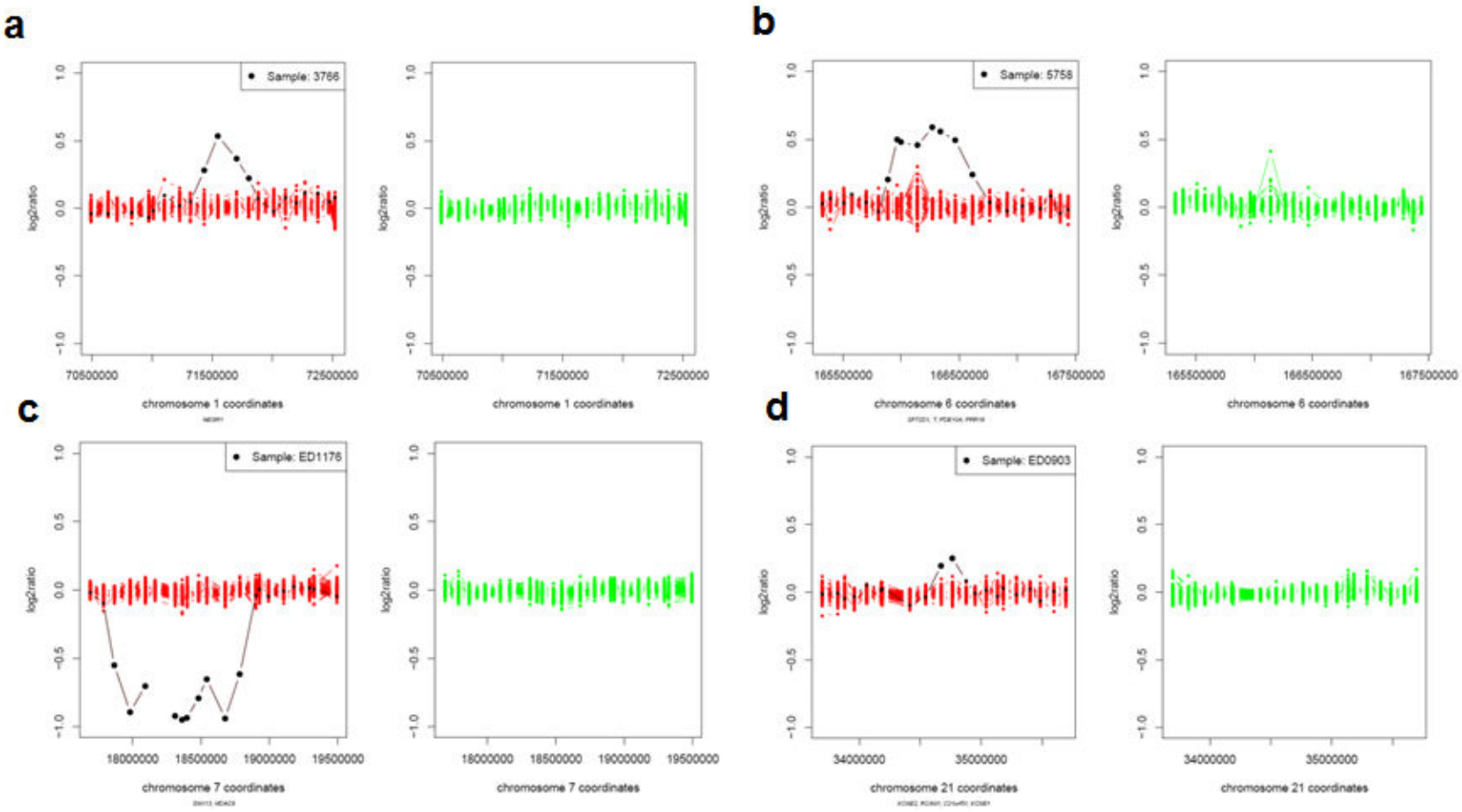


Figure 4.11 Schizophrenia cohort-specific CNVRs containing brain-related or neuronal-related genes. a) Duplication at *NEGR1* (Neuronal growth factor 1) **b)** Duplication at *PDE10A* (Phosphodiesterase 10A) **c)** Deletion disrupting *HDAC9* (histone deacetylase 9) **d)** Duplication spanning *DSCR1/RCAN1* (Down Syndrome Critical Region 1)

4.5 Recurrent SCZ-Specific Variants in Extended Cohort

Among the SCZ-specific variants detected by WGTP, the majority were singletons in the cohort of 91 SCZ samples (see last column in Table 4.1). To detect recurrent SCZ-specific CNVs, we extended our cohort using CNV calls made from the Affymetrix SNP 6.0 array as provided by ISC and University of Edinburgh. The dataset consists of an additional 206 Scottish schizophrenia samples hybridized on the Affymetrix array (SCZ_{Affy}n206). Combined with our original WGTP SCZ cohort of 91 (SCZ_{WGTP}n91), the extended cohort consists of 297 patient samples ($SCZ_{WGP\text{-}Affy}$n297).

A recurrent SCZ CNVR was defined as any CNV region that was detected in the original 91 WGTP dataset (SCZ_{WGTP}n91), and was present in two or more patients in the combined cohort of 297 ($SCZ_{WGP\text{-}Affy}$n297).

A total of 26 recurrent SCZ CNVRs were identified, with a frequency of 2 to 6 patients out of 297 (0.67- 2.02%) (Table 4.3). Although they were all relatively rare events, the observation of these recurrent, putative disease variants is consistent with the genetically heterogeneous model of schizophrenia, and the frequency range is comparable to the rare disease variants observed in previous CNV studies on schizophrenia (Rujescu et al. 2008; Stefansson et al. 2008).

We focussed on 7 variants which were reported in the same patient in both WGTP and Affymetrix platforms (Figure 4.12). Among them, a region at chromosome 21 (CNVR-480) and another at chromosome 9 (CNVR-297) were detected in 4 and 5 patients respectively. Both of these recurrent CNVRs had a frequency of >1% in the disease cohort, and were found to overlap with functionally relevant candidate genes namely *RCAN1* (Down syndrome critical region 1) and *OLFM1* (olfactomedin 1).

Table 4.3 Recurrent SCZ-specific variants as detected in 297 SCZ samples. CNVs were detected by WGTP and/or Affymetrix platforms. Regions overlapping with more than one entry in the Database of Genomic Variant (DGV) for >50% were indicated in the last column. *(**red**: Samples detected by both platforms; blue: samples detected by WGTP only; green: samples detected by Affymetrix only)*

WGTP Identifier	WGTP coordinates	AFFY coordinates	Samples	Genes	DGV
CNVR_97	chr3:189546564-189761158	chr3:189564871-189682182	7183,3945	LPP	
CNVR_123	chr4:91083114-91244292	chr4:91202170-91309914	3443,4255	MMRN1	DGV
CNVR_173	chr5:110212742-110577198	chr5:110337010-110448345	850,4707	TSLP,WDR36	
CNVR_225	chr7:17795285-18873121	chr7:17674099-18707719	5541,5425	HDAC9,SNX13	
CNVR_336	chr12:17168087-17349294	chr12:17172155-18038013	1295,4372,3409		
CNVR_337	chr12:17761729-18154145	chr12:17172155-18038013	3409,4372	RERGL	
CNVR_415	chr16:82947011-83160475	chr16:82977842-83091199	3652,7245	ATP2C2,COTL1,KIAA1609	
CNVR_52	chr2:86130487-86470772	chr2:86237140-86417803	3789,7196	IMMT,MRPL35,POLR1A,PTCD3,REEP1	DGV
CNVR_64	chr2:188772738-189096411	chr2:188736466-189061332	5324,3107	GULP1	
CNVR_91	chr3:143326111-143612526	chr3:143303643-143554513	4710,6704	GK5,TFDP2,XRN1	DGV
CNVR_184	chr6:1994540-2304998	chr6:2001510-2293508	3815,3549	GMDS	
CNVR_185	chr6:7088878-7351773	chr6:7069194-7174697	7294,3511	CAGE1,RIOK1,RREB1,SSR1	
CNVR_188	chr6:29361859-29468501	chr6:29031959-29446150	3071,5367	OR12D3,OR14J1	
CNVR_243	chr7:88344723-88962788	chr7:88495346-89067351	1295,4008	ZNF804B	DGV
CNVR_260	chr8:2303622-2411266	chr8:2318174-2570606	1784,3768,7182,4804,3339		DGV
CNVR_264	chr8:13581606-13896999	chr8:13543770-13664796	5541,4227		DGV
CNVR_267	chr8:47038752-47556339	chr8:47516254-47655712	1784,7403		DGV
CNVR_297	chr9:137112069-137464520	chr9:137200474-137862491	323,4179,5758, 3593,6269	C9orf62,OLFM1	DGV
CNVR_298	chr9:137590463-137867305	chr9:137200474-137862491	3503,3593,6269,5758	CAMSAP1,GLT6D1,KCNT1,LCN9,PAEP,SOHLH1	DGV
CNVR_347	chr12:131115601-131701464	chr12:131210873-131630315	3071,3944	DDX51,EP400,EP400NL,GALNT9,MUC8,NOC4L	
CNVR_362	chr13:113166604-113658050	chr13:113543086-113654850	6638,3890	ATP4B,DCUN1D2,FAM70B,GAS6,GRK1,TFDP1,TMCO3	
CNVR_375	chr14:103592317-103986703	chr14:103860158-104009366	3812,5390,3802,5660,4106,6736	KIF26A	DGV
CNVR_398	chr16:2870891-3098404	chr16:2937256-3042293	3802,1334	CCDC64B,CLDN6,CLDN9,FLYWCH1,FLYWCH2,HCFC1R1,IL32,KREMEN2,MMP25,PAQR4,PKMYT1,THOC6,TNFRSF12A,ZSCAN10	
CNVR_450	chr19:7034753-7162312	chr19:6843781-7059064	6289,4281,3443	INSR,ZNF557	DGV
CNVR_469	chr20:29267569-29480559	chr20:29334284-29732674	3377,1559,3987	DEFB115,DEFB116,DEFB117,DEFB118,DEFB119,DEFB121	
CNVR_480	chr21:34576661-34852873	chr21:34631628-34827865	899,3546,3728,7171	C21orf51,KCNE1,KCNE2,RCAN1	

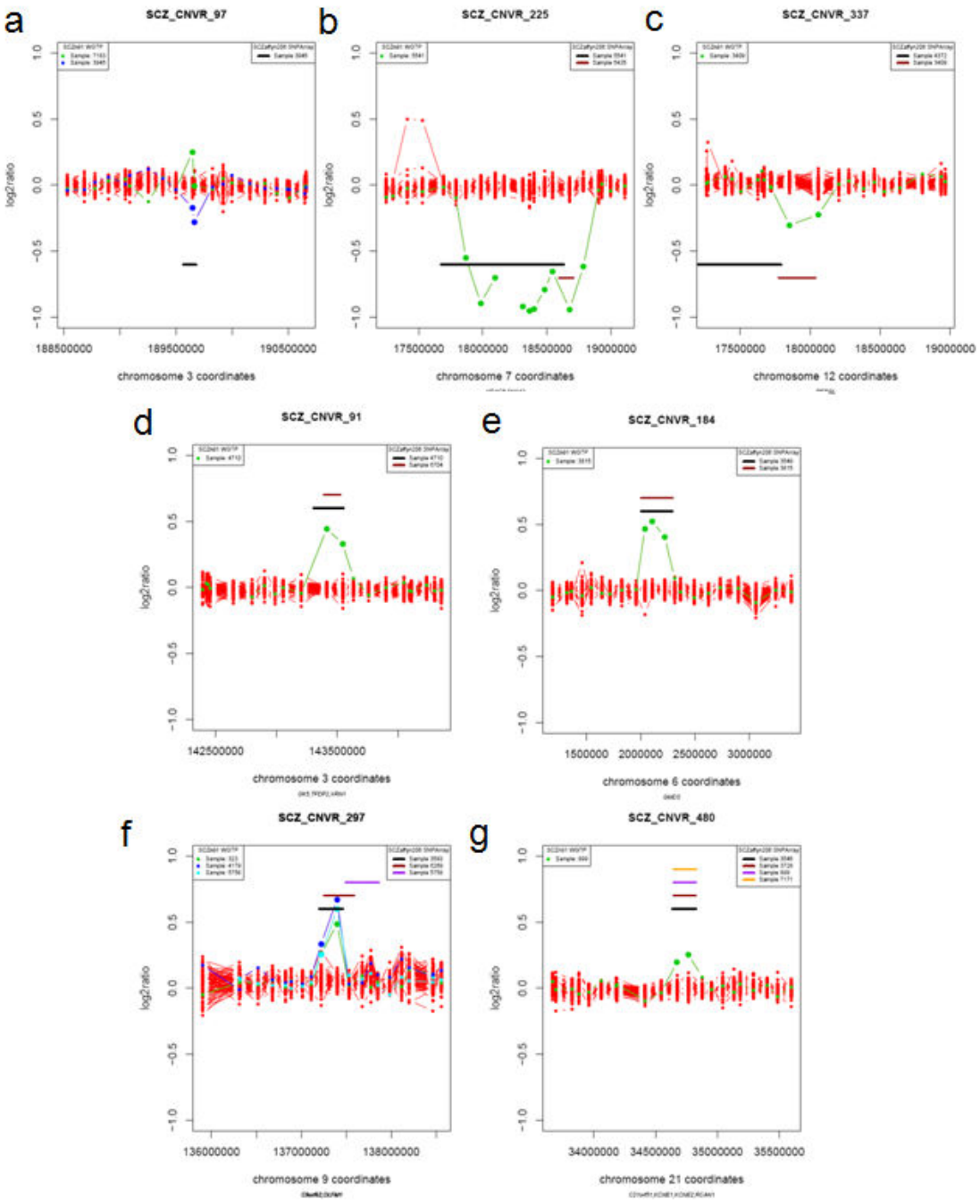


Figure 4.12 Recurrent SCZ-specific CNVR regions detected by both WGTP and Affymetrix platforms. The first 5 regions were identified in 2 SCZ samples: **a)** chr3; 189 Mb (*CNVR_97*) **b)** chr7; 176-187 Mb *(CNVR_225)* **c)** chr12; 171-181Mb ***(****CNVR_337)* **d)** chr3; 143 Mb (*CNVR_91*) **e)** chr6; 1.9-2.3 Mb *(WGTP_CNVR_184).* The final 2 regions occurred in 5 and 4 samples in the SCZ cohort respectively **g)** chr9; 137 Mb *(CNVR_297)*; **h)** chr21; 346Mb *(CNVR_480)*

4.5.1 A CNVR at Down Syndrome Critical Region 1 (RCAN1/DSCR1)

CNVR-480 is a duplicated locus affecting four patients (899, 3546, 3728 & 7171), one detected in SCZ_{WGTP}n91 and another three detected in SCZ_{Affy}n206 (Figure 4.12f). The 3 cases from the Affymetrix platform revealed the same breakpoints at chr9:137200474-137862491. The ~200 kb duplication region harbours two potassium channel genes *KCNE1* and *KCNE2*. Moreover, the gene Down Syndrome Critical Region 1 (*DSCR1*), also known as *RCAN1*, is located at the 3' breakpoint of the duplication according to Affymetrix SNP data. *RCAN1* has been linked to Down Syndrome and Alzheimer's disease (Hoeffer et al. 2007; Porta et al. 2007; Keating et al. 2008).

The gene *RCAN1* encodes a regulator of calcineurin, a calmodulin-dependent protein phosphatase. Both animal models and human studies have established the role of the calcineurin-dependent cascade in neuronal signal transduction and synaptic plasticity, as well as its involvement in psychosis (Eastwood et al. 2005). Genes that encode subunits of calcineurin, for example *PPP3CC*, have been demonstrated as associated with schizophrenia (Eastwood et al. 2005). Furthermore, *RCAN1* was recently shown to affect the expression of GSK-3beta (Ermak et al. 2006), another well-established schizophrenia and bipolar candidate gene (Koros and Dorner-Ciossek 2007). Knockout mouse models of *RCAN1* showed impairment in spatial learning and memory (in particular long-term potentiation), as well as sensorimotor deficits (Hoeffer et al. 2007).

4.5.2 A Variant Near Olfactomedin1 and other Recurrent CNVRs

CNVR-297 is another duplication region detected in three patients (5758, 323 & 4179) SCZ_{WGTP}n91, and in two additional samples (3593 & 6269) in SCZ_{Affy}n206. The duplication at 9q34 (Figure 4.12g) is downstream of the rat neuronal olfactomedin-related ER localized protein *OLFM1*, a gene abundantly expressed in the brain. Characterization of the CNV by oligonucleotide array revealed the same breakpoint in the patients (5758, 323 and 4179). We subsequently genotyped this variant in an extended cohort of 304 cases and 309 controls but failed to show significant association of the CNV with disease.

The other 5 recurrent regions were detected in 2 out of 297 samples. These regions harboured the following genes: LIM-domain containing preferred translocation partner in lipoma *LPP* (Figure 4.12a); histone deacetylase 9 (*HDAC9*), a histone deacetylase, which couples neuronal activities to muscle cells (Mejat et al. 2005); sorting nexin (*SNX13)*, a protein involves in G protein signalling and intracellular trafficking (Zheng et al. 2006) (Figure 4.12b); Ras-related and estrogen-regulated growth inhibitor-like protein *RERGL* (Figure 4.12c); Glycerol kinase 5 *GK5*, Transcription factor *DP-2* (TFDP2) and exoribonuclease *XRN1* (Figure 4.12d) and finally GDP-Mannose 4,6-Dehydratase *GMDS* (Figure 4.12e).

We evaluated the specificity of the recurrent CNVRs in the disease cohort by comparing them to the Database of Genomic Variants (DGV), a public resource cataloguing copy number variants and other structural rearrangements in normal individuals. 11 out of the 26 recurrent regions overlap (>50%) with more than one DGV CNV entries.

Whilst recognizing that some of these recurrent CNVRs overlap with DGV entries, such an overlap does not necessarily refute disease association for two reasons: 1)

Incomplete penetrance of genetic variants in schizophrenia suggests a disease-causing variant could be present in normal individuals although at a lower frequency. For example, known SCZ-associated loci, such as the 15q11.2 deletion, were shown to overlap with DGV entries. 2) Schizophrenia is complex psychiatric illness, occurring in 1% in population, with possibility of late disease onset. Since the apparently normal individuals reported with variants in DGV usually did not undergo rigorous neurological examinations, their disease status and background remain questionable.

In summary, in the extended cohort of $SCZ_{WGP\text{-}Affy}$n297, we detected 26 recurrent regions as putative disease variants. Candidate genes such as *RCAN1* were highlighted due to the relatively high frequency in cases and its biological relevance. Nevertheless, to establish disease association, a substantially larger sample size would be required.

4.6 Frequent Copy Number Variations in SCZ and LBC

4.6.1 Variance-based Clone-by-Clone Cohort Comparison

To investigate the role of frequent CNVs in disease, our first approach was to perform pair-wise comparisons on the SCZ versus LBC array CGH data for each autosomal BAC clones on the array, based on a statistic measure "Vst" (with reference to the F-statistic "Fst" in population genetics) (Tills 1977).

Vst is derived from the individual and combined variance of two sets of quantitative data (i.e. log2ratio in the two cohorts), adjusted by the respective number of observations (i.e. DNA samples), and is summarized as:

$$\mathrm{Vst} = \left\{ \mathrm{V_{total}} - \frac{(\mathrm{V_{P1}} \times \mathrm{N_{P1}} + \mathrm{V_{P2}} \times \mathrm{N_{P2}})}{(\mathrm{N_{P1}} + \mathrm{N_{P2}})} \right\} \div \mathrm{V_{total}}$$

V_{total} = Total variance of the two dataset

V_{P1} = Variance of the first population (e.g. Schizophrenia)

V_{P2} = Variance of the second population (e.g. LBC)

N_{P1} = Number of samples in the first population

N_{P2} = Number of samples in the second population

Vst scores, ranging from 0 to 1, were computed for each autosomal BAC clone spotted on the array. This variance-based approach has previously been used in the WGTP study on 269 HapMap samples (Redon et al. 2006) to examine population differentiation with respect to CNVs (Figure 4.13a). High Vst implies a considerable degree of differentiation among populations and vice versa. Regions previously reported with high

Vst among population, for example the copy number variant at *CCL3L1* associated with HIV-1 susceptibility, were detected with Vst scores up to 0.5 - 0.8 for the affected clones (Redon et al. 2006).

In contrast, Vst scores computed from schizophrenia versus LBC dataset suggests the case and control cohorts have minimal variation difference across the genome (Vst scores < 0.2) (Figure 4.13b). This remarkable similarity can be explained by the fact that schizophrenia and LBC are effectively the same ethnic population. It also implies that no single region of the genome, at the resolution of the WGTP array, shows substantial cohort-wide differentiation between SCZ versus LBC.

To conclusively exclude any Vst-based detection of copy number differentiation between the cases and controls, we examined the top hits of the Vst analysis. First all BAC clones were ranked according to Vst score and a Vst distribution for all clones was plotted. The top 1% clones, each with a Vst score above the threshold of 0.06853, were further filtered for regions with consecutive clones, based on the assumption that single clones with a technical artefact will have high probability of being in the tail of the Vst distribution and therefore should be discarded from the Vst analysis.

19 consecutive-clone regions were included in the top 1% Vst distribution. If these top Vst regions were true (i.e. demonstrating real population-wide differentiation between SCZ and LBC), we should be able to detect CNVs in samples with extreme log2ratio. It is also probable that any such CNV would be small in size compared to the size of a BAC clone, thus giving only a slight change of clone variance with relatively low Vst scores (compared to previous studies). We tested for the presence of cryptic CNVs in five of the top Vst regions using a high-resolution custom designed Nimblegen oligo array, hybridized using 10 available DNA samples from the SCZ cohort. Samples were

hybridized in pairs such that for each test Vst region, there was at least one DNA pair with high versus low log2ratio (Table 4.4).

4 of the 5 regions tested showed no signs of CNV. The Vst region at chr10 55-56Mb was detected with a ~50 kb CNV which overlaps the intron of gene *PCDH13* (protocadherin 13). However, occurrence of this CNV does not correlate with the log2ratio of the two clones at the Vst region, and it is concluded that the CNV was found by chance with no relation to Vst. In summary, none of the Vst regions were detected with CNVs that could be explained by the log2ratio.

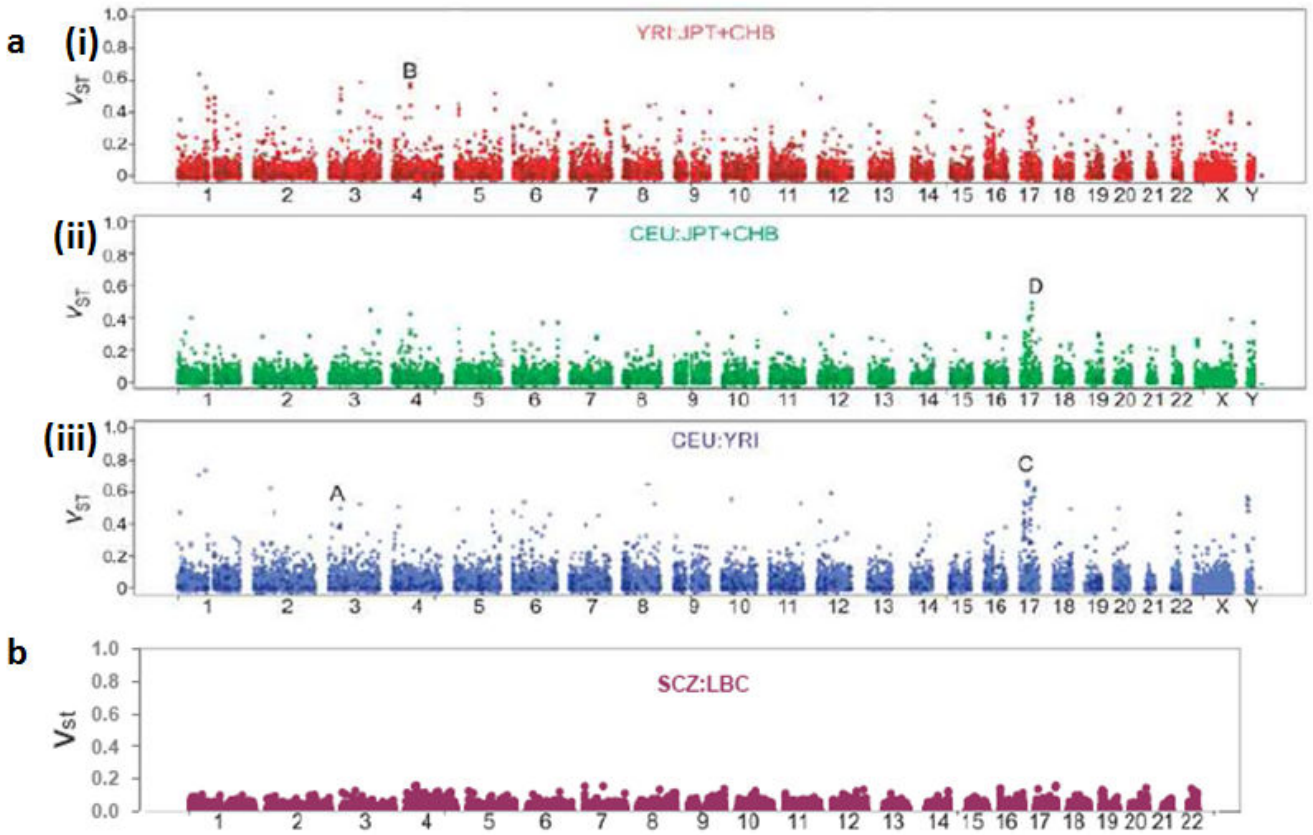


Figure 4.13 Vst scores to identify clones showing SCZ and LBC differentiation in the WGTP analysis. a) Vst scores from previously published HapMap WGTP CNV data (Diagram adopted from Redon *et. al.*). Comparisons were made between populations: **(i)** Yoruba (African) Versus Japanese/Chinese (Asian) **(ii)** Utah (European) Versus Japanese/Chinese (Asian) and **(iii)** Utah (European) Versus Yoruba (African). **b)** Vst scores derived from WGTP log2ratio in the current study, comparing schizophrenia versus control Scottish samples. Vst scores for all clones along the genome showed minimal differentiation (Vst below 0.2).

Table 4.4 Validating 5 top Vst-regions using custom-design oligonucleotide array. Each top Vst region had two consecutive clones with Vst score above the threshold of significance. For each Vst hybridization, a sample with high log2ratio (blue) was hybridized against another sample with low log2ratio (yellow). The final columns showed the minimum/maximum log2ratio of that clone for all samples in the cohort.

chr_corrdinates	signs of CNV?	clones	DNA Pair A		DNA Pair B		DNA Pair C		DNA Pair D		DNA Pair E		min	max
			4398	3341	3584	3789	241	4179	3652	5386	7180	5390		
chr5:84194412-84418390	no	Chr5tp-6E5	-0.0012	0.1146	0.0628	-0.0156	0.0373	0.0667	-0.0247	0.0582	-0.0254	0.0915	-0.0759	0.2033
		Chr5tp-16D5	0.0270	0.0897	0.0176	-0.0216	0.0368	-0.0068	-0.0342	0.0282	-0.0005	0.0684	-0.0754	0.1204
chr10:55963328-56236143	yes, but not in expected sample	Chr10tp-2H6	0.0099	0.0660	0.0935	0.0398	0.0428	0.0490	**-0.0192**	**0.0697**	-0.0430	0.1080	-0.0998	0.1793
		Chr10tp-5H2	-0.0439	-0.0304	-0.0470	-0.0677	0.0122	-0.0230	**0.0159**	**-0.0501**	0.0433	0.0156	-0.1770	0.1567
chr10:69928325-70219374	no	Chr10tp-2E7	0.0002	0.0746	-0.0495	-0.0745	-0.0238	0.0005	-0.0082	0.0435	0.0493	0.1126	-0.0977	0.1137
		Chr10tp-3C7	-0.0138	0.1468	-0.0396	-0.0414	0.0513	0.0695	0.0325	0.0483	-0.0109	0.0896	-0.1194	0.1468
chr16:229695-801844	no	Chr16tp-11B4	0.0946	0.0580	0.2023	0.0378	0.1668	-0.0066	0.0600	0.1321	-0.1565	0.1177	-0.1565	0.2023
		Chr16tp-12E7	0.0526	0.0905	0.1085	-0.0225	0.1995	0.0024	0.0611	0.1185	-0.1484	0.1167	-0.1484	0.1995
		Chr16-32k-3E2	0.1380	0.0333	0.1224	0.0032	0.1899	0.0910	0.0636	0.0352	-0.0798	0.1491	-0.1636	0.2275
		Chr16-32k-2H14	0.1126	0.0627	0.1709	-0.0218	0.2415	0.1336	0.0488	0.1390	-0.0882	0.2213	-0.1198	0.3709
chr22:48767142-49057642	no	Chr22-32k-1J4	0.0224	0.0197	0.1257	0.0024	0.0564	0.0529	-0.0118	0.0540	0.0200	0.0794	-0.0628	0.1558
		Chr22tp-8F6	0.0829	0.0871	0.1937	-0.0182	0.1411	0.0105	0.0158	0.0784	0.0070	NA	-0.1003	0.1937

4.6.2 CNV Genotyping with Bivariate Clustering

A second approach to detect frequency bias in common variants is to genotype CNV based on log2ratio clustering, followed by association tests on the resulting contingency tables of genotype distributions. For data clustering, we employed a model-based, expectation maximization hierarchal clustering algorithm known as Mclust (R package). The algorithm was performed on log2ratio of consecutive BAC clones with high correlation (Pearson correlation >0.5). This method is more effective for estimating genotypes at multi-allelic and complex regions than CNVFinder, which simplifies CNV status to a trichotomous classification of “deletion”, “normal” and “duplication” with respect to the reference DNA. Consistent with our previous analyses, we used a consecutive clone approach (bivariate clustering), decreasing the likelihood of false positives as a result of individual variability of any single clone.

Combined WGTP data on cases (n=91) and controls (n=92) were used for bivariate clustering. A total of 577 clones (out of 26,954 autosomal BAC clones on the WGTP array) were “genotypable” (i.e. resulting in more than one cluster when clustered with its downstream neighbour). The number of clusters was set at a range of 2 to a 9. Examples of bivariate clustering diagrams showing DNA samples with deletion (a), duplication (b) and a multi-allelic region (c) are shown in Figure 4.14.

We genotyped each schizophrenia and control sample for the 577 “genotypable” regions (except samples classified with high uncertainty which were removed in our algorithm). A chi-square test was then performed for the resulting genotype distribution, with the null hypothesis that SCZ and LBC were drawn from identical population.

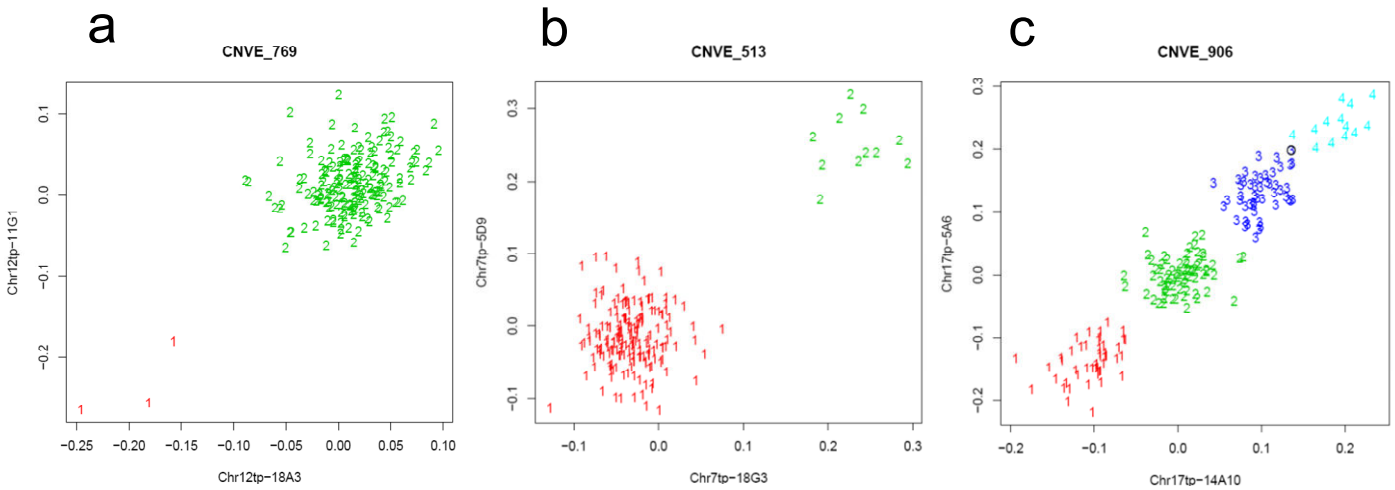


Figure 4.14 Examples of bivariate clustering based on log2ratio of consecutive clones. a) A region with majority of samples centred at log2ratio of 0.0 ("normal", cluster 2) with a minority of samples as deletion ("deletion", cluster 1); **b)** A region with a majority of samples centred at log2ratio of 0.0 ("normal", cluster 1) with a minority of samples as duplication ("duplication", cluster 2); **c)** A multi-allelic region with multiple clusters (cluster 1,2,3 and 4) with different copy numbers. *(x-axis: log2ratio of a particular BAC clone on the WGTP array; y-axis: log2ratio of an adjacent clone)*

31 of the 577 “genotypable” clones demonstrated significant difference in CNV genotype distribution (p-value < 0.05) following bivariate clustering (Appendix D-3). The clustering results for each of the 31 clones were manually inspected. 3 regions were selected for further investigation based on manual clustering and candidate gene involvement (Figure 4.15).

Two of the selected regions (Region 1 & 2) were genotyped in an extended cohort of cases and control DNA samples using PCR-based techniques. The two candidate regions overlap the gene neuronal adhesion molecule Close Homolog of L1 (*CHL1*) at 1p36 and the nicotinic receptor fusion gene *CHRFAM7A* (see Chapter 5) at 15q13 (see Chapter 5).

Region 3 was located near another cluster of highly variable segmental duplication on chromosome 17. The multi-allelic region partially overlapped N-ethylmaleimide-sensitive factor (*NSF*), a gene in the NMDA-receptor signalling complex which is important in membrane vesicle fusion, neurotransmitter release, trafficking of AMPA receptors and other major roles in synaptic plasticity (Haas 1998). Further experiments were performed to explore the genomic structure of the region (see Chapter 6).

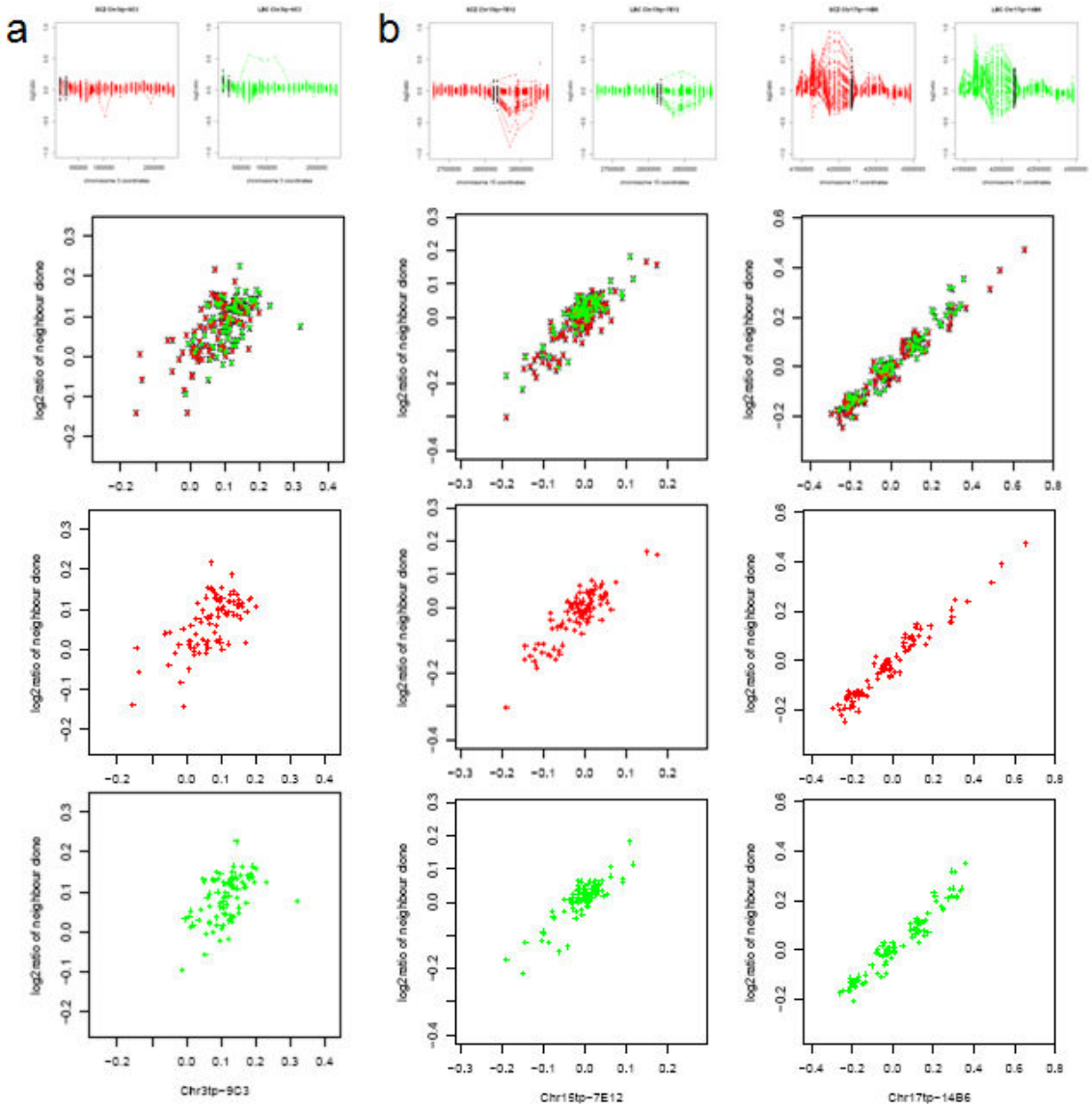


Figure 4.15 Three regions with significant difference of genotype distributions between SCZ and LBC as detected by bivariate clustering. **a)** 3p26 near *CHL1*: 3 CNV genotypes. **b)** 15q13 near *CHRFAM7A:* 2 main CNV genotypes. **c)** 17q21 near *NSF*: multi-allelic region with at least 5 CNV genotypes. *(Red: SCZ; Green: LBC*)

4.7 Chapter Summary and Discussion

In this chapter we described results from a population-based genome-wide CNV screen for schizophrenia patients and matched normal controls. The initial array CGH study was based on 91 patients and 92 controls, with hybridization performed on whole-genome tiling path BAC array. A subset of CNVs was validated by quantitative PCR and Nimblegen oligo-nucleotide array experiments. We also integrated CNV calls on a subset of 39 patients generated from Affymetrix SNP 6.0 array hybridizations, with data provided by the ISC and University of Edinburgh. The analysis was divided into two components: (a) Identification of rare, disease-causing variants in the case cohorts as putative disease candidates; & (b) Assessment of frequent copy number polymorphisms to investigate possible disease association.

A number of recent studies suggested that rare CNVs are responsible for a significant proportion of schizophrenia. In particular, Walsh *et al.* reported that in their case-control cohort, rare genic copy number variants existed as 15% of CNVs in cases but only as 5% in control (Walsh et al. 2008). If this were extrapolated to the general disease-control populations, as many as two thirds of the rare genic variants identified specifically in a case-control study should be implicated in schizophrenia pathogenesis. Accordingly, we focused our analysis of the WGTP data on such rare genic variants in the disease cohort.

We identified 113 rare, schizophrenia specific candidates not detected in the matched control set. We also replicated the finding at the 15q11.2 recurrent deletion region previously demonstrated as schizophrenia-associated in a large-scale study (Stefansson et al. 2008), demonstrating the ability of our approach to reveal potential schizophrenia-associated CNVs. However, to prove disease association of any of these variants, one would need a substantially larger case-control cohort.

Instead, we pursued a candidate CNV discovery approach, and aimed to catalogue SCZ-specific rare variants as a collective set of disease candidates. Studying such rare copy number mutations will lead to a better understanding of the genetics of schizophrenia: first, accumulating information on rare CNVs in SCZ patients will help in discriminating true recurrent disease variants from benign mutations; second, the rare disease variants in many different patients or families may converge in biological pathways, providing insight into the disease phenotype.

An important approach to assess clinical relevance of this set of rare CNVs is to examine the loci with respect to gene content (Lee et al. 2007). For instance, rare case-specific variants with the presence of morbid OMIM gene(s) with brain-related function, or gene(s) previously linked to a neuropsychiatric disorder, would be considered plausible schizophrenia candidates responsible for disease phenotypes. Among the disease-specific variants we detected a number of such candidate genes, including *PIK3C3*, *SGCE, OXTR* and *RCAN1*, all providing functional relevance to psychiatric disorders.

In the second half of the chapter, we described the genotyping of common copy number polymorphisms by a hierarchical model-based clustering algorithm. Genotype distributions in cases and controls were compared using a nonparametric test with the null hypothesis that patients and controls were drawn from the same distribution of copy numbers. A total of 31 CNV regions were identified with a distribution significantly different in cases versus control (chi-square test, p-value <0.05). By applying this case-control association approach to our CNV study, we face the limitation of a) the adequacy of our sample size to confer enough statistical power; and b) the accuracy of converting CNV measurements (as log2 fluorescent ratio) to CNV genotypes. We therefore subsequently genotyped two of these candidate regions, using more precise PCR-based methods for copy number

detection, and in the extended case-control cohorts of ~ 300 + 300 samples (discussed in Chapter 5).

Regarding experimental design, of particular importance to a population-based case-control CNV screen was the appropriate choice of controls (McCarroll and Altshuler 2007). A well-matched control cohort is crucial for a number of reasons: First, a number of studies showed that private rare CNVs associated with specific continental ancestry were common (Khaja et al. 2006; Locke et al. 2006; Redon et al. 2006), perhaps more so than for private SNPs (Jakobsson et al. 2008). To detect rare variants as disease candidates we employed an ethnically well-matched Scottish control cohort for our Scottish patient cohort.

Similarly, well-matched case and control ethnicity is critical for tests of association. The frequency of CNV of interest may differ across populations. Other genetic variants which could alter liability to disease (e.g. the presence of inversion predispose to CNV) may also be population-specific. Moreover, in some cases of CNV disease association such as *CCL3L1* in HIV-1 progression, the copy number *per se* was not a main disease determination factor, rather the gene dose relative to the population average was critical (Gonzalez et al. 2005). All of these again argue for the significance of a well-matched control cohort.

Other practical issues related to case-control study include the differential quality, source and experimental treatment of DNA samples, which could influence CNV data quality in cases versus control. As described in this chapter, our study did face a limitation regarding sample quality. This translates into a bias of CNV detection sensitivity in control versus cases, and has complicated statistical disease association testing, including the calculation of overall disease burden of CNVs.

Finally, we would emphasize on using a control cohort that is free from psychiatric disease for proper schizophrenia CNV study. In this work, the Lothian Birth Control individuals (Deary et al. 2007) have been thoroughly examined for neuropsychiatric functions, with neurological tests performed at the age of ~90 (far beyond the normal age of onset of schizophrenia), and therefore should be free from psychiatric illness. Nevertheless, since schizophrenia is complex and genetic variants could have incomplete penetrance, disease-related variants may still occur in the control cohort.

Despite a few inevitable constraints, our data generated a number of validated, potentially disease-associated rare variants in the schizophrenia cohort. A number of these rare variants harbour psychiatric-disease related genes. Further bioinformatic work to explore these rare variants using a functional biological pathway approach (e.g. the Gene Ontology pathway) could provide more insight into the disease pathways.

Chapter 5

Genotyping Two Candidate CNVs in an Extended Case-Control Cohort

5.1 Candidate I: *CHL1* (Close Homolog of L1) at 3p26

5.1.1 Functional Significance of *CHL1*

The first CNV genotyped in the extended replication cohort was located 5' to the gene Close homolog of L1 on chromosome 3p. Close homolog of L1 (*CHL1* or *CALL*) belongs to the L1 family of cell adhesion molecules, a subgroup of the immunoglobulin superfamily (Leshchyns'ka et al. 2006). The mammalian L1 family comprises four neural recognition molecules: *L1*, *CHL1*, neurofascin and *NrCAM*. Together with other families of neural recognition molecules, such as cadherins and integrins, the four L1 family proteins play important roles in neural development, neuronal regeneration and synaptic activities (Maness and Schachner 2007). Members of the L1 family have been implicated in a wide spectrum of brain abnormalities and neuropsychiatric disorders. For instance, *L1CAM* is responsible for several X-linked mental retardation syndromes, as well as increased risk of schizophrenia (Kurumaji et al. 2001). *NrCAM* is the binding partner of the synapse associated protein *SAP102 (Davey et al. 2005)*, a MAGUK family member involved in non-syndromic X-linked mental retardation (Laumonnier et al. 2007). Moreover, allelic variants of *NrCAM* have been implicated in autism (Hutcheson et al. 2004; Sakurai et al. 2006; Marui et al. 2008) and addiction vulnerability (Ishiguro et al. 2006).

CHL1 is expressed in subpopulations of neurons and glial cells in both the central and peripheral nervous system during early development, starting from embryonic day 13 in mice (Hillenbrand et al. 1999). The protein is involved in a variety of neuronal functions including: (i) the promotion of neurite outgrowth (Holm et al. 1996; Hillenbrand et al. 1999), (ii) neuronal migration (Buhusi et al. 2003; Demyanenko et al. 2004; Nishimune et al. 2005) and survival (Chen et al. 1999); (iii) modulation of

synaptic plasticity (Montag-Sallaz et al. 2002; Morellini et al. 2007); (iv) synaptic targeting via the clathrin dependent pathway (Leshchyns'ka et al. 2006); and (v) the regulation of axon regeneration (Chaisuksunt et al. 2000; Zhang et al. 2000; Jakovcevski et al. 2007).

CHL1 was demonstrated to have involvement in mental retardation and schizophrenia in both human and mouse genetic studies. In human, *CHL1* was one of the genes deleted in the 3p-syndrome characterized by mental and psychomotor retardation with congenital anomalies. Haplo-insufficiency of the gene was suggested to be responsible for the cognitive phenotype in this syndrome (Angeloni et al. 1999). *CHL1* was also identified at the translocation breakpoints in a patient with non-specific mental retardation (Frints et al. 2003). In addition, two association studies have established a link between a mis-sense polymorphism in the signal peptide region of *CHL1* (Leu17Phe) and schizophrenia (Sakurai et al. 2002; Chen et al. 2005).

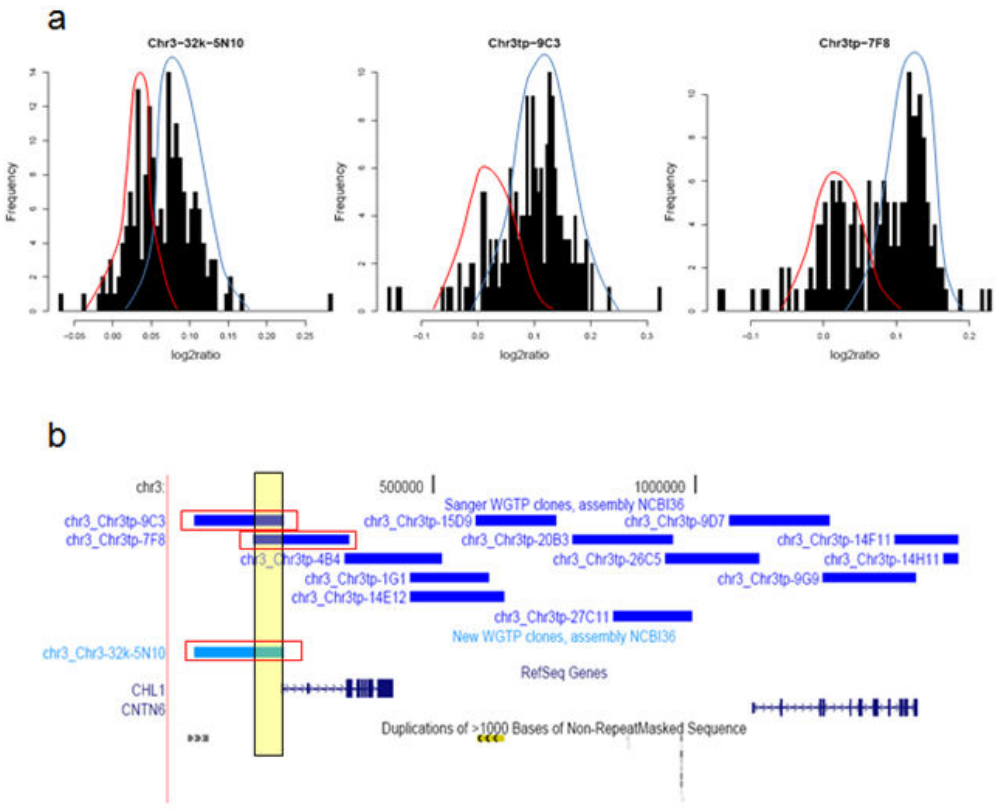
In *CHL1* knockout mice, a delayed reactivity to environmental stimuli was observed, suggesting a deficit in extracting relevant information from the environment in these mice (Morellini et al. 2007). Chl1-/- mice were also less aggressive and have altered social behavior in exploration tests including open field, resident-intruder test and urine marking (Frints et al. 2003; Morellini et al. 2007). In sensorimotor gating functions, Chl1-/- mice exhibited impaired prepulse inhibition (PPI) of the acoustic startle response (Irintchev et al. 2004), a key endophenotype of schizophrenia. In terms of cognitive ability, synaptic plasticity in juvenile Chl1-/- mice showed enhanced perisomatic inhibitory synaptic transmission, coupled with impaired long-term potentiation (LTP) in the hippocampus (Nikonenko et al. 2006). In adults, hippocampus synaptic plasticity was normal, but there was an enhanced basal synaptic transmission in the dentate gyrus (Morellini et al. 2007). Finally, in the Morris

Water Maze test, Chl1-/- mice displayed an altered exploratory pattern (Montag-Sallaz et al. 2002; Frints et al. 2003), suggesting cognitive processes involving spatial memory could be impaired. Altogether, *CHL1* knockout mouse models suggested mutations of the gene could result in schizophrenia by disrupting key neurological functions related to disease phenotypes.

5.1.2 Structure of the *CHL1* 5' Deletion Polymorphism

A copy number polymorphism 5' of *CHL1* was detected via the log2 ratio distributions of three consecutive clones at the most distal end of chromosome 3p (Chr3-32K-5N10, Chr3tp-9C3, & Chr3tp-7F8) (Figure 5.1). Visually at least two peaks could be detected in the log2ratio histograms of the three affected clones (Figure 5.1a & b). Furthermore, the log2ratio distributions for three clones showed strong positive correlation with each other (Pearson correlation of log2ratio >0.6), but not with other neighbouring clones (Figure 5.1c).

By plotting the log2ratio of one of the three clones against another and applying bivariate clustering algorithm, the data clustered into three classes (Figure 5.2). Cluster A and B were non-discrete, with only a small shift in log2ratio, indicate the presence of a small CNV. We hypothesized that the 3 clusters represented normal (cluster A), heterozygous deleted (cluster B) and homozygous deleted (cluster C) samples respectively, with the reference DNA sample (HapMap NA10851) being heterozygous deleted.



c

Clone	Position on Chr3	Pearson Correlation with Clone Chr3tp-9C3
Chr3-32k-5N10	132207.5	0.62811
Chr3tp-9C3	132217.0	1.00000
Chr3tp-7F8	251366.0	0.61511
Chr3tp-4B4	425974.0	0.09403
Chr3tp-1G1	533672.5	-0.08985
Chr3tp-14E12	547467.0	-0.00283
Chr3tp-15D9	658724.5	0.02715
Chr3tp-20B3	863307.0	0.05340
Chr3tp-27C11	920543.5	0.11604
Chr3tp-26C5	1031880.0	-0.01170

Figure 5.1 A copy number polymorphism at chromosome 3p26.3 was detected by three BAC clones (a) Log2ratio histograms* of the 3 BAC clones at 3p26.3: *Chr3-32k-5N10* (left), *Chr3tp-9C3* (middle) and *Chr3tp-7F8* (right). Bi-allelic distribution (two peaks) in the histograms indicates a common CNV in the genomic region. (**b)** Location of the three clones (red box) and their overlapping region (yellow box) with respect to the location of the gene *CHL1*. **(c)** Pearson correlation coefficients of log2ratio indicated high correlation of *Chr3tp-9C3* with *Chr3-32k-5N10* and *Chr3tp-7F8,* but not with other neighbouring clones.

* histograms generated from combined SCZ & LBC cohort data

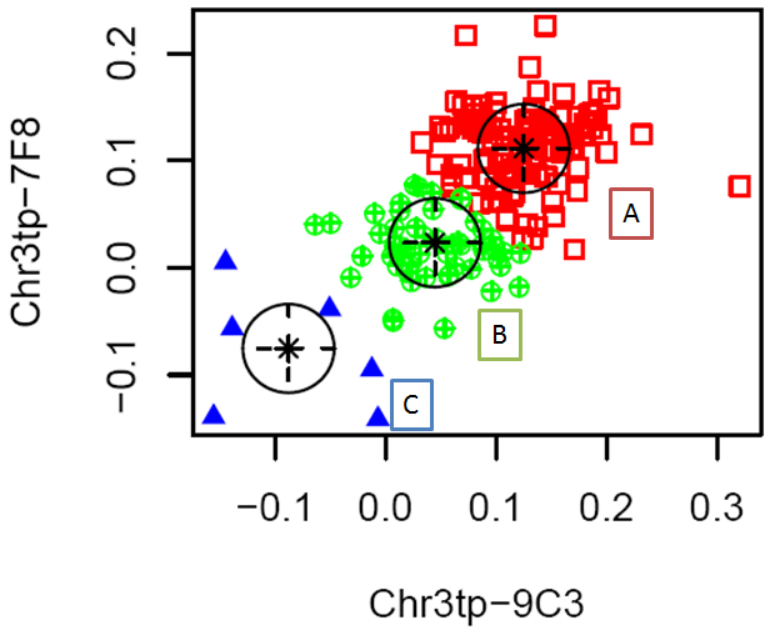


Figure 5.2 CHL1 CNV genotypes from bivariate clustering. Applying a model-based hierarchical clustering algorithm on clones Chr3tp-7F8 and Chr3tp-9C3 resulted in three major clusters, which correspond to (A) normal, (B) heterozygous deletion and (C) homozygous deletion genotypes.

The location of the *CHL1* CNV was determined and breakpoints were sequenced in five HapMap samples that were either heterozygous or homozygous carriers of the deletion polymorphism (NA12144, NA12892, NA18576, NA18976 and NA10851) (Figure 5.3). All samples generated the same breakpoint structure, with distal and proximal breakpoints localised to *Alu* repeat sequences (Figure 5.3a). The CNV occurred at approximately 4000 bp 5' of the gene *CHL1*, with the distal breakpoint estimated at chr3:204,000, and the proximal breakpoint at chr3:209,500. Recently released high-resolution data from the Nimblegen oligonucleotide array platform (Genome Structural Variation Consortium 2008) further confirmed the deletion size in the heterozygous reference DNA sample NA10851 (Figure 5.3b). The size of the CNV was estimated to be around 5.5 kb.

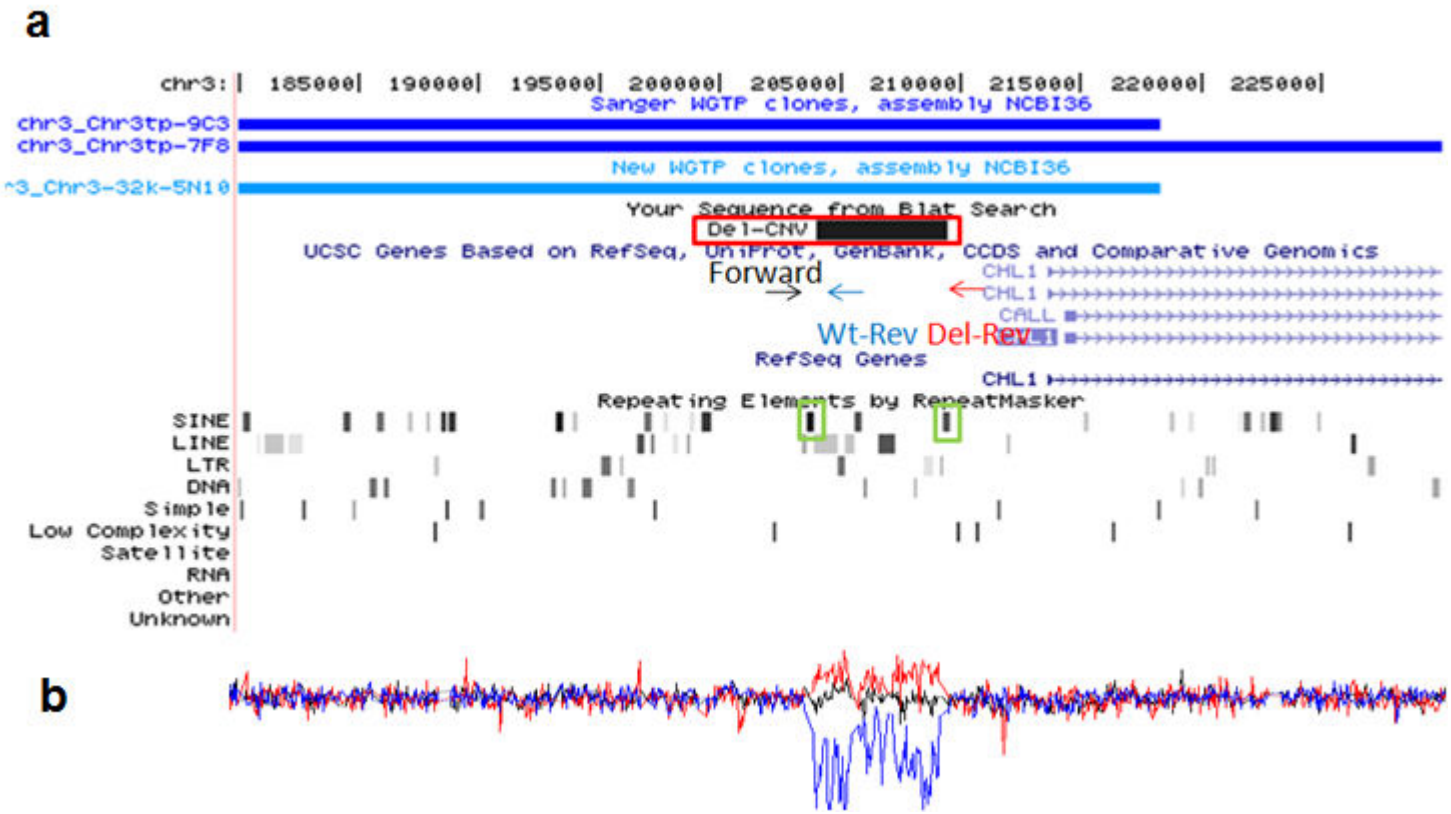


Figure 5.3 Genomic location of the *CHL*1 5' CNV. The CHL1 5' deletion polymorphism (black box) was validated using both PCR and Nimblegen oligonucleotide array. **(a)** PCR determined the CNV breakpoints to fall on a pair of SINE *Alu* elements (green boxes). Arrows represent the locations of the PCR primer series used for CNV genotyping and validation. →forward primers; ← reverse primers for the wild-type allele; ← reverse primers for the deletion allele **(b)** Nimblegen oligonucleotide array confirmed the size and location of CNV to be chr3:204,000-209500 and revealed both deletions and duplications in different samples.

5.1.3 Correlation of *CHL1* CNV to mRNA Expression

To assess the relationship of this 5' deletion polymorphism with *CHL1* mRNA expression, we extracted gene expression data from a recent publication (Stranger et al. 2007) assessing the effect of genetic variation (SNPs and CNVs) on expression phenotypes for 210 unrelated individuals from International HapMap Project (IHMC 2005).

In the analysis performed by Stranger *et al.* on a whole genome expression array platform (Illumina Sentrix Human-6 Expression BeadChip), the expression of the *CHL1* probe *GI_27894375-S* was shown to correlate with the presence of a CNV within the *CHL1* gene (at clone *Chr3tp-4B4*). This correlation was one of the five most significant associations between CNV and expression phenotypes in the CEU population as detected at the 0.05 permutation threshold. *Chr3tp-4B4* covers the 3' end of CHL1, and is immediately downstream of the three clones (*Chr3-32K-5N10*, *Chr3tp-9C3*, & *Chr3tp-7F8*) that span the 5.5 kb deletion polymorphism we described.

We next investigated expression data of the clones that span the deletion polymorphism of interest (*Chr3tp-9C3, & Chr3tp-7F8)*[3], by comparing them to published HapMap WGTP array CGH data from the same samples (Redon et al. 2006). As shown in Figure 5.4, correlation between array CGH log2ratio and the expression level of the *CHL1* probe (*GI_27894375-S)* was most prominent in CEU samples (COR=0.519, Pearson Correlation). Correlation was less significant in YRI or CHB +JBP samples.

[3] the third clone *Chr3-32K-5N10* was not present in the HapMap array CGH data and therefore was not analyzed

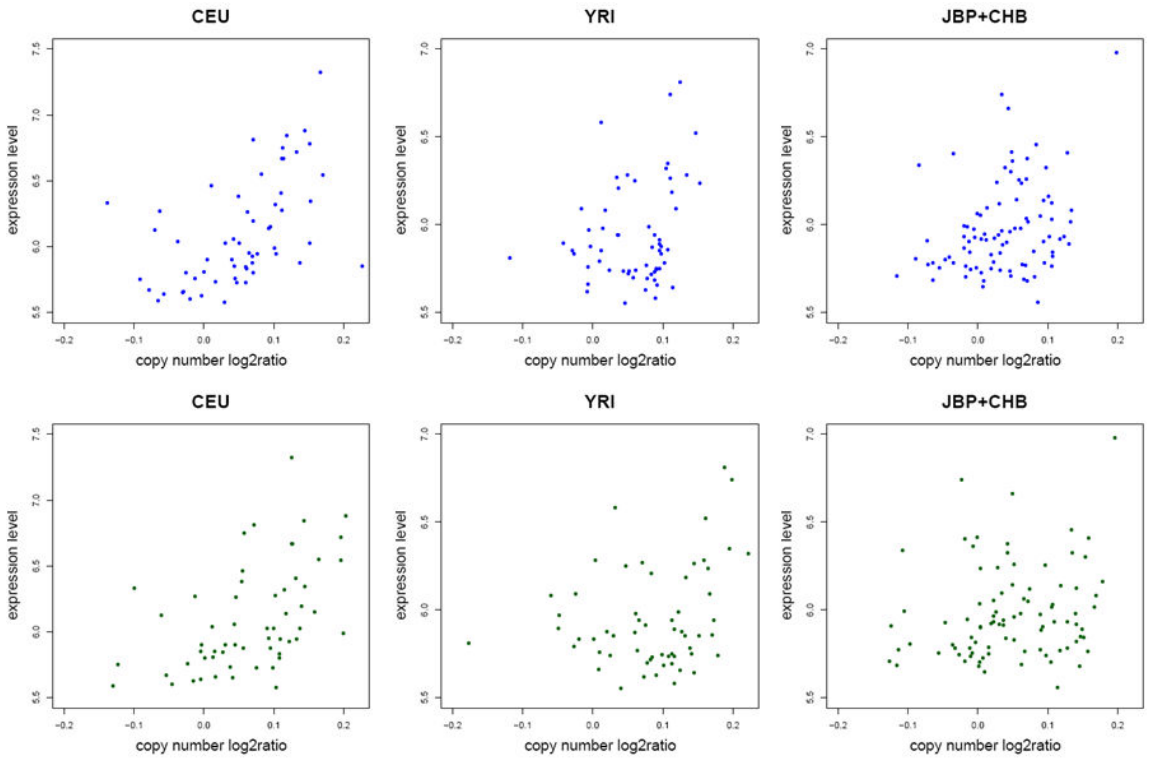


Figure 5.4 *CHL1* expression level against BAC clone log2ratio in HapMap samples. *CHL1* expression level (as measured by probe *GI_27894375-S* on Illumina Sentrix Human-6 Expression BeadChiparray) was plotted against log2ratio data from three BAC clones spanning the *CHL1* 5' CNV. Expression data was plotted against log2ratio of clones: **(a)** *Chr3tp-9C3 &* **(b)** *Chr3tp-7F8*

Left: CEU- Utah samples with European ancestry; Middle: YRI- Yoruba samples with African ancestry; Right: JBP+CHB Japanese and Chinese samples with Asian ancestry

All HapMap CEU samples were then genotyped for the *CHL1* 5' deletion polymorphism using a PCR strategy with one forward primer and two reverse primers spanning the deletion, generating a ~700 bp PCR band for a normal allele, or a ~400 bp band for the deletion allele. PCR genotyping allowed each sample to be categorized into discrete CNV genotypes of: homozygous deletion (0), heterozygous deletion (1) or normal (2) (Figure 5.5).

Considering only the CEU samples with European ancestry (Figure 5.6), the correlation of expression with clone log2ratio was 0.519 (Pearson Correlation) (Figure 5.6a). The comparison of expression with discrete CNV genotypes also showed a positive association with a p-value of 0.0436 (Figure 5.6b). In summary, our analysis revealed a potential positive correlation between a novel 5' *CHL1* CNV with expression of the *CHL1* gene in individuals of European ancestry.

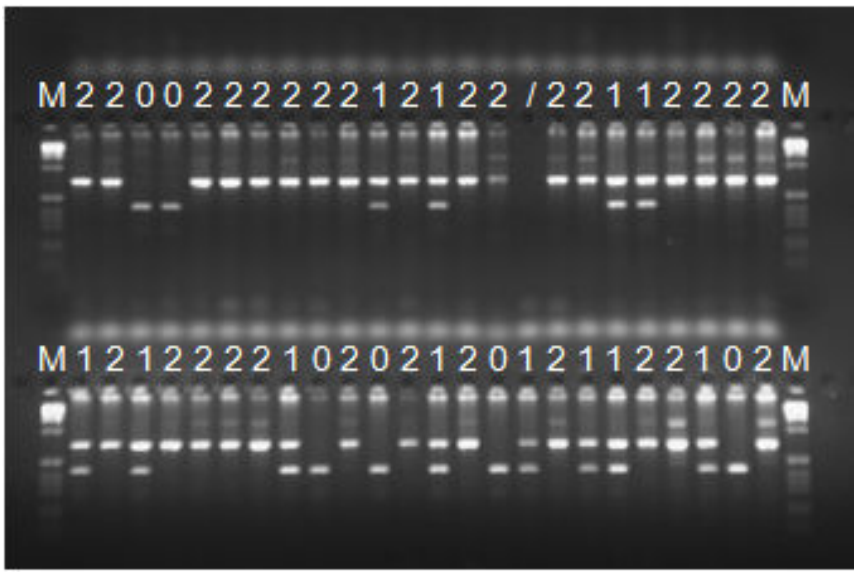


Figure 5.5 Genotyping of the *CHL1* 5' deletion polymorphism in HapMap samples. Genotyping was performed by PCR on 90 HapMap CEU samples. The ~700 bp band represents the "normal" allele and the ~400 bp band represents the deletion allele.

(M: marker; 0: homozygous deletion; 1: heterzygous deltion; 2: normal allele)

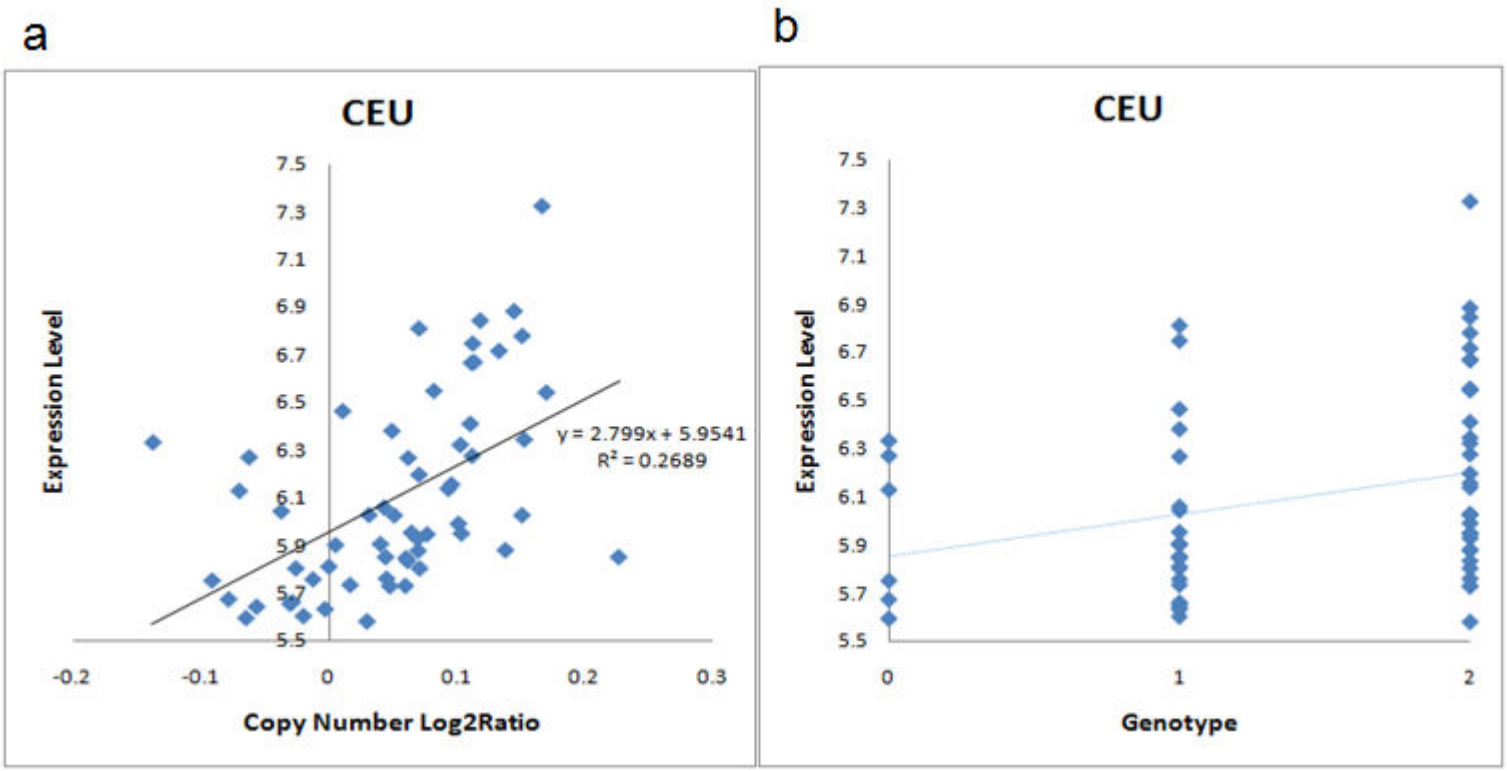


Figure 5.6 *CHL1* gene expression against copy number of the *CHL1* 5' CNV in HapMap CEU samples. (a) *CHL1* expression plotted against array CGH log2ratio signal demonstrate positive association (Pearson correlation= 0.519). **b)** *CHL1* expression plotted against discrete CNV genotypes resulted from PCR demonstrates positive association (p-value= 0.0436)

(0: homozygous deletion; 1: heterozygous deletion; 2: normal)

5.1.4 Genotyping CHL1 CNV in a Larger Case-Control Cohort

Our initial whole-genome schizophrenia CNV screen on 91 cases and 92 controls revealed possible association of the deletion polymorphism with disease status (p-value=0.00876) (section 4.6 & Table 5.1). Prompted by this enrichment of *CHL1* 5' deletion polymorphism in patients, and the correlation of *CHL1* expression with the 5' CNV genotypes, we attempted to replicate such findings in an extended case-control cohort.

354 SCZ DNA and 376 LBC DNA samples were genotyped using the same PCR strategy as described above. CNV genotype distribution [-/-: +/-: +/+] in the disease cohort was [12 (*3%*): 125 (*33%*): 239 (*64%*)], compared with [11 (*3%*): 121 (*34%*): 222 (*63%*)] in the control cohort (Table 5.2). The result did not show any association between CNV genotypes with disease (chi-square test, p-value = 0.964). Allele frequency distribution also demonstrated a lack of association (chi-square test, p-value = 0.906, data not shown). In conclusion, despite the initial positive association and the relevant biological evidence, we were not able to extend the findings of *CHL1* 5' deletion enrichment in disease to a larger case-control cohort.

Table 5.1 *CHL1* 5' CNV distribution determined by the original genome-wide array CGH data. CNV genotypes for 91 schizophrenia patients and 92 controls were estimated by log2ratio bivariate clustering of clones *Chr3tp-9C3* and *Chr3tp-7F8*. Samples with high uncertainty* (uncertainty value > 0.1), mostly between clusters A and B, were removed from analysis. A significant association of deletion carriers to disease status was observed (p-value = 0.00876)

CNV Status	SCZ	LBC
Deletion Carrier (Cluster B + C)	34 (45.9%)	19 (24.4%)
No Deletion (Cluster A)	40 (54.1%)	59 (75.6%)
Total	74	78

**uncertainty values (0-1) were generated for each data point by the model-based hierarchical clustering algorithm in R Mclust package*

.

Table 5.2 *CHL1* 5' CNV distribution in an extended case-control cohort determined by PCR genotyping. (a) Results for 376 DNA samples from the Schizophrenia Cohort. **(b)** Results for 354 DNA samples from the Lothian Birth Control Cohort. No association of the CNV with disease was detected (p-value=0.984).

a

SCZ Cohort	Plate DN30417	Plate DN30418	Plate DN30419	Plate DN30420	SUMMARY CNV Status	No. of samples	% of Samples
Homozygous Deletion	5	2	3	2	Homozygous Deletion	12	3.19%
Heterozygous Deletion	30	33	26	36	Heterozygous Deletion	125	33.24%
Normal (No Deletion)	59	59	64	57	Normal (No Deletion)	239	63.56%
Total	**94**	**94**	**93**	**95**		**376**	**100%**

b

LBC Cohort	Plate DN30396	Plate DN30398	Plate DN30399	Plate DN30400	SUMMARY CNV Status	No. of samples	% of Samples
Homozygous Deletion	1	5	4	1	Homozygous Deletion	11	3.11%
Heterozygous Deletion	31	35	20	35	Heterozygous Deletion	121	34.18%
Normal (No Deletion)	56	51	64	51	Normal (No Deletion)	222	62.71%
Total	**88**	**91**	**88**	**87**		**354**	**100%**

5.2 Candidate II: *CHRFAM7A* (CHRNA7-FAM7A fusion gene) at 15q13

The second candidate genotyped in the replication cohort was a deletion polymorphism at *CHRFAM7A*. The fusion gene *CHRNA7-FAM7A* (*CHRFAM7A*) and its counterpart alpha7 neuronal nicotinic acetylcholine receptor gene (*CHRNA7*) reside at the 15q13-14 locus, a region that has been widely studied in the genetics of schizophrenia. Linkage analysis and association studies have implicated this region in schizophrenia and its endophenotypes (see Table 5.4), and in a number of neurological conditions including bipolar disorder (Edenberg et al. 1997), epilepsy (Elmslie et al. 1997) and Alzheimer's Disease (Carson et al. 2008).

5.2.1 Genomic Architecture at chromosome 15q13-14

The 15q13-14 locus has a complex genomic structure with segmental duplications and an evolutionary history of structural rearrangements (Riley et al. 2002; Sharp et al. 2008) (Figure 5.7). *CHRNA7* and *CHRFAM7A* are ~1.6 Mb apart in this locus. *CHRFAM7A* is part of the segmental duplication BP4 (breakpoint 4 of the Prader-Willli Syndrome), which has three large regions showing high homology (≥95%) with BP5 (breakpoint 5) (Figure 5.7 panel 4), the downstream segmental duplication immediately adjacent to *CHRNA7* (Sharp et al. 2008). *CHRFAM7A* was partially derived from *CHRNA7* and partially from *FAM7A*, a novel gene which was known to have duplicated to several locations on chromosome 15 (Riley et al. 2002). The last 5 exons of *CHRNA7* and *CHRFAM7A* share >99% homology (Gault et al. 1998). The formation of this hybrid gene *CHRFAM7A* was suggested to be a recent event specific to the human evolutionary history (Riley et al. 2002; Taske et al. 2002), since *CHRFAM7A* is absent in other primates such as the chimpanzee (Locke et al. 2003).

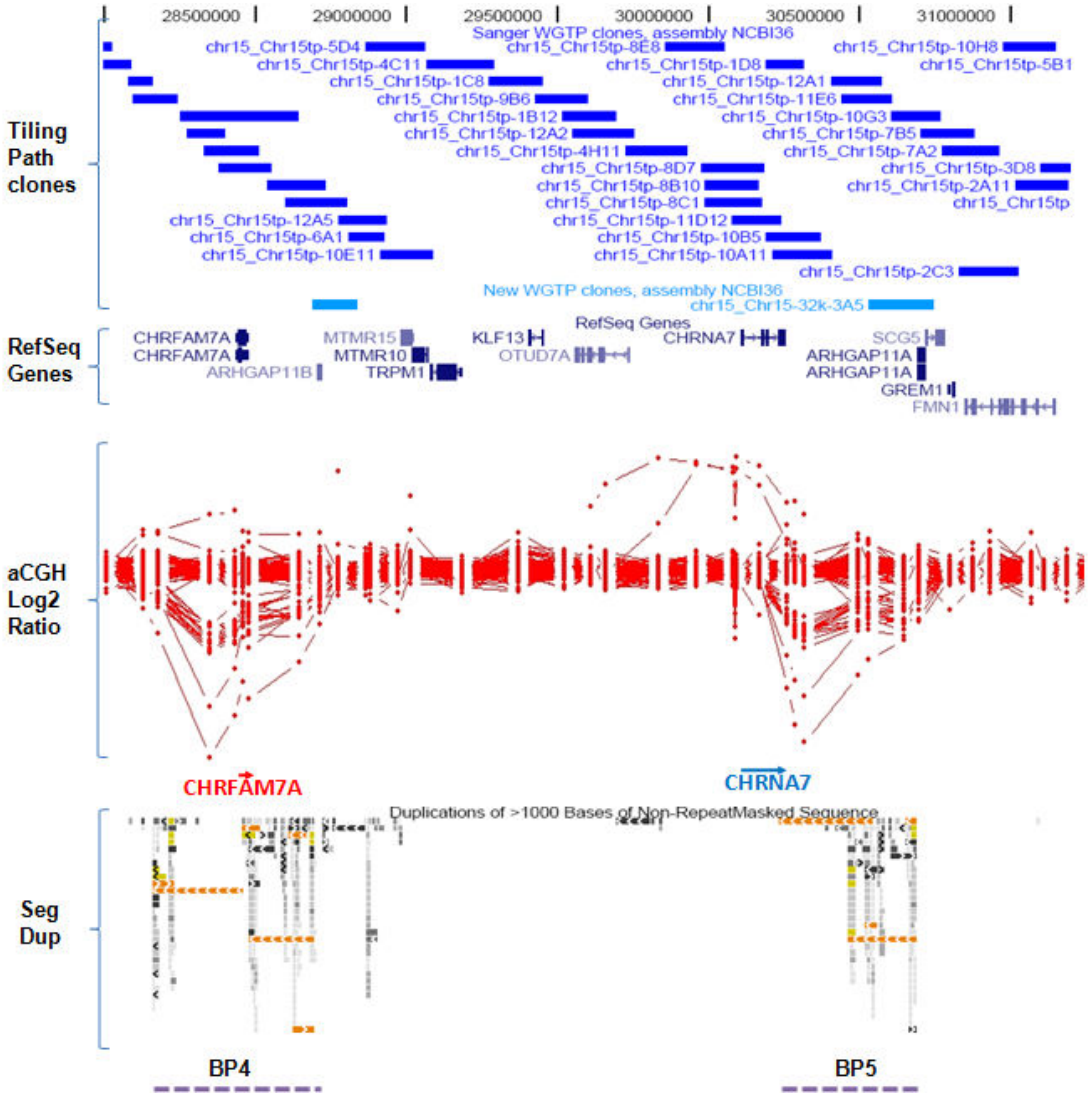


Figure 5.7 The 15q13-14 genomic locus. *(from top to bottom)* WGTP array BAC clones (blue blocks) (panel 1); RefSeq genes (panel 2); Array CGH data for 91 schizophrenia patients (Each red line represents the genomic profile of one patient) (panel 3); segmental duplication structures aligned to genomic locus (panel 4). The location of BP4 and BP5 (breakpoint 4 and breakpoint 5 of the Prader Willi Syndrome) are shown as purple dotted lines. They represent a pair of segmental duplication blocks with high homology to each other.

CHRNA7 and *CHRFAM7A* are shown in opposite orientations in the public DNA sequence database (NCBI Build 36), suggesting an inversion during the emergence of *CHRFAM7A* in the human lineage. Flomen *et al.* presented a model of the evolution of *CHRFAM7A* as a series of deletions, inversions and duplications, with postulated ancestral and intermediate structures (Flomen et al. 2008). It was also proposed that the *CHRFAM7A* exists as polymorphic inversion structures in the current population (consistent with previous findings from Riley et. al), and that the *CHRFAM7A* null allele represents the persisting ancestral chromosome. The 3 known allele variants of *CHRFAM7A* in humans are depicted in Figure 5.8.

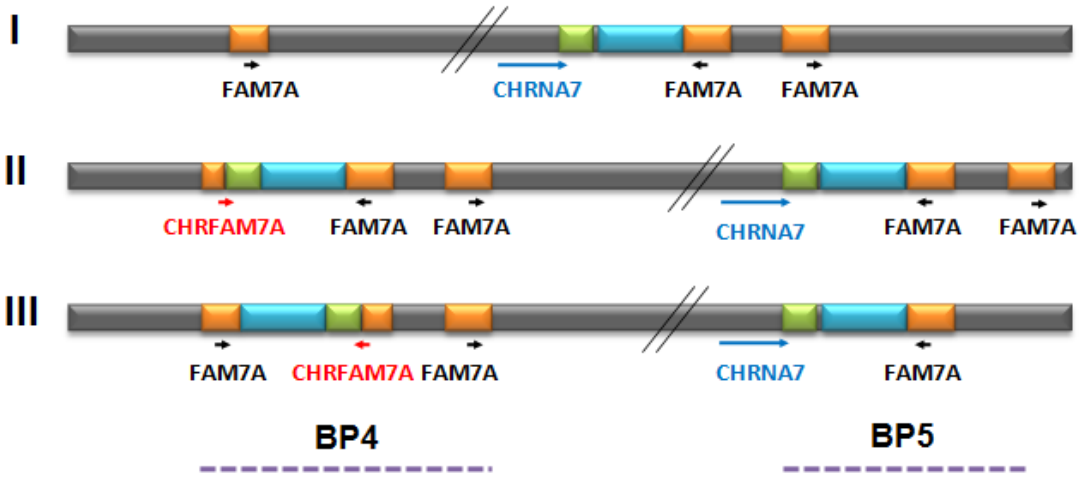


Figure 5.8 The three existing polymorphic structures of *CHRFAM7A* allele at 15q13-14. Structures as proposed by Flomen *et al.:* The first structure (I) is a null allele with no *CHRFAM7A*, possibly representing the ancestral allele which has persisted in the current population. The second structure (II) was formed subsequent to the acquisition of the hybrid gene *CHRFAM7A*. *CHRFAM7A* and *CHRNA7* are in the same orientation. The third structure (III) represents the inverted allele with respect to structure II, with *CHRNA7* and *CHRFAM7A* in opposite orientation. This is the allele as shown in the current genome build (NCBI36).

Structural rearrangements at 15q13-14 have been linked to several clinical phenotypes. Sharp *et al.* identified the 1.5 Mb recurrent deletion from BP4 to BP5, together with an overlapping 3Mb deletion, in patients with mental retardation and epilepsy (Sharp et al. 2008). Two recent genome-wide investigations of copy number variations in schizophrenia also identified the 15q13.3 recurrent deletions as a susceptibility locus (see section 1.6) (ISC 2008; Stefansson et al. 2008). With regard to the polymorphic *CHRFAM7A* gene, copy number of this gene was investigated in a combined cohort of patients (n=453) diagnosed with schizophrenia, bipolar affective disorder, schizoaffective or other psychoses (Flomen et al. 2006). When all disorders were considered as a single psychotic phenotype, the null *CHRFAM7A* allele was significantly associated with disease (P = 0.04) (association was not significant when the psychoses were analyzed separately). Furthermore, a 2-bp deletion in exon 6 of *CHRNA7*, which has recently been demonstrated as a surrogate marker of the inversion polymorphism of the upstream *CHRFAM7A* gene (Flomen et al. 2008), has previously been associated with bipolar disorder and major depressive disorder (Lai et al. 2001a; Lai et al. 2001b; Hong et al. 2004). In two other independent association analyses, the same 2 bp-deletion was involved in P50 sensory gating (Raux et al. 2002) and episodic memory performance (Dempster et al. 2006), both endophenotypes of schizophrenia, although association was not detected in another study with schizophrenia and other psychoses (Flomen et al. 2006). Combined evidence from these studies indicates that structural polymorphisms at chromosome 15q13-14 are likely candidates for a number of psychiatric disorders.

5.2.2 Molecular Genetic Studies linking 15q13-14 to Schizophrenia

In 1997, Freedman *et al.* first reported genetic linkage of the locus to a psychiatric phenotype (Freedman et al. 1997). In the study with nine multiplex families, a microsatellite marker in intron 2 of *CHRNA7* (D15S1360) was found to co-segregate with auditory evoked potential deficit (P50), an endophenotype related to sensory gating frequently observed in schizophrenia patients. The linkage signal to P50 was strong (LOD score of 5.9), but the signal was much weaker (LOD score decrease to 1.33) when the linkage was analyzed with the disease itself. The same microsatellite marker was then genotyped in other schizophrenia patient cohorts and the positive linkage was replicated by some (Leonard et al. 1998; Riley et al. 2000; Liu et al. 2001) but not in all studies (see Table 5.3).

Subsequently a number of linkage and association studies were performed (Table 5.3), some demonstrating positive linkage or association of the region (and of *CHRNA7* or *CHRFAM7A* polymorphisms) to schizophrenia, periodic catatonia, impaired auditory gating deficit, and/or other psychiatric disorders. However, inconsistent linkage results and lack of associations were also reported. In particular, in a recent large-scale association studies by Sanders *et. al* (1,870 cases and 2002 controls) (Sanders et al. 2008), as well as another comprehensive association study with linkage disequilibrium analysis (Iwata et al. 2007), both groups specifically investigated the *CHRNA7* gene/region, and found no association of any SNP or haplotype block with schizophrenia.

In summary, reports on genetic linkage analysis and association studies in the past decade have largely demonstrated positive results. Together with the recent compelling evidence of genomic rearrangement of this region in schizophrenia patients, the 15q13-14 locus is strongly suggested as a plausible candidate for schizophrenia pathogenesis.

Nevertheless, the specific mutation or polymorphism functionally involved in disease etiology and pathway (as opposed to markers or genomic regions) has yet to be identified.

Table 5.3 Previous reports associating the 15q13-14 locus to schizophrenia and related psychiatric disorders.

Author	Year	Study Type	Genes Under Investigation	Disease/ Phenotype	Sample Size	Result
Freedman	1997	Linkage	CHRNA7	SCZ (with P50 sensory gating deficit)	9 multiplex families	Microsatellite Marker D15S1360 in intron2 of CHRNA7 cosegregates with P50 sesory gating deficity with a LOD= 5.3. A smaller LOD score (Z=1.3) with disease
Leonard	1998	Linkage	CHRNA7	SCZ	20 families	Significant linkage at marker D15S1360 (P < 0.0024)
Neves-Pereira	1998	Linkage	CHRNA7	SCZ, SCZ AFF and others	5 multiplex families	No evidence of linkage Did not replicate findings previous postive finding with Did not replicate findings with D15S1360
Curtis	1999	Linkage	CHRNA7	SCZ	54 families	No evidence of linkage with 3 markers D15S1360, L76630, and ACTC Did not replicate previous positive findings
Riley	2000	Linkage	CHRNA7	SCZ	20 families	Microsatellite markers spanning 15q were genotyped Positive nonparametric linkage signals dectected at D15S1360,D15S1010 and D15S1043 Significant transmission disequilibrium detected at D15S1360
Stober	2000	Linkage	CHRNA7	SCZ with periodic catatonia	12 multiplex families	Evidence of linkage at the chromosome 15q candidate region (LOD score 2.75-2.89)
Stassen	2000	Association	CHRNA7	SCZ, BP and SCZ AFF	127 cases, 129 controls	Confirmed association between marker D15S1360 and schizophrenia. Also significant association with marker L76630 (P<0.0001). No significant difference between smokers and non-smokers
Gejman	2001	Linkage	CHRNA7	SCZ	68 families	20 microsatellite markers spanning chromosome 15 detected linkage at 15q11.2-q25. Single markers D15S659, ACTC, D15S150 and D15S1039 were implicated
Liu	2001	Linkage	CHRNA7	SCZ	52 families	6 microstallete markers at 15q11-14 genotyped. Positive linkage detected at D15S976 and D15S1360
Freedman	2001	Linkage	CHRNA7	SCZ (with P50 sensory gating deficit)	61 parent-child triads in 50 families	Significant linkage disequilibrium (P < 0.007) found at D15S165 within 1Mb of CHRNA7A and CHRFAM7A
Xu	2001	Linkage	CHRNA7	SCZ	31 families or trios	3 microsatellite markers genoypted. D15S165 and L76630 shows evidence of linkage
Tsuang	2001	Linkage	CHRNA7	SCZ, SCZ AFF	392 cases	Linkage detected between D15S165 and D15S1010
Meyer	2002	Linkage	CHRNA7	SCZ with periodic catatonia	1 multiplex family	Evidence strongly support linkage at 15q13-22 in a multiplex family, but gene CHRNA7 was excluded
Gault	2003	Mutation Screening & Association	Region surrounding CHRNA7 & CHRFAM	SCZ	171 cases (families), 185 control	33 polymorphisms were found in patients and controls. 21 variants in exons, but non-synonymous changes were rare
Iwata	2007	LD Analysis & Association Study	CHRNA7 and CHRFAM7A	SCZ	188 cases (188 controls)	14 microsatellite markers geneotyped, one deviate from HWE in patients but not in control. Overall no linkage detected

Shi J.	2007	Family Based Association	19 cholinergic genes	BP and SCZ AFF (BP type)	I. 474 cases (152 families) II. 83 cases (22 multiplex families)	Some SNPs showed nominally significant transmission bias, but none reach genome-wide significance after correction by multiple testing
Sanders	2008	Association	CHRNA7 (and 14 other SCZ genes)	SCZ	1,870 cases (2002 controls)	No significant association with any of the 14 genes using tag SNPs and SNPs previously reported to be associated with SCZ
Association Study of the Promoter Region of CHRNA7						
Leonard	2002	Mutation Screening & Association	Promoter Region of CHRNA7	SCZ	166 cases and 165 controls	Evidence of linkage of functional promoter variants with SCZ ((trasncription analysis subsequent to genoytping)
Houy	2004	Mutation Screening & Association	Promoter Region of CHRNA7	P50 sensory gating deficit	111 cases, 85 controls	Significant association of the -194 C allele (T/C ratio <0.45) with P50 but not SCZ. Did not replicate previous result of core promoter polymorphism from Leonard et. al.
Li	2004	Mutation Screening & Association	Promoter Region of CHRNA7	SCZ	249 cases, 273 controls	New polymorphisms were identified at the promoter region but no association detected
Association Study of the 2-bp deleiton in exon 6 of CHRNA7						
Lai	2001	Association	CHRNA7 and CHRFAM7A	Major Depressive Disorder	72 cases, 103 controls	Mildly significant association of the 2-bp deletion genotypes (P=0.027) and alleles (P=0.037) with major depressive disorder
Lai	2001	Association	CHRNA7 and CHRFAM7A	SCZ	146 cases, 151 controls	No significant association of 2-bp deletion with disease
Hong	2004	Association	CHRNA7 and CHRFAM7A	BP	77 cases, 135 controls	Mildly significant association of 2-bp deletion with BP (P=0.044)
Raux	2002	Association	CHRNA7 and CHRFAM7A	SCZ & P50 sensory gating deficit	70 cases, 77 controls	2bp deletion is not associated with SCZ, but is associated with the P50 gating deficit
Flomen	2006	Association	CHRNA7 and CHRFAM7A	SCZ, BP and SCZ AFF	453 cases, 197 controls	Mildly significant association (P = 0.04) of CHRFAM deletion genotype with disease, not significant after multiple testing. No association of the 2-bp deletion at exon 6 of CHRN7A w ith disease
Dempster	2006	Association	CHRNA7 and CHRFAM7A	SCZ with Episodic Memory Performance	251 cases	significant association of the 2bp deletion with episodic memory performance in SCZ patients
Association Study of Nicotinic Receptor Genes with Smoking in Schizophrenia						
De Luca	2004	Association	CHRNA7	SCZ with smoking	177 patients (108 smokers, 69 non smokers)	Marker D15S1360 associated with smoking in schizophrenia patients (p=0.015)
Faraone	2004	Association	CHRNA7 and 15 nicotinic acetylcholine receptor genes	SCZ with smoking		Significant linkage of the group of genes to SCZ (p = 0.039). CHRN7A alone shows marginally significant association (p = 0.095)

5.2.3 Functional Significance of *CHRNA7* and *CHRFAM7A*

As with other nicotinic acetylcholine receptors (nAChRs), *CHRNA7* is a ligand-gated ion channel that binds nicotine with low affinity. The receptor admits calcium ions into cells and mediates fast signal transmission at synapses. nACHRs are thought to assemble with other subunits to form a pentameric holoreceptor (Lindstrom 1996), although *CHRNA7* was demonstrated to form homo-oligomeric channel receptors in vitro (Couturier et al. 1990; Leonard and Bertrand 2001). *CHRNA7* plays important roles in development and in synaptic plasticity (Dani and Bertrand 2007).

Recent biochemical studies have linked the alpha7 receptor to pathways and molecules with high relevance to schizophrenia. Hancock *et al.* established the role of neuregulin-Erb signalling in targeting *CHRNA7* for axonal surface expression (Hancock et al. 2008). In a study of the CA1 region of hippocampus, the brain-derived neurotrophic factor (*BDNF*), another well-accepted schizophrenia candidate, was shown to decrease the nicotinic receptor's response in interneurons (Fernandes et al. 2008). Furthermore, antisense knockdown of *CHRNA7* in the rat demonstrated deficits in learning and memory (Curzon et al. 2006), while overexpression of the receptor in mice was suggested to improve spatial memory-related performance (Ren et al. 2007). Impaired attention (Young et al. 2007) and sensory gating functions (Young et al. 2007) were also reported in mouse models.

The function of the hybrid gene *CHRFAM7A* is not entirely understood. The gene is transcribed and mRNA detected in various tissues including the brain (Gault et al. 1998), although *CHRFAM7A* is expressed at a lower level in the brain compared to that of *CHRNA7* (Agulhon et al. 1999; De Luca et al. 2006). De Luca *et al.* also compared postmortem expression of the *CHRNA7* and *CHRFAM7*A in schizophrenia, bipolar and

controls and found a decreased *CHRNA7*:*CHRFAM7A* mRNA ratio in bipolar disorder (but not in schizophrenia) compared to controls (De Luca et al. 2006). It had been proposed that *CHRFAM7A* could affect the oligomerization of alpha-7 receptor channel, or have other gain of function or dominant negative effect on *CHRNA7* (Raux et al. 2002).

5.2.4 Copy Number Polymorphism at the *CHRFAM7A* Region

A copy number polymorphism spanning the gene *CHRFAM7A* was detected via the log2 ratio distributions of a series of clones at 15q13.2-13.3. The copy number region covers the location of the segmental duplication blocks breakpoint 4 (BP4), which has high homology to the more distal breakpoint BP5 (refer to section 5.2.1).

Applying bivariate-clustering algorithm to *Chr15tp-2G5* and *Chr15tp-2G2,* the two BAC clones that span the gene *CHRFAM7A*, two major clusters were observed (Figure 5.9). Since previous literature indicates the presence of a null *CHRFAM7A* allele in the current population (Flomen et al. 2006), we postulated that the upper cluster (cluster A) represents samples with two copies of the "normal" allele each having one copy of *CHRFAM7A*, while the lower cluster (cluster B) represents samples carrying deletion of the *CHRFAM7A* allele (the "ancestral" allele).

We assigned each of the 91 SCZ DNA samples and 92 LBC DNA samples from the array CGH screen to one of the two clusters (Figure 5.10). The ratio of samples in cluster A: cluster B in the schizophrenia cohort was 70:19 (normal:deletion) (n=89, with two samples in between clusters removed due to high uncertainty), compared to the ratio of 82:10 (normal:deletion) in the control. The enrichment of cluster B samples in the schizophrenia cohort indicates an excess of null (deletion) *CHRFAM7A* allele in the

disease population (p-value = 0.054682). We therefore genotyped the deletion polymorphism in an extended cohort of 300+ patients with equivalent number of matched controls.

With Taqman qPCR primers and probes spanning the breakpoint betjween *FAM7A* exons (D' to A) and the *CHRNA7* exons (5-10), we were able to genotype the copy number of the hybrid gene *CHRFAM7A* independent of the copy number of *CHRNA7*. The primers and probes were first tested on Hapmap individuals with known log2 ratio, and in each sample qPCR fold change correlated well with fold change estimated from array CGH log2 ratio (data not shown).

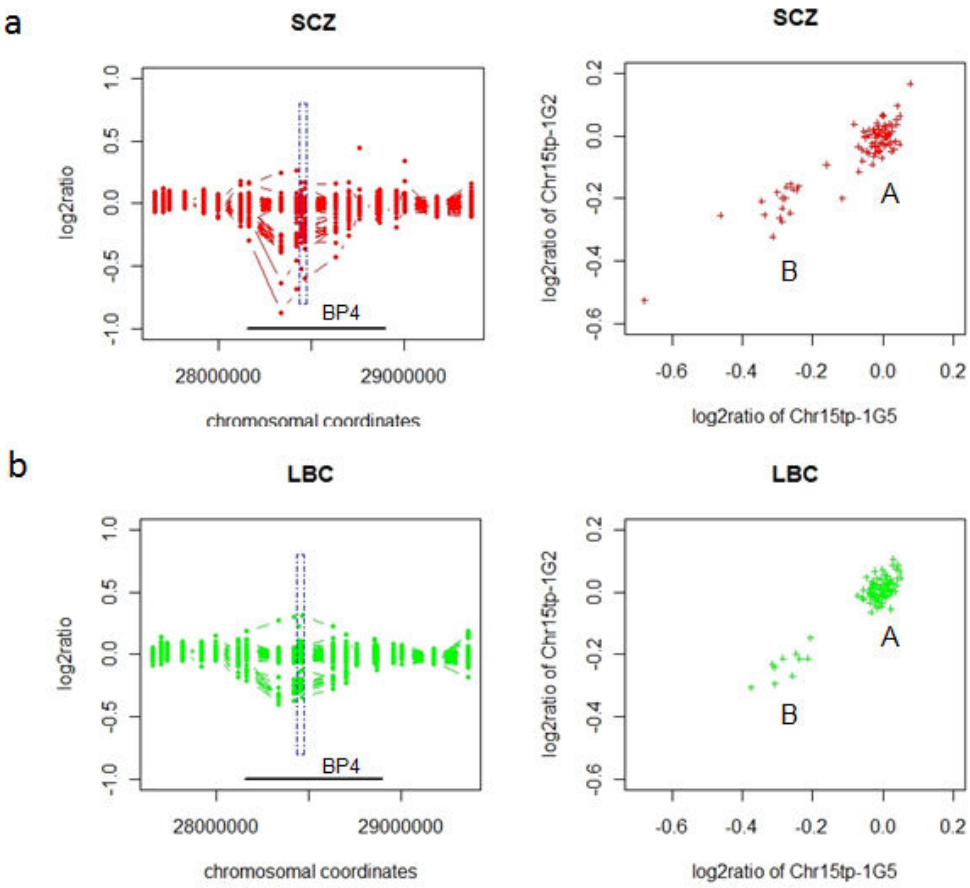
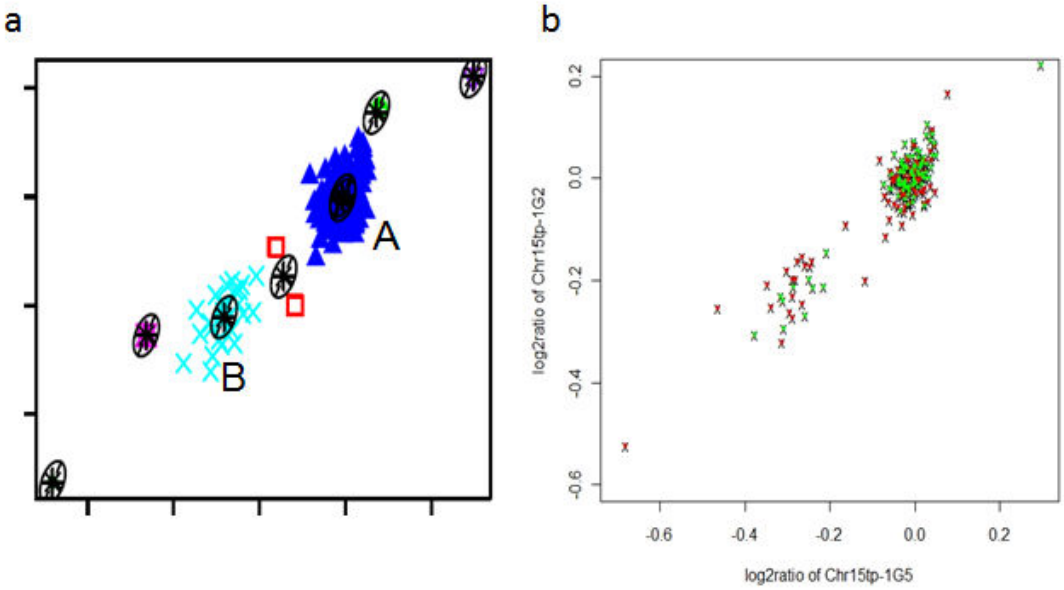


Figure 5.9 Copy number polymorphism at 15q13.2-13.3 spanning the gene *CHRFAM7A*. Array CGH data for **a)** SCZ cohort (n=91) & **b)** LBC cohort (n=92). On the left panel log2ratio was plotted against the genomic position. Location of *CHRFAM7A* was highlighted with the blue box. BP4 (breakpoint 4 of the Prader-Willi Syndrome) was shown as black line. On the right panel log2ratio of two clones spanning *CHRFAM7A* (Chr15tp-1G2 and Chr15tp-1G5) were plotted against each other. Two major clusters were detected in both SCZ and LBC cohorts.



c

	Cluster A ("Normal")	Cluster B ("Deletion")	Total
SCZ	70	19	89
LBC	82	10	92
	Chisq.test	P-value	0.054682

Figure 5.10 *CHRFAM7A* CNV genotype distributions in cases versus controls. (a) CNV genotypes were estimated from array CGH log2ratio by bivariate clustering of BAC clones *Chr15tp-1G2* and *Chr15tp-1G5*. **b)** Assigning SCZ (red) and LBC (green) samples to the two major genotype clusters. **c)** Sample counts and chi-square test statistics* show a mildly significant association of deletion carriers with disease (p=0.05).

**uncertainty values (0-1) were generated for each data point by the model-based hierarchical clustering algorithm in R Mclust package. Two samples with high uncertainty values (between cluster A and B) were removed before statistical analysis.*

Next we genotyped the *CHRFAM7A* CNV in an extended replication cohorts using qPCR. Results were obtained for 262 schizophrenia patients and 270 matched controls (each with 2 replications). A typical qPCR experiment was shown in Figure 5.11 (see Appendix D-4 for the complete set of qPCR results). Table 5.4 summarizes the genotype frequency distributions in cases and controls.

Our genotyping attempt in an extended cohort suggested that CNV genotypes at *CHRFAM7A* have no statistically significant association with disease (by genotype frequency: chi-square test, p-value = 0.693; by allele frequency: chi-square test, p-value=0.672), although by direct observation there seem to be a slight trend of increased deletion allele in the disease cohort.

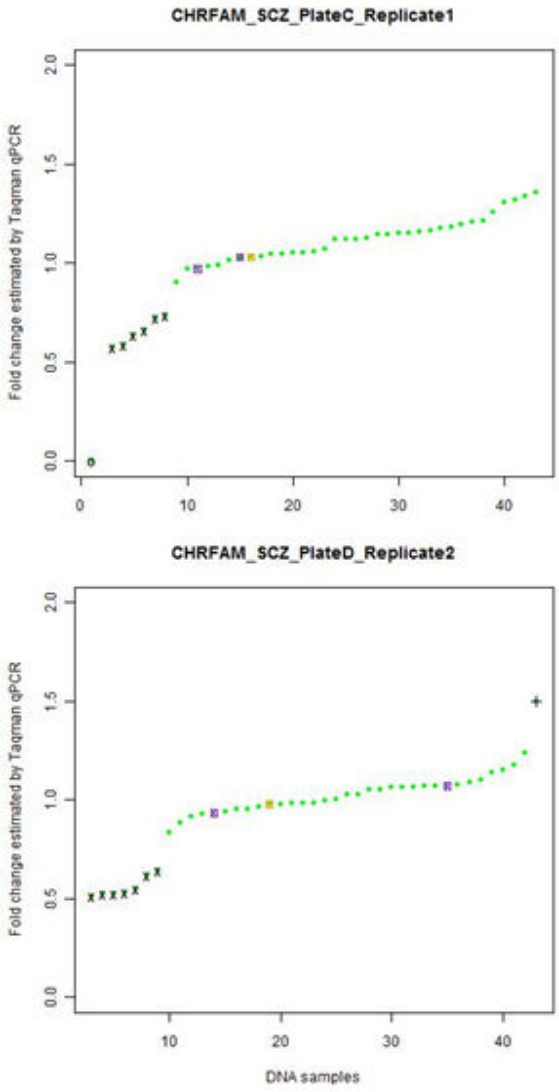


Figure 5.11 Analysis of a typical Taqman qPCR experiment to determine *CHRFAM7A* genotypes. In each 96-well plate 40 samples (+2 controls) were genotyped. Fold change of all samples in the same experiment (sorted in ascending order) were plotted on the same graph (green dots) and CNV genotypes were estimated by the change of the slope of the graph.

o: Homozygous deletion samples; x: heterozygous deletion samples; +: duplication samples : control sample NA10851; : control sample NA12776

Table 5.4 *CHRFAM7A* CNV genotype distributions in the extended case and control cohorts. qPCR genotyping results for **(a)** SCZ replication cohort (n=270) (**b)** LBC replication cohort (n=262). Chi-square test of the CNV genotype distributions suggests no statistical significant difference between the two cohorts (p-value=0.0693).

a

	Plate	Homodel	Hetdel	Normal	Duplication	Total
1	SCZ_PlateA	0	7	32	1	40
2	SCZ_PlateB	0	6	32	0	38
3	SCZ_PlateC	1	6	30	0	37
4	SCZ_PlateD	0	7	29	1	37
5	SCZ_PlateE	3	4	31	2	40
6	SCZ_PlateF	0	7	32	0	39
7	SCZ_PlateH	1	4	34	4	39
	No. of Samples	**5**	**41**	**220**	**8**	**270**
	% of Samples	**1.85%**	**15.19%**	**81.48%**	**2.96%**	

b

	Plate	Homodel	Hetdel	Normal	Duplication	Total
1	LBC_PlateA	0	4	34	1	39
2	LBC_PlateB	0	7	22	1	30
3	LBC_PlateC	0	8	29	0	37
4	LBC_PlateE	0	5	32	2	39
5	LBC_PlateF	2	6	30	1	39
6	LBC_PlateG	0	4	33	1	38
7	LBC_PlateH	0	4	36	0	40
	No. of Samples	**2**	**38**	**216**	**6**	**262**
	% of Samples	**0.76%**	**14.50%**	**82.44%**	**2.29%**	

5.3 Chapter Summary and Discussion

This chapter describes targeted genotyping and characterization of two schizophrenia candidate copy number polymorphisms, and an attempt to replicate findings from our initial array CGH screen in an extended case-control cohort.

The *CHL1* 5' deletion polymorphism at 3p26 was fully characterized by a combination of CGH data clustering and PCR strategies, together with published gene expression and high-resolution CNV data. We demonstrated that the small (~5 kb), frequent deletion located 5' upstream of *CHL1* could partially explain expression level of the gene. The deletion was significantly associated with schizophrenia in our initial screen on 91 patients and 92 controls, but the enrichment was not observed in the extended cohort of 300+ patients and matched controls.

The second candidate, a copy number polymorphism at *CHRFAM7A* at 15q13, demonstrated recurrent deletion and duplication in the population. The highly complex genomic architecture with segmental duplication and inversion made it difficult to devise a simple PCR assay for CNV genotyping. A Taqman qPCR strategy with specific primers and probes for *CHRFAM7A* was used to determine gene dosage. Our initial whole-genome CNV screen demonstrated an enrichment of the deletion phenotypes. In an extended replication cohort, qPCR genotypes for *CHRFAM7A* were obtained for 250+ cases and matched controls. No significant association was detected between CNV genotypes and disease, although there was slight bias of deletion genotypes by direct observation.

Neither CNV polymorphism showed significant association in the extended cohort of ~300 cases and ~300 controls. We postulate the following to explain our inability to replicate the initial associations:

First, our initial CNV screen was modest in sample size, and could result in statistical bias leading to false-positive associations. It is worth mentioning that we see a high plate-to-plate variation (irrespective of disease status) in our subsequent PCR genotyping of the *CHL1* 5' CNV (see Table 5.2), indicating the relevance of random population sampling. This emphasizes the need for replication in CNV association studies, for instance a two-stage approach of polymorphism identification with subsequent replication, as in most SNP association studies. Replication is particular important for studies with small initial sample size (e.g. for cost-effective genome-wide CNV discoveries as in our study), and is pertinent to detecting associations of frequent copy number polymorphisms, as well as the relatively rare but recurrent copy number changes.

Alternatively, the common polymorphism candidates we targeted could have real impact on disease risks but with small effect sizes. The lack of association in subsequent replication attempts may point to the need for yet larger samples to statistically detect such association with a small odds ratio. This relates to the common-disease common-variant (CD-CV) hypothesis (Collins et al. 1997), which suggests that common disease traits like schizophrenia could be the consequence of common allelic variants with small to modest effects, typically under less stringent selection pressure (Carlson et al. 2004; Zondervan and Cardon 2004). Consistent with such an hypothesis, in the case of *CHRFAM7A* CNV, we see a non-statistically significant bias of patient samples towards

deletion genotypes, which could represent a small increased disease risk as conferred by the deletion allele.

Finally, we cannot rule out non-sampling-related artifacts in the original CNV screen and analysis. For example, in the initial screen, accuracy of copy number quantification is compromised by converting continuous measurements (fluorescent intensity-based data) into discrete CNV genotypes (McCarroll and Altshuler 2007). In addition, variability in CNV measurements between cohorts, e.g. due to DNA quality, as implicated in our study, could have a potential impact on genotype classification (Barnes et al. 2008). Knowing the true underlying genotype is important to validate association signals (Todd 2006). To this end, one could use targeted quantification methods such as Taqman qPCR to deliver a more acceptable accuracy level in larger-scale genotyping, a strategy we employed. Recently, more robust, high-throughput and low-cost methods have been developed and proved valuable in disease association studies, even for multi-allelic loci with a wide range of copy numbers (see section 1.2 for discussion of such methods). Improved statistical methods have also been applied to intensity-based data for more accurate CNV genotype classifications (Barnes et al. 2008).

Chapter 6
CNVs and the
NMDA Receptor Complex

6.1 NMDA Receptor Complex, Schizophrenia and Cognition

Molecules at the synapse, the specialized junction between neurons, play a fundamental role in learning and memory, neural development and neural degeneration (Nakanishi et al. 1998; Kandel 2000). Alterations to these synaptic proteins can result in different human cognitive and behavioral phenotypes, including a number of neuropsychiatric disorders (Laumonnier et al. 2007; Ramocki and Zoghbi 2008; Sudhof 2008). Within the synapse proteome, the NMDA Receptor Complex is of particular interest due to its central role in glutamatergic signalling for synaptic transmission in the brain. The NMDA receptor complex (NRC), also known as MAGUK (membrane-associated guanylate kinase) Associated Signalling Complex (MASC), encompasses a key set of synaptic proteins in the postsynaptic density (PSD) (Husi et al. 2000; Husi and Grant 2001; Collins et al. 2006), centered around the NMDA receptor (*NMDAR,* N-Methyl D-Aspartate ionotropic receptor) (composed of NR1 and NR2 subunits) and associated MAGUK scaffolding proteins (e.g. *PSD-95*, *PSD-93*, *SAP-102* and *SAP-97*) (Figure 6.1).

The NRC/MASC complex comprises of 186 proteins (Collins et al. 2006) (Appendix F). A number of them (43 out of 186) have demonstrated involvement in synaptic plasticity (Grant et al. 2005), the changes in synaptic properties in response to neuronal stimulation, which form the basis of learning and memory formation. 54 of these synaptic proteins were associated with one or more nervous system disorders (Grant et al. 2005). Unraveling the functional complexities of these components is critical for our understanding of diseases, behaviors and the learning processes (Pocklington et al. 2006).

Pertinent to the investigation of psychiatric disorders, NRC/MASC components have provided support for the glutamate hypothesis of schizophrenia (section 1.5).

Hypofunction of the NR1 subunit of NMDAR in mice, for instance, displayed behavioral abnormalities mimicking schizophrenia (Mohn et al. 1999). Recent studies on copy number variations in schizophrenia further implicate the role of NRC/MASC proteins in schizophrenia. Candidate genes involved in schizophrenia CNV loci, for example *DLG2* (Walsh et al. 2008), *ERBB4* (Walsh et al. 2008), *NRXN1* (Rujescu et al. 2008), *CHRNA7* (ISC 2008; Stefansson et al. 2008) *and CYBIP1* (Stefansson et al. 2008) (section 1.6 and Figure 1.8), all converge to a direct or indirect involvement in the NRC/MASC complexes.

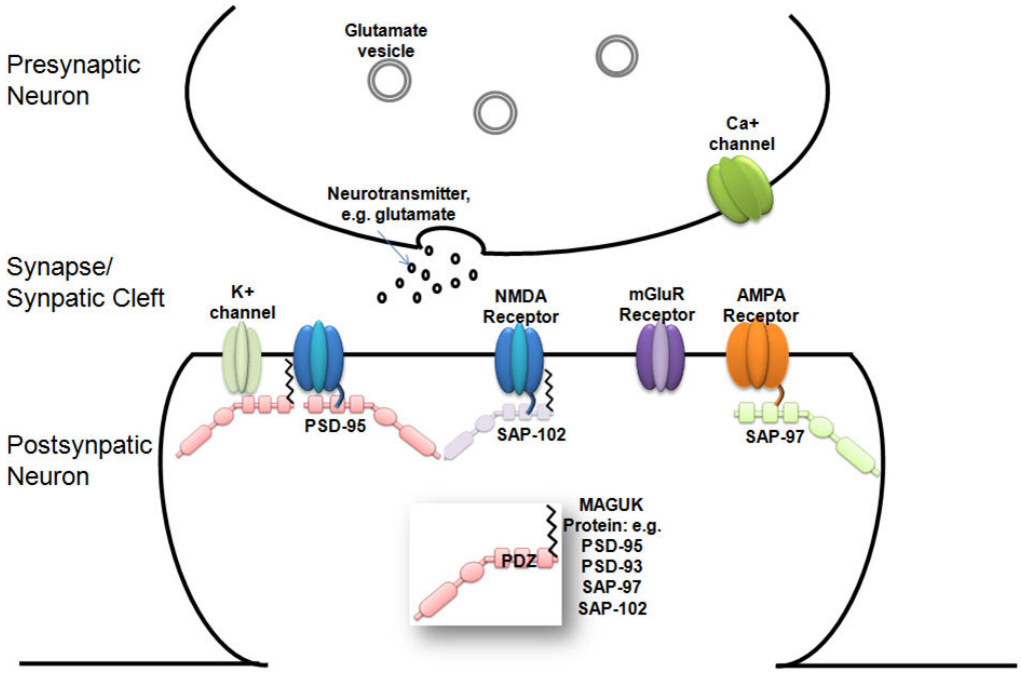


Figure 6.1 Schematic diagram of a glutamatergic excitatory synapse. The NMDA receptor (NMDAR) is an ionotropic glutamate receptor embedded in the postsynaptic membrane. Other glutamate receptors include the AMPA Receptor (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) and the mGluR receptor (metabotropic glutamate receptor). NMDAR and its associated proteins, such as the MAGUK proteins, were clustered at the postsynaptic neuron forming part of the postsynaptic density (PSD).

6.2 Copy Number Variation and the NMDA Receptor Complex

We assessed whether genes encoding NRC/MASC components are located at copy number variable loci. 186 NRC/MASC signalling complex components were mapped to the genomic coordinates of WGTP BAC clones for comparison against known CNVs in 269 normal HapMap individuals. 20 out of 186 (11%) MASC/NRC complex components were detected within CNV regions in one or more samples (herein refer as MASC CNVs) (Table 6.1, Figure 6.2).

The 20 MASC CNVs were characterized into 5 categories according to a previously defined classification system (Redon et al. 2006)[4]. Fifteen of them were classified as simple deletions, duplications or deletions with duplications. These CNVs were at relatively low frequency in the population (on average 1.4% of HapMap samples affected, corresponding to ~4 individuals out of 269).

The remaining MASC CNVs were complex or multiallelic, affecting the majority of the HapMap population. The most frequent variants were located near the genes *NSF* (N-ethylmaleamide-sensitive factor) (Figure 6.2q), *RPL13* (ribosomal protein L13) (Figure 6.2p) and *GRM5* (metatropic glutamate receptor 5) (Figure 6.2j). In particular, *NSF* was identified as a multi-allelic variant with at least 4 discrete CNV genotypes. Due to the interesting genomic architecture of this locus and the biological relevance of the *NSF* gene, we investigated the genomic region in more detail (see section 6.3).

[4] The classification system categorizes CNVs into 5 categories, namely: “deletions”, “duplications”, “deletions and duplications”, “multi-allelic” and “complex rearrangement”

Table 6.1 CNVs detected at 20 genes encoding NRC/MASC signalling complex components. The most frequent CNVs are highlighted in yellow, and the CNVs affecting core components of the MASC complex in blue.

Gene	Gene Description	Chr	Start	End	CNV Clones (% in HapMap)	Category
GNB1	guanine nucleotide binding protein (g protein), beta	1	1706590	1812355	Chr1tp-25A7(7.43%)	C
PKLR	pyruvate kinase, liver and rbc	1	153526254	153537843	Chr1tp-21C5(1.49%), Chr1tp-7H12(1.49%)	A
DLG1	**discs, large homolog 1**	**3**	**198255819**	**198509844**	**Chr3tp-20C11(13.38%)**	**E**
MOG	myelin oligodendrocyte glycoprotein	6	29732788	29748128	Chr6tp-22B5(0.37%), Chr6tp-1F1(0.37%)	A
YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	7	75794053	75826252	Chr7tp-9C8(3.72%)	C
HRAS	v-ha-ras harvey rat sarcoma viral oncogene homolog	11	522242	525591	Chr11tp-17H2(0.37%)	B
SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	11	780478	786221	Chr11tp-13G5(0.37%)	B
SLC1A2	solute carrier family 1 (glial high affinity glutamate transporter), member 2	11	35229329	35398186	Chr15tp-5C1(0.37%)	A
DLG2	**discs, large homolog 2, chapsyn-110**	**11**	**82843701**	**85015962**	**Chr11tp-13D9(0.37%), Chr11tp-18E5(0.37%)**	**A**
GRM5	**glutamate receptor, metabotropic 5**	**11**	**87881006**	**88438761**	**Chr11tp-5E4(4.83%), Chr11tp-11C4(13.01%), Chr11tp-3E11(13.38%)**	**E**
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	12	6513918	6517797	Chr12tp-17B1(0.37%)	A
PDPK1	3-phosphoinositide dependent protein kinase-1	16	2527971	2593190	Chr16tp-11D7(2.6%)	C
GRIN2A	**glutamate receptor, ionotropic, n-methyl d-aspartate 2a**	**16**	**9754762**	**10184112**	**Chr16tp-12H9(0.74%)**	**C**
MAPK3	mitogen-activated protein kinase 3	16	30032928	30042131	Chr16tp-3F5(3.35%)	E
GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	16	57298538	57325747	Chr16tp-12E8(1.49%), Chr16tp-3F2(1.49%)	B
RPL13	**ribosomal protein l13**	**16**	**88154595**	**88158451**	**Chr16tp-1H4(21.56%)**	**E**
NSF	**n-ethylmaleimide-sensitive factor**	**17**	**42023391**	**42189997**	**Chr17tp-2F6(64.31%), Chr17tp-2F9(56.51%)**	**D**
DLGAP1	**discs, large (drosophila) homolog-associated protein 1**	**18**	**3486030**	**3870135**	**Chr18tp-8F6(0.37%), Chr18tp-3A2(0.74%), Chr18tp-6E9(0.74%)**	**A**
AKT2	v-akt murine thymoma viral oncogene homolog 2	19	45428064	45483105	Chr19tp-6A8(0.37%), Chr19tp-4C5(0.37%)	A
PLCB1	phospholipase c, beta 1 (phosphoinositide-specific)	20	8061296	8813547	Chr20tp-5C4(0.37%), Chr20tp-6E3(0.37%), Chr20tp-7D8(0.37%), Chr20tp-4C4(0.37%), Chr20tp-4B2(0.37%), Chr20tp-6F2(0.37%), Chr20tp-3B9(0.37%), Chr20tp-6B1(0.37%)	A

* CNV categories: *A*: deletions, *B*: duplications, *C*: deletions and duplications, *D*: multi-allelic and *E*: complex re-arrangement.

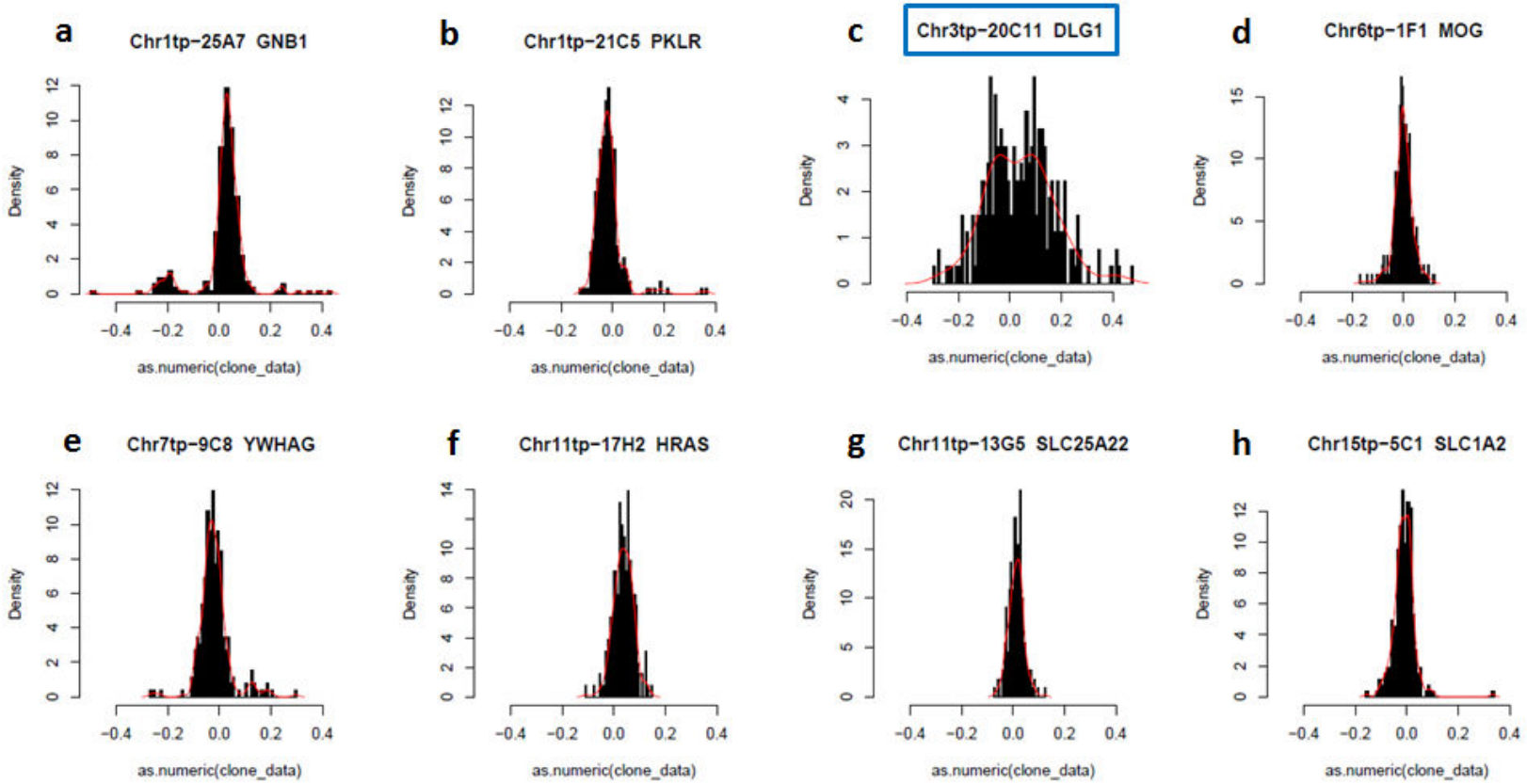


Figure 6.2 (to be continued)

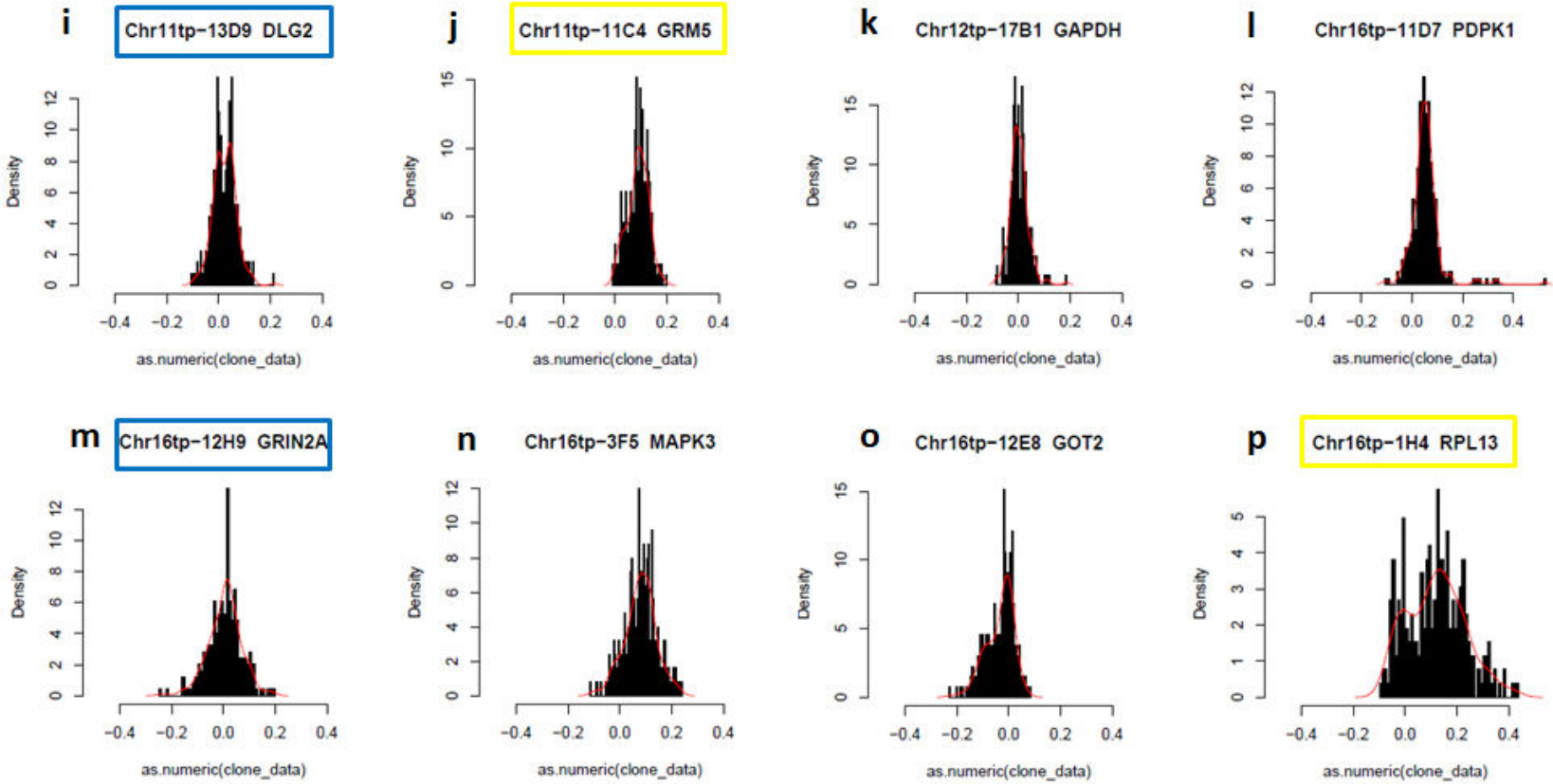


Figure 6.2 (to be continued)

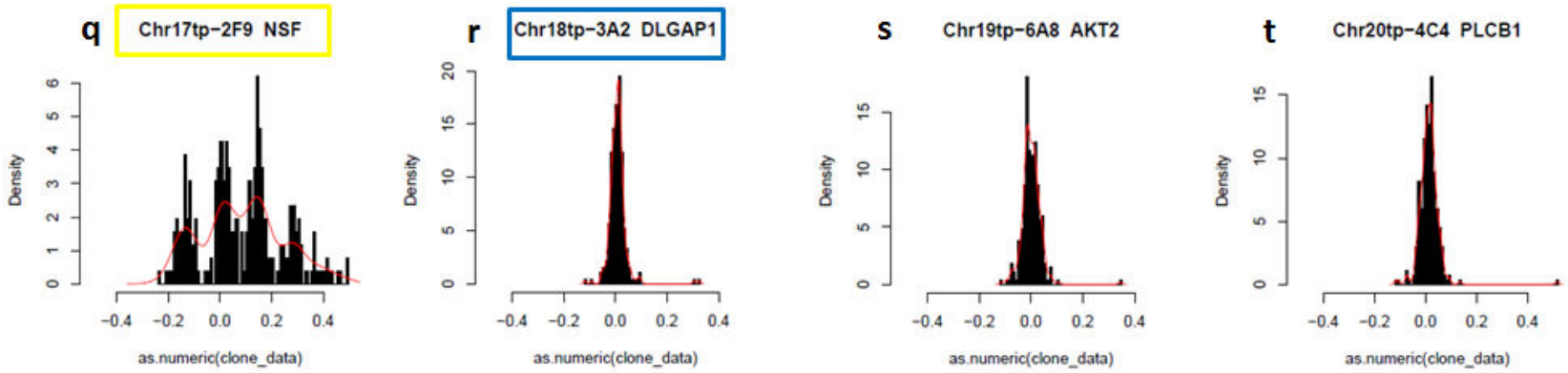


Figure 6.2 Log2ratio distributions for WGTP clones reporting CNV among 269 HapMap samples in 20 MASC regions. The most frequent CNVs were highlighted in yellow, and the CNVs affecting core components of the MASC complex in blue. *a: GNB1: ; b: PKLR; c: DLG1; d: MOG; e: YWHAG; f: HRAS; g: SLC25A22;h: SLC1A2; i: DLG2; j: GRM5; k: GAPDH; l: PDK1; m: GRIN2A; n: MAPK3; o: GOT2; p: RPL13; q: NSF; r: DLGAP1; s: AKT2; t: PLCB1*

The core NRC/MASC components, i.e. the NMDA receptor subunits and the closely associated MAGUK scaffolding proteins, were of particular interest for understanding cognition and diseases. Among the 20 MASC CNVs, 4 overlapped with the core NRC/MASC genes, detected as variants at *DLG1*, *DLG2*, *DLGAP1* and *GRIN2A* (Figure 6.3; Table 6.1 in yellow).

The CNV at *DLG1* (Discs, Large Homolog 1 (Drosophila)) showed a complex pattern, as displayed by a broad log2ratio distribution of clone *Chr3tp-20C11* (Figure 6.4a). At WGTP resolution, we were not able to resolve the precise CNV location and discrete copy number for individuals. The WGTP data was subsequently compared to the recently-released oligonucleotide array data (Genome Structural Variation Consortium 2008). We confirmed the presence of a frequent intronic CNV of ~5 kb within *DLG1*.

In addition, a duplication involving *DLG2* (Discs, Large Homolog 2 (Drosophila)) was detected in one individual in the WGTP platform. Oligo array data revealed additional CNVs in more individuals (deletions) (Figure 6.4b). *DLGAP1* (Discs Large-Associated Protein 1) was also identified within a larger duplication, affecting 3 consecutive BAC clones in the WGTP data (Figure 6.4c). A number of deletions were revealed by the oligo array data from the CNV Consortium.

Finally, *GRIN2A* (glutamate receptor, ionotropic, n-methyl d-aspartate 2a), encoding the NR2A subunit of the NMDA receptor, was detected with both duplications and deletions in the population by WGTP platform (Figure 6.4d), although no CNV was detected using higher-resolution oligonucleotide array.

In summary, copy number variants at multiple components of the NRC/MASC complex exist among normal, healthy individuals, despite the significance of these proteins in

normal brain functions such as synaptic plasticity and cognition. We further investigated the *NSF* CNV locus as it was the most frequent MASC CNV (section 6.3).

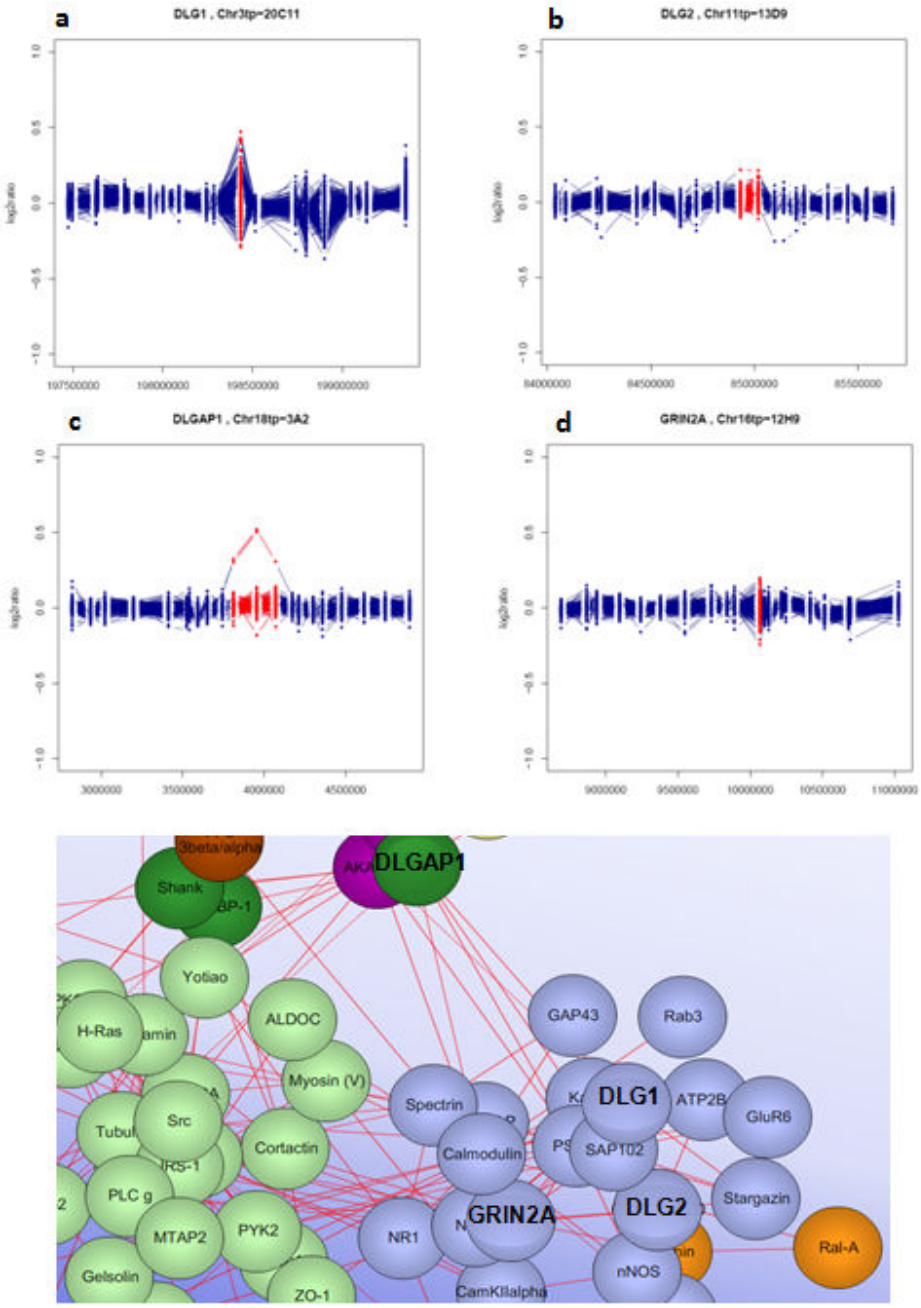


Figure 6.3 CNVs affecting core components of the NRC/MASC signalling complex. (a)-(d): WGTP array CGH profiles of the four CNVs in 269 HapMap individuals. (a) CNV at *DLG1* with complex pattern in the HapMap population b) Duplication at *DLG2* in one sample c) Large duplication at *DLGAP1* in 2 samples d) CNV at *GRIN2A* detected as both duplication and deletion (e) The four core NRC components within the network of NRC/MASC signalling complex (diagram modified from www.genes2cognition.com)

6.3 CNV at 17q21 near N-ethylmaleamide-Sensitive Factor (NSF)

6.3.1 Introduction to the 17q21 locus near *NSF*

We detected a multi-allelic variant near *NSF*, the most frequent CNV in the NRC/MASC CNV screen. The *NSF* CNV region is highly variable with a complex genomic architecture and association with mental illness, prompting us to further investigate the region. Structural variants in the region have been identified, including a well-studied common inversion of ~900 kb (Stefansson et al. 2005), resulting in two designated alleles H1 and H2. The H2 allele was suggested to be under positive selection. The inversion encompasses several genes involved in brain development including *NSF* (N-ethylmaleamide-sensitive factor, a membrane fusion protein), *CRHR1* (corticotrophin releasing hormone receptor), *IMP5* (intramembrane protease 5, a presenilin homologue), *MAPT* (microtubule associated protein tau) and *STH* (saitohin, an intronless gene within MAPT) (Stefansson et al. 2005). Among these, *NSF* is involved in the NRC/MASC complex, and *MAPT* is a component of the broader postsynaptic density (PSD) complex (Collins et al. 2006).

The H1/H2 inversion suppresses recombination, leading to divergence of the two alleles. The H2 haplotype is shown to exist only in populations with history of European admixture (Evans et al. 2004; Stefansson et al. 2005). Overall, H2 shows little genetic variation (Hardy et al. 2005). In contrast, many H1 variants have been described, with the most common reported variants being single nucleotide polymorphisms (SNPs) or small insertions/deletions that modify *MAPT* (tau) expression or splicing mechanisms (Rademakers et al. 2005; Caffrey et al. 2006; Caffrey et al. 2007; Myers et al. 2007; Hayesmoore et al. 2008).

The H1 & H2 inversion has been described in the context of several neurodegenerative diseases associations. The H2 haplotype was reported to have a protective role in a number of brain diseases, including progressive supranuclear palsy and corticobasal degeneration (de Silva et al. 2001; Pastor et al. 2004; Pittman et al. 2005; Rademakers et al. 2005), Parkinson's disease (Goris et al. 2007; Zabetian et al. 2007) and frontotemporal dementia (Borroni et al. 2005; Ghidoni et al. 2006) (more comprehensive reviews in (Rademakers et al. 2004; Kalinderi et al. 2008)). H2 was also demonstrated to have higher expression of a variant of *MAPT* mRNA transcript known as 4R, where the fine balance of 4R transcript with its 3R counterpart is critical in certain neurodegenerative diseases (Kalinderi et al. 2008).

More recently the 17q21 locus has been shown to be involved in a newly defined genomic disorder. A recurrent microdeletion of the genomic locus was detected in patients with learning disability (Koolen et al. 2006; Sharp et al. 2006; Shaw-Smith et al. 2006). The critical deletion region (~420 kb) was similar in all patients identified. In deletion carriers where parental DNA is available, the deletions were shown to be *de novo*. Moreover, for all patients identified so far, a H2 parental background is a necessary factor for the deletion to occur (Koolen et al. 2008). Clinical cases of mental retardation patients with the reciprocal duplication of the microdeletion syndrome have been detected (Kirchhoff et al. 2007), demonstrating the genomic instability of the region.

6.3.2 Known Genomic Structure of 17q21

The structure of the 17q21 genomic interval has been extensively described by Cruts *et al.* and Pittman *et al.* (Pittman et al. 2004; Cruts et al. 2005). There are three main LCR modules in the region, each with various subunits of duplication blocks (Figure 6.4). These LCRs flank both the H1/H2 inversion and the microdeletion syndrome breakpoints (Stefansson et al. 2005; Koolen et al. 2006; Sharp et al. 2006; Shaw-Smith et al. 2006). Specific segmental duplicons were attributed to the genomic instability of the region. H1 and its sub-haplotypes, together with the inverted H2 allele, differ in terms of the number and the orientation of the duplication blocks. Copy number polymorphisms within H1 at the *NSF* locus, defining several H1 subclades, have previously been reported (Stefansson et al. 2005). The 17q21 locus is a case in point for characterizing the relationship between segmental duplications, copy number changes, genomic instability and their phenotypic consequence.

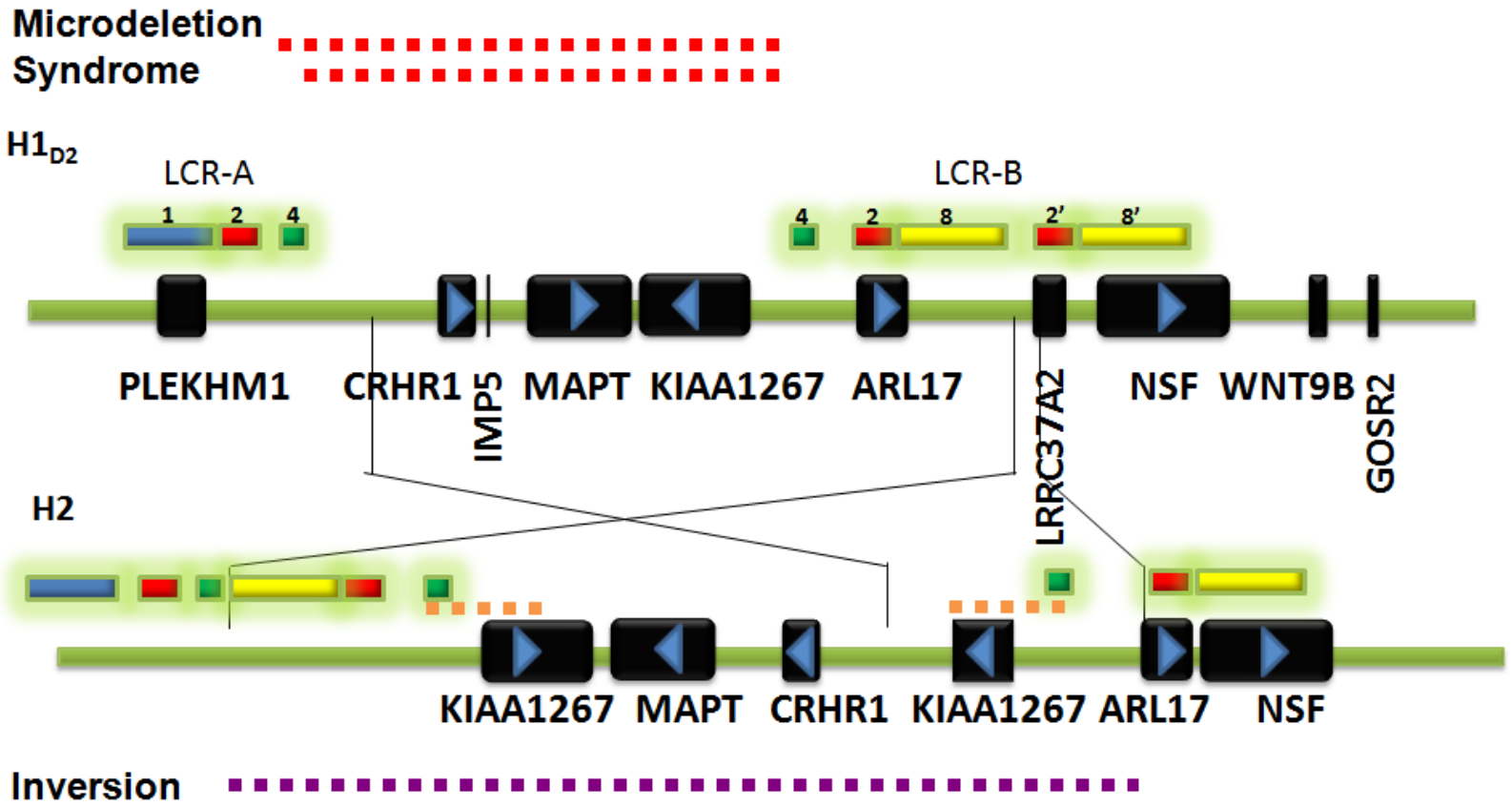


Figure 6.4 Schematic representation of the 17q21 locus. The H1-D2 allele (as in the reference assembly, UCSC genome browser) and the inverted H2 allele are mapped with low copy repeats (LCR) and major segmental duplication blocks. Estimated locations of the H1/H2 inversion and the 500 kb microdeletion syndrome with learning disability is shown as dotted purple and red lines respectively.

6.3.3 Array CGH Data Reveals Two Major CNVs at 17q21

The MASC CNV screen provided evidence for copy number variations near the *NSF* gene at 17q21. Here we present a detailed molecular and cytogenetic study on the genomic structure of the chr17q21 locus. Analysis of the WGTP array data on 269 HapMap samples from three ethnicities[5] delineated two major blocks of CNV: CNV_{NSF} (Figure 6.5, right) and $CNV_{KIAA1267}$ (Figure 6.5, left) (with reference to the genes involved).

6.3.3.1 $CNV_{NSF:}$ Copy number Variant at 5' end of NSF

CNV_{NSF} (~41.7 Mb to 42.2 Mb NCBI36) is part of the low copy repeat LCR A (segmental duplicon subunits 2 and 8) (Cruts et al. 2005). It encompasses the truncated duplication of exons 1-13 of the gene *NSF* (N-ethylmaleamide-sensitive factor). This CNV has been demonstrated previously through sequence analysis in the Icelandic population (Steffanson et al), and has led to the definition of 3 H1 sub-alleles designated as ***H1-D0-1*** or ***H1-D1*** (both having 1 copy of NSF), ***H1-D2*** (2 copies of NSF) and ***H1-D3*** (3 copies of NSF).

Bivariate clustering of the log2ratio of two clones overlapping *NSF* (*Chr17tp-2F9* and *Chr17tp-12B1*) confirmed the multi-allelic nature of CNV_{NSF}, but revealed higher complexity than previously expected. Given that the reference DNA used for array hybridization has a three copies of the sequence (demonstrated by FISH experiments), samples from Yoruba are estimated to have diploid copy numbers varying from 2 to 5 (4 distinct genotype signals), while samples from Asia carry diploid copy numbers varying from 2 to 7 (6 distinct genotype signals) (Figure 6.6). Data from the European samples is less easy to interpret, due to homology of BAC sequences to the second CNV block

[5] The three populations are CEU: Samples from Utah with European ancestry; YRI: Samples from Yoruba with African ancestry; JPT+CHB: Samples from Japanese and Chinese with Asian ancestry

$CNV_{KIAA1267}$, but they also seem to display diploid copy numbers varying from 2 to 7. Combining the evidence, a maximum of 7 copies of the truncated *NSF* (diploid state) was identified, suggesting the existence of a ***H1-D4*** allele with 4 copies of 5' *NSF* (haploid state). The presence of such allele was confirmed in a CEU sample by FISH experiments (section 6.3.3).

6.3.3.2 $CNV_{KIAA1267:}$ Copy number Variant at KIAA1267

$CNV_{KIAA1267}$ (left, ~41.5 Mb to 41.7 Mb), the second major CNV block, overlaps with subunit 4 of LCR-A (Cruts et al. 2005), and a brain-expressed gene *KIAA1267* (Oliveira et al. 2004). The existence of structural variation at the region has previously been implicated in the studies of H1 and H2 inversion haplotypes. The H2 inverted allele and the H1-D1 allele (a variant of H1) were suggested to have overlapping but not identical duplications at this region (Stefansson et al. 2005).

Clustering the log2ratio of BAC clones at $CNV_{KIAA1267}$ (*Chr17tp-2G12* and *Chr17tp-3D7*) generated 3 distinct clusters in the CEU samples, corresponding to 2, 3 and 4 copies of $CNV_{KIAA1267}$. In Yoruba or Asian samples, on the other hand, the CNV is non-variant. In conclusion, the higher copy number of $CNV_{KIAA1267}$ is European specific.

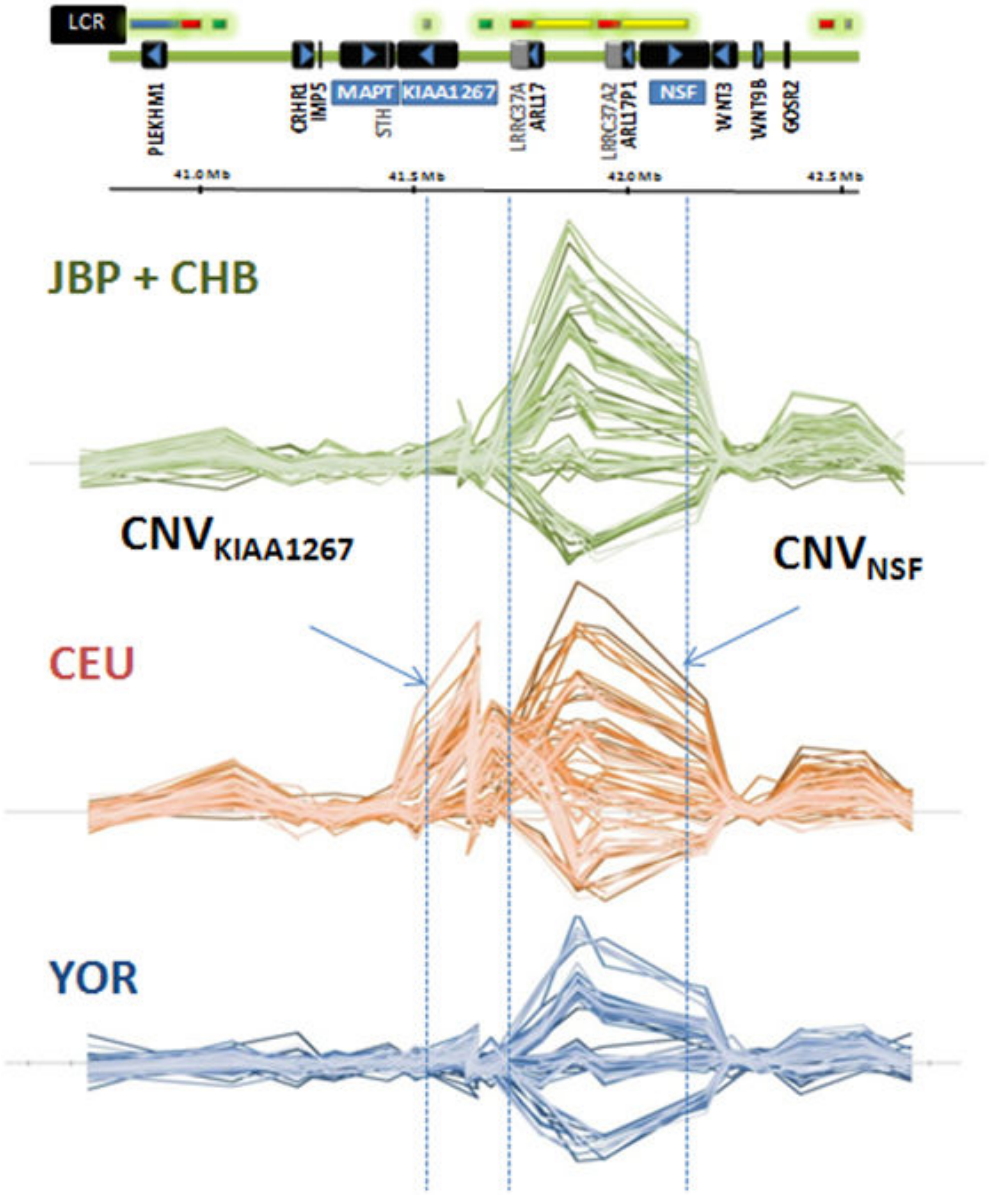


Figure 6.5 Array CGH genomic profiles of the 3 HapMap ethnic groups at chr17q21. Top panel: Schematic representation of the region chromosome 17 40.8 Mb- 42.5 Mb, with locations of genes and low copy repeats for reference. Bottom panel: Genomic profiles from WGTP data on samples of Asian ancestry (JPT+CHB), samples of European ancestry (CEU) and samples of African ancestry (YOR). Two major blocks of CNVs are indicated in the diagram (*left*: $CNV_{KIAA1267}$; right: CNV_{NSF}).

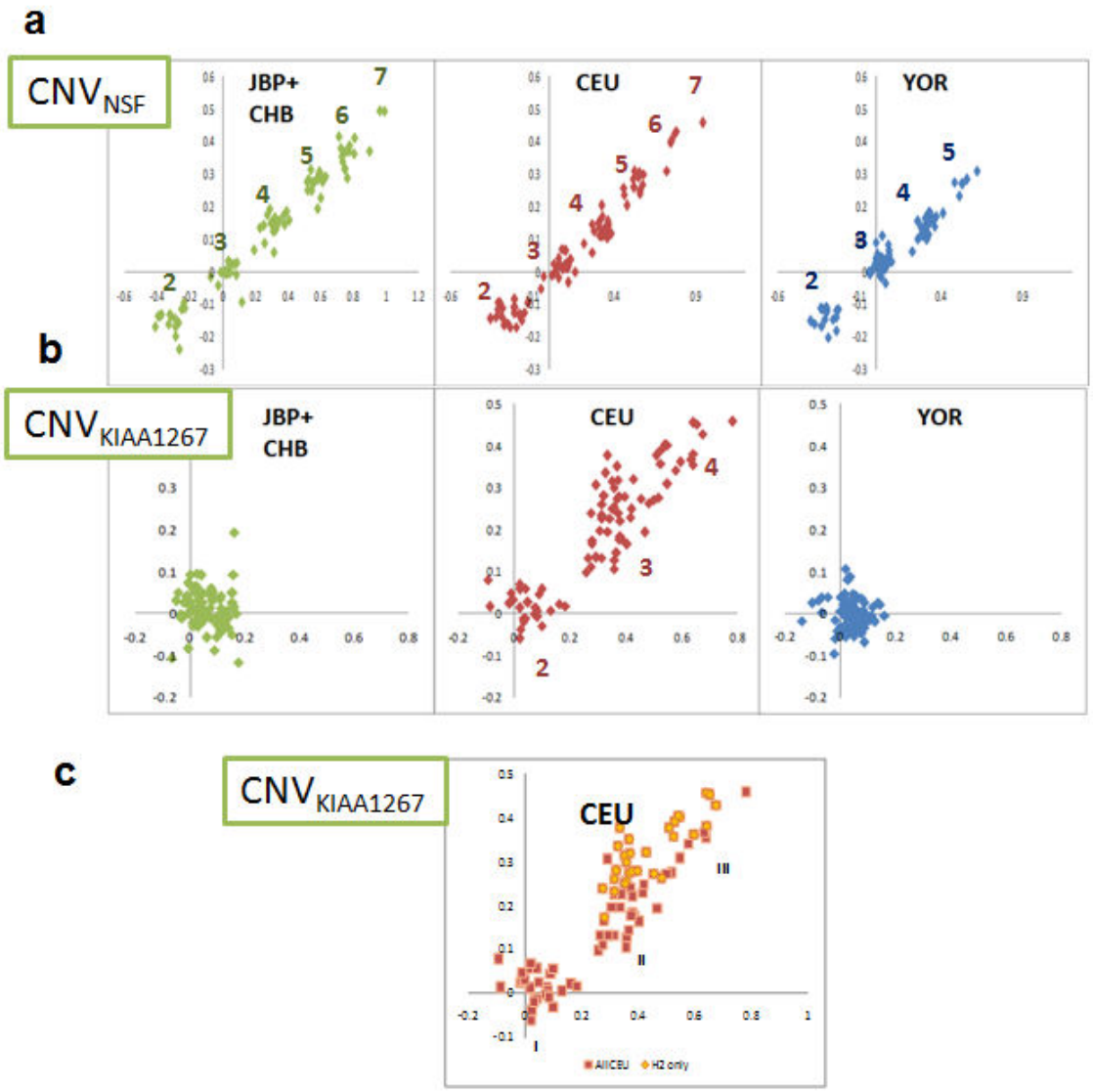


Figure 6.6 Population differentiation at two major CNV blocks at 17q21 Bivariate clustering of log2ratio from WGTP data demonstrates population bias at CNV_{NSF} and $CNV_{KIAA1267}$. **a)** *CNV_{NSF}: 2-7 copies (diploid) in JPT+CHB and CEU; 2-5 copies (diploid) in YOR. **b)** **$CNV_{KIAA1267}$: 2-4 copies in CEU; non-variant in JPT+CHB or YOR (2 diploid copies in each individual). **c)** H2 carriers (orange) are mapped exclusively to the higher copy number clusters (II and III) of the CEU diagram representing $CNV_{KIAA1267}$. The remaining samples at cluster II and III belongs to an H1 allele variant, H1-D1.

* Bivariate clustering of log2ratio from *Chr17tp-14B1* (x-axis) and *Chr17tp-2F9* (y-axis)
** Bivariate clustering of log2ratio from *Chr17tp-2G12* (x-axis) and *Chr17tp-3D7* (y-axis)
JPT+CHB: samples from Japan and China with Asian ancestry; CEU: samples from Utah with European ancestry; YOR: samples from Yoruba with African ancestry

6.3.4 Resolving CNVKIAA1267 with SNP and High Resolution Oligo Array Data

The WGTP data demonstrated a European-specific pattern at $CNV_{KIAA1267}$. Coincidentally, H2 from the common inversion was also found in the European population or population with European admixture only (Evans et al. 2004; Stefansson et al. 2005). To explore a relationship between $CNV_{KIAA1267}$ and H2, we integrated the HapMap WGTP CNV results with corresponding SNP data (IHMC 2005). SNP rs9468, a surrogate marker for the H1/H2 inversion, was used to categorize samples into H1 or H2 carriers. Samples carrying one or more H2 alleles were plotted on the CEU diagram of $CNV_{KIAA1267}$ (Figure 6.6c). These H2 carriers (orange) all mapped to higher copy numbers of $CNV_{KIAA1267}$ (cluster II & III).

However, $CNV_{KIAA1267}$ is only partially explained by H2. Non-H2 variants were also observed at the high-copy-number clusters of $CNV_{KIAA1267}$. This suggests heterogeneity at the CNV locus. High resolution oligonucleotide from HapMap individuals (Genome Structural Variation Consortium 2008) confirms at least two types of duplications at the $CNV_{KIAA1267}$ locus (Figure 6.7). The smaller duplication was explained by H2 carriers. The larger CNV would involve the H1-D1 variant, an allele defined by previous literature as having a duplication overlapping the region. Like H2, this H1-D1 allele is absent in non-European populations (see Figure 6.6b, diagrams for JBP+CHB/YOR).

This is the first demonstration of a non-H2 variant being European specific, and it raises questions of the evolutionary relationship between this H1-D1 variant and the H1/H2 inversion.

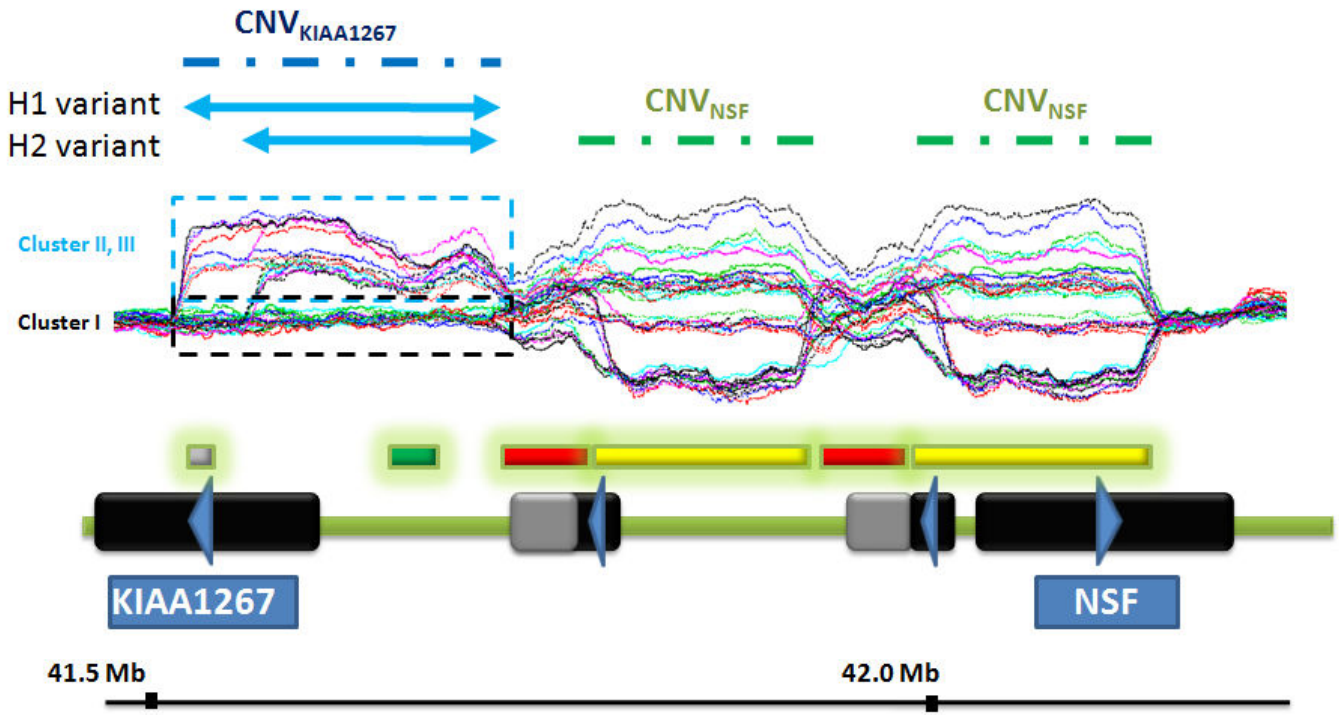


Figure 6.7 High resolution oligo array CGH profiles for 20 individuals at chr17 41.5 Mb- 42.2 Mb. Top panel: Nimblegen oligonucleotide array data for 20 HapMap CEU samples resolved the two CNV blocks: CNV_{NSF} and $CNV_{KIAA1267}$. $CNV_{KIAA1267}$ is heterogeneous, comprising of a larger variant (corresponding to H1-D1 carriers) and a smaller variant (corresponding to H2 carriers). Bottom panel: Schematic representation of the region with locations of genes and low copy repeats for reference.

6.3.5 Validating CNVNSF and CNVKIAA1267 by qPCR and FISH

Given the complexity and multi-allelic nature of the described copy number variations, array CGH data at Chr17q21 is best interpreted with independent platforms to provide quantitative and qualitative measurements in terms of the number, size, distance and orientation of CNVs.

First, quantitative real-time PCR (qPCR) was performed to validate the multi-allelic variants CNV_{NSF} and $CNV_{KIAA1267}$. Genotypes of CNV_{NSF} ranging from 2 to 6 copies (diploid) were validated, and genotypes of $CNV_{KIAA1267}$ from 2 to 4 copies (diploid) were confirmed (Figure 6.7).

To visualize the copy number and the orientation of rearrangement at the *NSF* locus, we performed fluorescent in situ hybridizations (FISH) using DNA fiber slides made from lymphoblastoid cell lines. Fosmid probes at the 5' duplicated (labeled with biotin-dUTP, green) and 3' non-duplicated ends (labeled with digoxigenin-dUTP, red) of the *NSF* CNV locus.

The haplotypes of different alleles from HapMap samples were visualized according to the number of green signals (Figure 6.8). Duplication of the 5' *NSF* is tandem, ranging from copy number of 1 to 4 (haploid) on a given allele. The presence of the ***H1-D4*** (4 copies of 5' *NSF*) as we described before is validated.

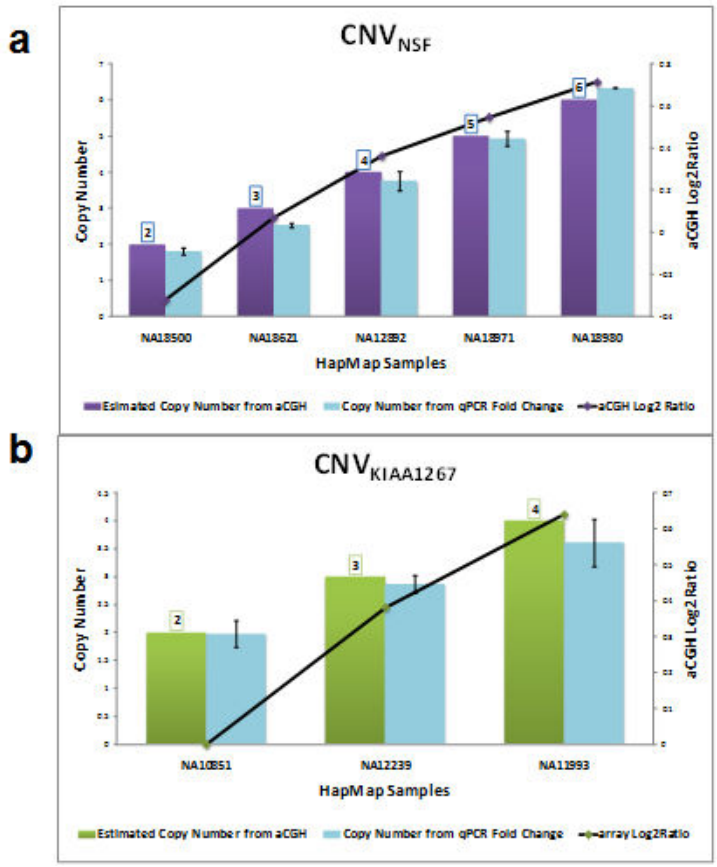


Figure 6.8 Quantitative PCR validation of $CNV_{KIAA1267}$ and CNV_{NSF}. a) CNV_{NSF} detected as 2 to 6 copies*. **b)** $CNV_{KIAA1267}$ detected as 2 to 4 copies*.
** diploid copy number*
blue: copy number calculated from qPCR; purple or green: expected copy number

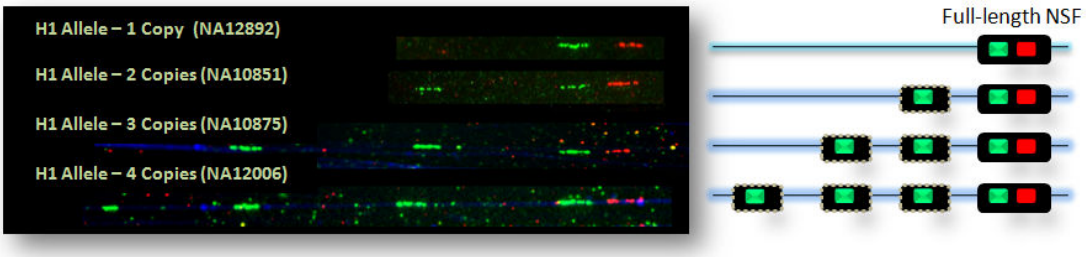


Figure 6.9 Fiber-FISH experiment to visualize copy number of CNV_{NSF}. A fosmid probe at the 5' *NSF* portion overlapping with CNV_{NSF} was labeled green and another fosmid at the 3' *NSF* invariant portion was labeled red. Alleles with tandem duplication from 1 to 4 copies* of CNV_{NSF} were observed
** haploid copy number*

6.3.6 Genotyping HapMap Individuals for 17q21 Structural Variants

Combining available genetic data (CNV and SNP) with the findings of various structural variations in the region, we genotyped all unrelated HapMap individuals for the 17q21 allele carried (children from the trios were removed from analysis). Table 6.2 demonstrates the trends of population differentiation. Horizontally across the table one sees a European specificity for both H2 and H1-D1 alleles, corresponding to higher copy numbers of $CNV_{KIAA1267.}$ Vertically down the table one sees an Asian and European bias for the highest copy numbers (6 or 7) of CNV_{NSF}.

Table 6.2 CNV genotypes for all unrelated HapMap individuals* at $CNV_{KIAA1267}$ and CNV_{NSF}.

CNV-NSF	H2 allele			H1-D1 allele			Other H1 Variants			Population Total		
Copy Number	ASIA	CEU	YOR	ASIA	CEU	YOR	ASIA	CEU	YOR	ASIA	CEU	YOR
2	0	0	0	0	15	0	30	15	20	30	30	20
3	0	11	0	0	12	0	34	11	58	34	34	58
4	0	4	0	0	4	0	50	22	36	50	30	36
5	0	7	0	0	0	0	34	9	6	34	16	6
6	0	2	0	0	0	0	26	4	0	26	6	0
7	0	0	0	0	0	0	6	0	0	6	0	0
Total	0	24	0	0	31	0	180	61	120	180	116	120

*90 ASIA samples, 58 CEU samples, 60 YOR samples

**offspring from the trios in CEU and YOR were removed from analysis so that only unrelated HapMap individuals were genotyped*

6.3.7 CNVNSF and CNVKIAA1267 in Schizophrenia Versus Control

In chapter 4.6 we introduced bivariate clustering of BAC clone log2ratio in the WGTP data to genotype CNV and test for disease association. The 17q21 region near *NSF* was one of the 31 regions across the genome with significant bias between schizophrenia patients and controls. We examined this region in the case and control cohorts in more detail with respect to CNV_{NSF} and $CNV_{KIAA1267}$.

We genotyped CNV_{NSF} and $CNV_{KIAA1267}$ using the same clones as in the HapMap analysis. Clusters were manually merged for CNV genotype counts. (Figure 6.9)

Both the Scottish SCZ and LBC cohorts demonstrate a full range of copy numbers for CNV_{NSF} (2-6+) and $CNV_{KIAA1267}$ (2-4), consistent with HapMap samples of European ancestry. For CNV_{NSF} there is a slight bias of patients at lower copy number but the difference is non-significant after clusters are manually merged (p value = 0.1931, chi-squared test). For $CNV_{KIAA1267}$ there is a nominally significant bias (p-value= 0.01689) of an association of low copy number with disease (Table 6.3). Nevertheless, the result is treated with caution since our case and control cohorts are modest in size and may therefore subject to sampling bias.

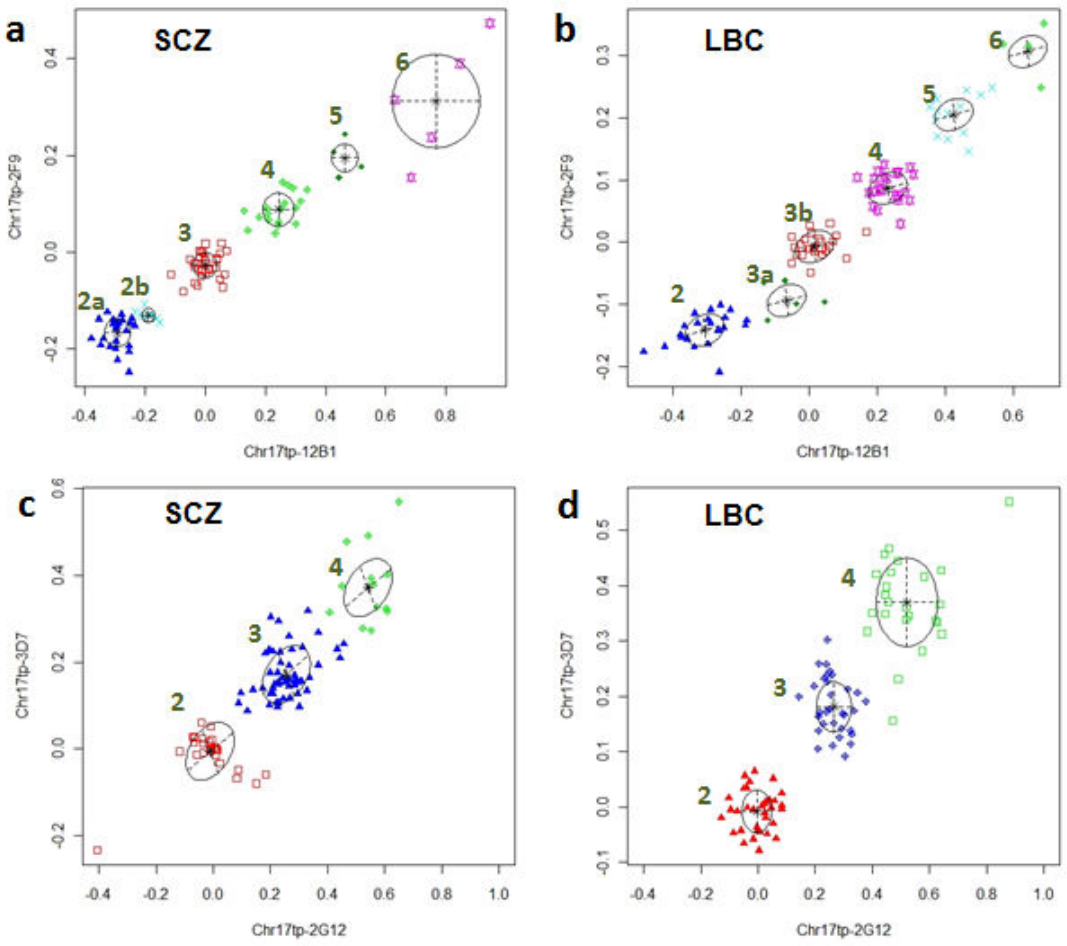


Figure 6.10 Comparing SCZ and LBC samples at $CNV_{KIAA1267}$ and CNV_{NSF}.
(a) & (b) CNV_{NSF} was detected as 2 to 6 copies in SCZ and LBC samples*.
(c) & **(d)** $CNV_{KIAA1267}$ was detected as 2 to 4 copies in SCZ and LBC samples**.
* Bivariate clustering of log2ratio from *Chr17tp-14B1* (x-axis) and *Chr17tp-2F9* (y-axis)
** Bivariate clustering of log2ratio from *Chr17tp-2G12* (x-axis) and *Chr17tp-3D7* (y-axis)

Table 6.3 SCZ and LBC genotype counts at $CNV_{KIAA1267}$ and CNV_{NSF}.

CNV-NSF Copy Number	SCZ	LBC	CNV-KIAA1267 Copy Number	SCZ	LBC
2	32	22	2	25	34
3	31	30	3	49	32
4	18	22	4	13	24
5+	9	17	5+	0	0
Total	90	91	Total	87	90

6.3.8 Evolutionary History of CNVNSF and CNVKIAA1267

To elucidate the relationships between CNVs and the more ancient segmental duplications in the 17q21 region, we classified segmental duplicons into ancestral or derived loci according to in silico predictions from Jiang *et al.* (Jiang et al. 2007) (Figure 6.10). Both CNV_{NSF} and $CNV_{KIAA1267}$ were relatively recent events, flanked by more ancient segmental duplicons (ancestral loci). Genomic regions near ancient segmental duplicons were suggested to be genetically more fragile, favoring evolution of the younger polymorphic CNVs in the complex region (Jiang et al. 2007). These ancestral SDs probably acted as predisposition factors for structural rearrangements in the region, leading to the formation of CNV_{NSF} and $CNV_{KIAA1267}$.

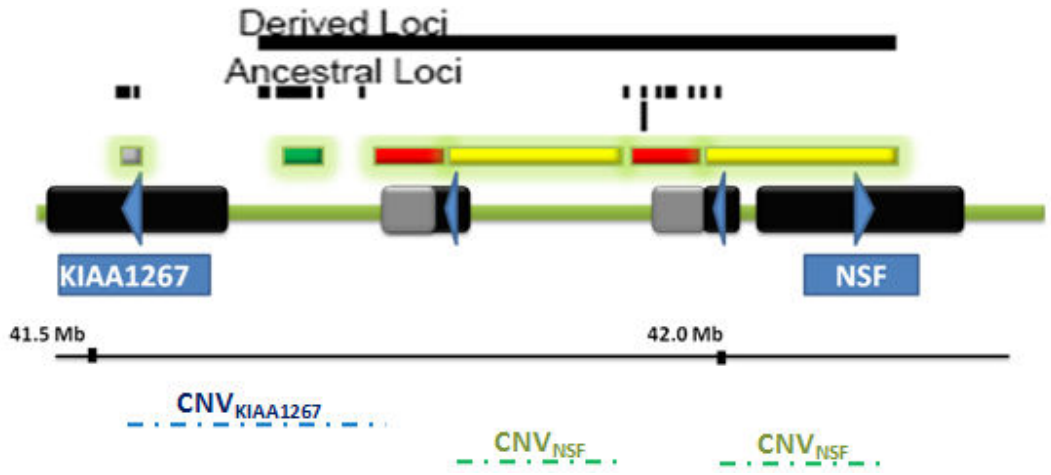


Figure 6.11 Locations of derived and ancestral loci of segmental duplications at 17q21. Locations of the two types of segmental duplications were depicted with respect to $CNV_{KIAA1267}$ and CNV_{NSF}.

6.4 Chapter Summary and Discussion

In this chapter we described a biological pathway approach to studying normal copy number variations in genes relevant to cognition and schizophrenia. Using NRC/MASC as a complex central to neurological functions, we assessed whether such components could be involved in copy number variations.

In the first part of this chapter we described 20 MASC CNVs detected in 269 apparently normal HapMap individuals. Four of these overlapped with core MASC components. Three were confirmed by high-resolution oligonucleotide array data. The presence of CNVs at such important synaptic proteins in healthy individuals suggests that (i) CNVs at these genes are benign and did not translate into pathogenic gene function (e.g. intronic CNVs may have no functional consequence); or (ii) neurological functions have high plasticity such that disrupting functions of one synaptic component could be compensated by other mechanisms in the network; or (iii) such CNVs did affect neurological functions, but the effect was not detrimental, perhaps contributing to variations in brain function, e.g. cognitive ability, among the population. The last hypothesis is particularly interesting for the genetics of neuroscience, and future studies investigating CNVs in human cognitive and behavior traits may reveal its validity.

On the other hand, investigating CNVs in such an important set of synaptic molecules provides us a baseline to evaluate existing and upcoming CNV screens of neurological diseases. A number of psychiatric disease CNV studies have reported variants at neuronal complexes as potential CNV candidates. A deletion at *DLG2*, for instance, was detected as a patient CNV in one genome-wide schizophrenia screen (Walsh et al. 2008). However, CNVs in the same gene were also identified in control samples in another schizophrenia study (ISC 2008), and in our analysis of normal HapMap

individuals. It is not surprising that disease-causing CNVs might exist in healthy controls, perhaps at lower frequencies, at slightly varied genomic locations, or with combinations of different interacting factors. Alternatively, such CNVs may be benign in both patients and controls. It is therefore important to characterize CNVs at key neuronal complexes in normal healthy individuals as a basis for disease comparison.

In the second part of the chapter, we described copy number variations near the *NSF* locus in HapMap samples from three ethnic groups. We revealed two major blocks of copy number variation, CNV_{NSF} and $CNV_{KIAA1267,}$ and further characterized them using quantitative real-time PCR, Fiber-FISH and oligo array CGH data. Interestingly, both CNVs showed population bias across the three ethnic groups examined. In particular, we revealed a H1 variant specific to the CEU samples, which is analogous to the European-specific H2 inverted allele. This H1 variant could represent an evolutionary remnant of the H1/H2 inversion, or a more recent structural rearrangement specific to the European population.

Chapter 7
General Discussion

7.1 Design of CNV Discovery and Association Study

My PhD thesis describes a survey of copy number variation associated with schizophrenia. Deciphering the genetic basis of schizophrenia has always been a major challenge in human genetics. In the past two years, CNV discovery and association studies have generated some promising results in the field of neuropsychiatric genetics, and more investigations on copy number variation in psychiatric diseases can be expected in the near future. Below I discuss the current and future trends of such CNV studies.

7.1.1 Multiple Approaches of CNV Study

Multiple genetics approaches and experimental designs are necessary to explore complex genetic traits such as schizophrenia. Our study focussed on CNV investigation with two components: a) a familial study of CNVs associated with schizophrenia in pedigrees with multiple affected individuals; and b) a population-based CNV investigation. There are many possible approaches for identifying CNVs in disease, including (i) looking for concordance in affected siblings or twins; (ii) contrasting CNVs in affected versus non-affected siblings of schizophrenia; (iii) identifying transmitting and non-transmitting CNVs in families or trios (to investigate the role of inherited or *de novo* CNVs) (Xu et al. 2008); iv) identifying common CNVs in families with high density of illness; (v) large-scale population-based case-control cohort CNV associations (ISC 2008; Stefansson et al. 2008); (vi) and investigating the less explored somatic CNVs in diseases. Furthermore, early discovery-based CNV studies revealing novel CNV loci are paving ways to more comprehensive case-control association studies, either at the genome-wide level or targeting particular genes of interest (eg. Rejuscu *et al.* reported a study of CNVs in Neurexin genes) (Rujescu et al. 2008). Combinations of these

approaches will reveal different aspects about the genetic architecture of the disease and the candidate genetic pathways underlying schizophrenia pathogenesis.

7.1.2 Future Large-Scale CNV Studies

Our CNV study on schizophrenia was based on modest sample sizes of 4 familial pedigrees and a case-control cohort of 91 patients and 92 controls. To achieve more definitive disease association, substantially larger sample sizes would be necessary. Recently published schizophrenia CNV studies demonstrated the importance of large-scale studies to identify novel rare recurrent CNV loci (ISC 2008; Stefansson et al. 2008) and to detect association. These risk loci have low frequency in the population, probably due to reduced fecundity as conferred by disease (Stefansson et al. 2008), and would require at least thousands of samples for sufficient statistical detection power. In contrast, common alleles are frequent but may only have marginal increase risks between case-controls, therefore considerable sample sizes would again be required to detect any association.

Collaboration among clinical and research centres for sample collections will largely facilitate schizophrenia genetic research. The most recent CNV studies are results of such large collaborative efforts, such as ISC (International Schizophrenia Consortium) (http://pngu.mgh.harvard.edu/isc/), SGENE (http://www.sgene.eu/) and GROUP (Genetic Risk and Outcome in Psychosis) (https://www.group-project.nl/). Combined, these 3 studies involved close to 8,000 psychosis samples (ISC 2008; Stefansson et al. 2008; Vrijenhoek et al. 2008). One limitation of such multi-centre studies would be the genetic heterogeneity and diagnostic differences between samples collected from different sources (Burmeister et al. 2008). Nevertheless, continued collection of patient and control cohorts, together with improved genotype-phenotype characterization, will be instrumental in rapid progress towards the understanding of schizophrenia.

7.1.3 Enrichment of Schizophrenia Subtypes

Another decision to make in CNV study design is the choice of suitable patient types or subtypes. One criterion is the inclusion of sporadic versus familial cases. The study of Xu *et al.*, for example, compared these two patient types in trios (parent-child sets) and demonstrated that *de novo* CNVs increased disease risk in sporadic cases (a 8 times increase was reported for cases versus control), but not in familial cases (Xu et al. 2008). The study suggested variable genetic architecture among the sporadic and the familial subtypes of patients, namely the involvement of *de novo* CNVs versus inherited CNVs. Although there were examples of *de novo* CNV involved in familial cases, for example in an affected sib pair in autistic spectrum disorder (as explained by post-zygotic mosiacism) (Cook and Scherer 2008), these were mostly exceptions rather than the rule. Consequently, by enriching sample cohorts to either sporadic or familial cases, CNV study may steer towards identification of different types and characteristics of CNVs.

Analogous to conventional linkage and association studies with SNPs, the use of endophenotypes in CNV studies would be another debatable strategy. Endophenotypes represent quantifiable trait markers for the inheritance of a given genotype (see section 1.5) (Gottesman and Gould 2003). Schizophrenia as a complex disease may comprise many diverse behavioural and cognitive traits, each corresponding to different underlying genetic variants. One rationale behind using endophenotypes in schizophrenia study was the advantage of bypassing traditional disease classification to obtain a more phenotypic homogenous group of patients. By sub-typing samples according to endophenotypes, one might increase detection power for the underlying genetic risk variants (Abrahams and Geschwind 2008). For example, the phenotype of P50 sensory gating deficit demonstrated association with a *CHRNA7A* polymorphism with a LOD score of 5.3, while association with schizophrenia disease as a phenotype resulted in a

much lower LOD score of 1.3 (Freedman et al. 1997). Magnetic Resonance Imaging (MRI) has also been used to determine association between functional variants of a serotonin transport gene and anxiety (Hariri et al. 2002). The disadvantage of using such endophenotypes would be the high cost involved, for example in performing imaging studies on every single patient. Other quantifiable traits, for instance executive function and visual memory, could also be incorporated as patient subtype enrichment (ISC 2008).

Finally, patients with different schizophrenia disease onset could be another subtype for enrichment in CNV study, and could alter the likelihood of detecting disease-causing variants. Evidence came from a study from Walsh *et al.*, who reported that patients with early or childhood onset were particularly enriched with rare copy number variants (Walsh et al. 2008). These early onset cases, or other schizophrenia subtypes representing more “severe” forms of the disease, may provide a good subset of patients for CNV detection (Vrijenhoek et al. 2008), while subsequent association studies could be performed in a larger, more general schizophrenia cohort. In addition, atypical clinical cases such as those co-occurring with mental retardation and dysmorphology may also be considered for patient enrichment, although CNVs detected in such cohorts may be linked to other phenotypes rather than schizophrenia *per se*.

7.2 Understanding the Genetic Model of Schizophrenia

7.2.1 Rare Variants Versus Common Variants

CNV studies have provided us with important insights into the genetic model of schizophrenia. One controversy surrounding schizophrenia genetics and many other complex disease traits is the debate between common disease-common variant (CD-CV) (Lohmueller et al. 2003) versus common disease-rare variant (CD-RV) models (Pritchard 2001; McClellan et al. 2007). The CD-CV model proposes that common alleles with small to moderate disease risks may have additive or multiplicative effect on schizophrenia (Cook and Scherer 2008), and have inspired many SNP association studies investigating common polymorphisms in schizophrenia. Recent large-scale association studies have successfully identified such risk variants, one of which being the polymorphism at *ZNF804A* associated with schizophrenia (O'Donovan et al. 2008).

The CD-RV model suggests that heterogeneity of schizophrenia comes from multiple rare variants, a theory which has received increasing attention in schizophrenia genetics during the past decade (Craddock et al. 2007; McClellan et al. 2007). Recent CNV studies further demonstrated that rare risk loci, individually or collectively, could predispose to schizophrenia (ISC 2008; Stefansson et al. 2008), supporting the CD-RV model. Our study replicated the findings of one of these rare loci (the 15q11.2 deletion), and provided a number of other putative disease-specific genetic variants. Apart from locus heterogeneity, CNV studies showed that even within the same disease-associated genomic locus there could be allelic heterogeneity (for example in the case of the Neurexin1 gene, section 1.6) (Rujescu et al. 2008). Some of these rare alleles could be recurrent in the population (either inherited or sporadic) but would remain at low

frequency due to the selective disadvantage they confer (Pritchard 2001), while others may be private events found only in single individuals or families.

On the other hand, current CNV discoveries, including our results, may be skewed towards the lower end of the minor allelic frequency spectrum (Ionita-Laza et al. 2008b), possibly due to the bias of CNV platforms and algorithms detecting relatively large mutations. As advances in CNV discovery techniques allow accurate detection of smaller, more frequent copy number polymorphisms, we may be able to detect common CNV alleles associated with schizophrenia. As presented in chapter 6 of this thesis, we attempted to test disease association of two common copy number polymorphisms, using targeted, PCR-based strategies, one of which was a small CNV of ~5.5 kb demonstrating correlation to gene expression.

Genetic studies of other complex diseases (e.g. in Alzheimer's disease) have demonstrated that the CD-CV and CD-RV theories need not be mutually exclusive in disease populations (McClellan et al. 2007). CNV studies on schizophrenia combined with SNP association have further suggested that these two models coexist, leading to the genetic heterogeneity in schizophrenia.

7.2.2 Incomplete Penetrance and Expressivity

Genetic investigations on schizophrenia have usually supported a complex mode of inheritance. Incomplete penetrance was demonstrated in a number of schizophrenia-associated CNV loci. The three novel recurrent deletions at 1q21, 15q11, 15q13, for example, were all detected in normal, healthy controls, suggesting penetrance of these genetic variants was not complete. In our familial CNV study of the *ABCA13* deletion (section 3.3), incomplete penetrance could explain the presence of deletions in one normal family member within the pedigree. This is consistent with the incomplete segregation in previous familial CNV studies, such as the *DISC1* translocation in an extended Scottish pedigree in which only 70% of translocation carriers were affected (Millar et al. 2003). Instead of a Mendelian mode of transmission, many structural variants demonstrate complex inheritance patterns with a variable degree of penetrance, as well as variable expressivity depending on interactions with other genetic and environmental factors.

7.2.3 Aetiological Overlap of Schizophrenia with Other Psychiatric Diseases

The well-documented link between the 22q11 deletion syndrome and schizophrenia demonstrated that the same structural variant can be associated with numerous psychiatric phenotypes, from schizophrenia to attention-deficit hyperactivity disorder, bipolar disorders and autism spectrum disorders (ASDs) (Antshel et al. 2007; Burmeister et al. 2008; Cook and Scherer 2008). These and other genetic loci argued for pleiotropic genetic effects among psychiatric disorders.

Our study, together with a number of CNV reports on schizophrenia in the past year, has reinforced this notion of aetiology overlap among neuropsychiatric diseases. The

15q11.2 recurrent deletion as detected in our patients and reported in Steffanson's study, for instance, is a region involved in Prader-Willi/ Angelman Syndrome with known cognitive and neurological impairment (Cassidy et al. 2000), and a subset of patients also exhibited autistic features (Hogart et al. 2008). The 15q13.3 recurrent deletion was previously associated with mental retardation and seizures (Sharp et al. 2008), as well as with ASDs (Veltman et al. 2005). Another recurrent schizophrenia CNV at neurexin1 (Kirov et al. 2008; Rujescu et al. 2008; Walsh et al. 2008) was also detected in autism (Autism Genome Project Consortium 2007) and mental retardation patients (Friedman et al. 2006). Furthermore, we detected multiple schizophrenia cohort-specific rare variants with candidate genes for multiple psychiatric diseases such as bipolar disorder (*PIK3C3*) (Stopkova et al. 2004), Myoclonus-Dystonia with obsessive compulsive disorder (*SGCE*) (Marechal et al. 2003; Misbahuddin et al. 2007), and mental retardation and Alzheimer's Disease (*RCAN1*) (Porta et al. 2007).

The CNV study in schizophrenia provided evidence that some genetic risk factors could give rise to diverse psychopathology falling across the traditional diagnostic boundaries. Alternatively, they may also reveal misdiagnosis of psychiatric diseases, or of variable expressivity of the genes involved (Cook and Scherer 2008).

7.3 Clinical Relevance of CNVs in Schizophrenia

7.3.1 CNV Findings Translating into Disease Classification

Investigating CNVs in psychiatric genetics has strengthened the view that the same genetic factors could predispose to diverse psychiatric diseases across traditional diagnostic categories. Schizophrenia and bipolar disorder, for instance, share symptoms of psychosis and mood disorders, with substantial overlap in genetic susceptibility, and are conventionally dichotomized into two psychiatric disorders (Craddock et al. 2006). In contrast, a single psychiatric disorder may comprise subtypes grouped into one disease category for diagnostic purposes. This concept of clinical diversity was recognized as early as the time of Kraeplin and Bleuler, who introduced subtypes or categories of schizophrenia (Bleuler 1911; Kraepelin 1919).

Future characterization of CNVs and other genetic variants in a wider range of psychiatric patients have the potential to revolutionize disease diagnosis. Traditional diagnostic criteria heavily rely on clinical signs and symptoms. These guidelines have presented a major challenge in studying psychiatric disorders (WHO 1992; American Psychiatric Association. 1994), since such diseases present no tangible lesion or phenotype, but rather complex behavioural and cognitive phenotypes. With information from CNV and other molecular genetics, future disease diagnosis could become more genetic-based, giving precise genotype-phenotype definition and more personalized treatment. Of equal importance, the information from genetic research could guide clinicians to a better understanding of phenotypes of the various psychiatric disorders. These will potentially have a substantial impact on clinical practice.

The more precise characterization of genotype-phenotype correlations through CNV studies will eventually translate back to advancement in molecular genetic research.

With high certainty of disease classification and refined phenotype descriptions, disease cohorts could be stratified to homogenous subtypes for genetic studies. This will in turn benefit the investigation of genetic variation associated with psychiatric phenotypes, leading to better understanding of the biology of schizophrenia.

7.3.2 Genetic Counselling and Therapeutic Potential

The translation of CNV findings into genetic testing and therapeutics remain a more distant target. In particular, the issue of genetic testing in schizophrenia is confounded by the complex genetic architecture and heterogeneity of the disease. Multiple rare variants - each having small effect size (as identified by current CNV studies) - are not particularly informative for genetic testing, nor are common variants with small to modest disease risks. Nevertheless, the accumulation of CNV data on pathogenic versus benign variants will be beneficial for clinical and research geneticists screening for variants in patient cohorts, for example by routine karyotyping and CGH experiments. Ultimately, high-throughput and reduced cost in sequencing techniques for individual genome sequencing could provide more complete knowledge of the catalogue of genetic risk factors in schizophrenia. Such techniques will also generate comprehensive genetic information on an individual basis for genetic testing for disease diagnosis and prevention.

Regarding CNV-based treatment of schizophrenia, multiple biological pathways, such as those involved in neuroligin-neurexin or neuregulin-ErRb signalling in the synapse, have started to emerge from recent CNV studies (reviewed in section 1.6). Some components of these signalling pathways are also known drug targets (Marchionni et al. 1996; Wang et al. 2008). Integrating current and future findings of CNVs into pharmacology could

rapidly progress the development of novel therapeutic strategies to treat psychiatric disorders.

7.4 Thesis Summary

This dissertation describes a multi-faceted study of copy number variants in schizophrenia, mainly based on array CGH platforms. In chapter 3 we presented a number of familial cases with schizophrenia and other psychiatric illnesses, and illustrated the search for CNVs carried by all affected members within family. Two rare CNV candidates were described, one on 1p36 with partial duplication of *H6PD* and *SPSB1*, and the other an intronic deletion on 7p12 within *ABCA13*. These two rare CNVs segregate with disease and could possibly play a role in disease pathogenesis. Extensive characterization of the 1p36 duplication revealed an interesting genetic architecture at the CNV breakpoints.

In chapter 4, we described a case-control population-based CNV study in schizophrenia. 91 patients and 92 ethnically matched controls were inspected for CNVs using WGTP BAC array. Subsets of CNVs were validated by qPCR and SNP genotyping CNV data. Based on previous literature establishing a role of rare variants in schizophrenia, we identified a number of rare variants specific to the disease cohorts overlapping plausible disease candidate genes. We also replicated the finding of the 15q11.2 novel schizophrenia-associated deletion as described by Steffanson *et al.*. Extending the findings into a larger CNV dataset, we explored the presence of recurrent CNVs as disease-causing candidates in schizophrenia. Of particular interest was the observation that multiple schizophrenia-specific rare CNVs overlap with candidate genes for various psychiatric disorders.

In the second half of Chapter 4 we presented an investigation of common copy number polymorphisms and schizophrenia. 31 out of 577 “genotypable” CNVs reported significant differences in genotype distributions between cases and controls. We further

genotyped two of these CNPs, a deletion 5' of *CHL1*, and a deletion spanning *CHRFAM7A*, by PCR-based strategies. Both candidates have significant links to schizophrenia with evidence from the literature. Neither candidate revealed a statistically significant bias in the extended case/control cohort of ~300+300. The two candidates and the genotyping experiments were described in Chapter 5.

Finally in Chapter 6, we targeted a set of functionally important molecules, the NMDA-receptor complex (NRC/MASC) in the post-synaptic density of neurons. Among the 183 NRC/MASC genes, we identified CNVs in 20 of these cognitive and behaviour-related genes in normal HapMap individuals. These copy number variants may confer a phenotypic effect on normal individuals by modulating synaptic functions through the NRC complex. Copy number variants near the *NSF* gene in 17q21 were further characterized. The two CNV blocks showed unusual population bias, and we identified a new European-specific H1 haplotypes with potential evolutionary significance.

7.5 Future Direction

The identification of CNVs as risk loci has provided a glimpse of the biological basis of schizophrenia. Looking forward, CNVs are just in the initial step of our understanding of schizophrenia genetics. Other independent risk factors in schizophrenia, including those that are genetic (SNPs, INDELs, and other structural variants such as inversions and translocations), epigenetic, environmental or stochastic, are all part of the comprehensive picture of this complex psychiatric illness. In addition, the same genetic locus may harbour multiple allelic variants of different types, further complicating the issue. The genetic architecture of schizophrenia, taking into account the number of variants at a given disease locus, their frequencies, and the penetrance of the genotype combinations, will add another layer of complexity to our comprehension of this disease (Pritchard and Cox 2002) (Figure 7.1).

Improvement of genomic technologies – in particular more accurate, higher-resolution and higher-throughput CNV detection platforms as well as advancement in sequencing technologies - will facilitate the search of such schizophrenia risk factors in a near future. New bioinformatic tools for CNV analysis will also be indispensable. Better algorithms for CNV detection and genotyping are needed (McCarroll and Altshuler 2007; Barnes et al. 2008; Ionita-Laza et al. 2008a), together with improved statistical methods to look for disease association (McCarroll and Altshuler 2007). Databases for CNV information will also be important in providing baseline comparison as the field is awash in CNV discoveries. For instance, the Database of Genomic Variants (DGV; http://projects.tcag.ca/variation/) and DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources; https://decipher.sanger.ac.uk/) are two public repositories collating CNV discoveries in normal and affected individuals

respectively. Databases dedicated to schizophrenia CNVs, such as one analogous to the Autism Chromosome Rearrangement Database (http://projects.tcag.ca/autism/), will also be valuable to the schizophrenia research and clinical community.

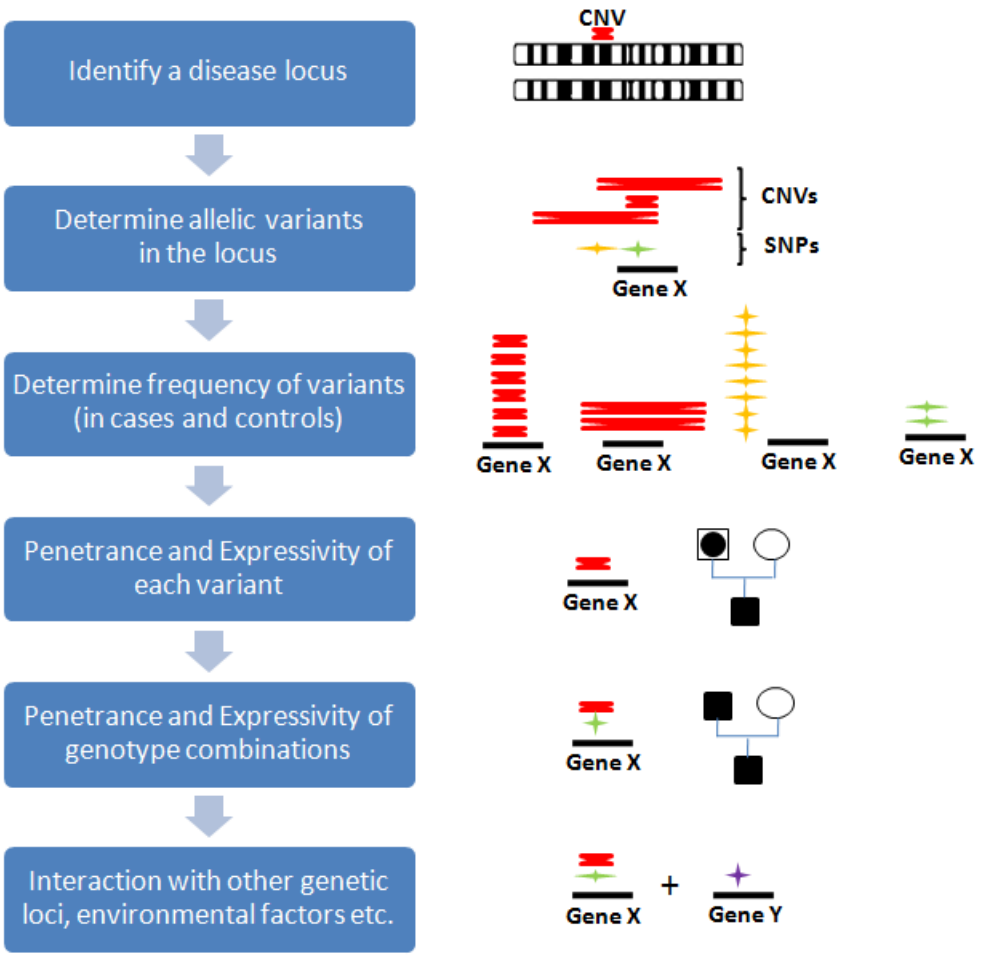


Figure 7.1 Identification of CNVs as disease risk loci for further characterization.

Finally, genetic studies alone are not sufficient to unravel such a complex human trait as schizophrenia. Multiple parallel approaches, including system biology, neuroscience and animal models, will benefit from the results of CNV investigations to deliver more valuable insights into schizophrenia. Candidate genes identified in CNV studies described in this thesis or in literature are prime candidates to be tested in animal models to assess behaviour and cognitive phenotypes, for example, or in neuronal cell cultures to confirm electrophysiological and neurological relevance. As such, the work described in this dissertation is part of the Genes2Cognition Program (http://www.genes2cognition.org/), and CNV findings described herein will be further investigated within the integrated, multi-faceted research program (Grant 2003; Croning et al. 2008).

References

Abi-Dargham, A., Gil, R., Krystal, J., Baldwin, R.M., Seibyl, J.P., Bowers, M., van Dyck, C.H., Charney, D.S., Innis, R.B., and Laruelle, M. 1998. Increased striatal dopamine transmission in schizophrenia: confirmation in a second cohort. *Am J Psychiatry* **155**(6): 761-767.

Abrahams, B.S. and Geschwind, D.H. 2008. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* **9**(5): 341-355.

Adams, D.J., Dermitzakis, E.T., Cox, T., Smith, J., Davies, R., Banerjee, R., Bonfield, J., Mullikin, J.C., Chung, Y.J., Rogers, J., and Bradley, A. 2005. Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nat Genet* **37**(5): 532-536.

Agulhon, C., Abitbol, M., Bertrand, D., and Malafosse, A. 1999. Localization of mRNA for CHRNA7 in human fetal brain. *Neuroreport* **10**(11): 2223-2227.

Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E., Hodges, M.D., Bhangal, G., Patel, S.G., Sheehan-Rooney, K., Duda, M., Cook, P.R., Evans, D.J., Domin, J., Flint, J., Boyle, J.J., Pusey, C.D., and Cook, H.T. 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**(7078): 851-855.

Albertson, D.G. and Pinkel, D. 2003. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* **12 Spec No 2**: R145-152.

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., and Lander, E.S. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**(6803): 513-516.

American Psychiatric Association. 1994. *Diagnostic and statistical manual of mental disorders : DSM-IV*. American Psychiatric Association, Washington, DC.

Andreasen, N.C. 1995. Symptoms, signs, and diagnosis of schizophrenia. *Lancet* **346**(8973): 477-481.

Angeloni, D., Wei, M.H., and Lerman, M.I. 1999. Two single nucleotide polymorphisms (SNPs) in the CALL gene for association studies with IQ. *Psychiatr Genet* **9**(3): 165-167.

Antshel, K.M., Aneja, A., Strunge, L., Peebles, J., Fremont, W.P., Stallone, K., Abdulsabur, N., Higgins, A.M., Shprintzen, R.J., and Kates, W.R. 2007. Autistic spectrum disorders in velo-cardio facial syndrome (22q11.2 deletion). *J Autism Dev Disord* **37**(9): 1776-1786.

Armour, J.A., Anttinen, T., May, C.A., Vega, E.E., Sajantila, A., Kidd, J.R., Kidd, K.K., Bertranpetit, J., Paabo, S., and Jeffreys, A.J. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* **13**(2): 154-160.

Armour, J.A., Palla, R., Zeeuwen, P.L., den Heijer, M., Schalkwijk, J., and Hollox, E.J. 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res* **35**(3): e19.

Armour, J.A., Sismani, C., Patsalis, P.C., and Cross, G. 2000. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res* **28**(2): 605-609.

Autism Genome Project Consortium. 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39**(3): 319-328.

Badner, J.A. and Gershon, E.S. 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* **7**(4): 405-411.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* **297**(5583): 1003-1007.

Baldwin, E.L., Lee, J.Y., Blake, D.M., Bunke, B.P., Alexander, C.R., Kogan, A.L., Ledbetter, D.H., and Martin, C.L. 2008. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet Med* **10**(6): 415-429.

Ballard, T.M., Pauly-Evers, M., Higgins, G.A., Ouagazzal, A.M., Mutel, V., Borroni, E., Kemp, J.A., Bluethmann, H., and Kew, J.N. 2002. Severe impairment of NMDA receptor function in mice carrying targeted point mutations in the glycine binding site results in drug-resistant nonhabituating hyperactivity. *J Neurosci* **22**(15): 6713-6723.

Barnes, C., Plagnol, V., Fitzgerald, T., Redon, R., Marchini, J., Clayton, D., and Hurles, M.E. 2008. A robust statistical method for case-control association testing with copy number variation. *Nat Genet* **40**(10): 1245-1252.

Bauman, J.G., Wiegant, J., Borst, P., and van Duijn, P. 1980. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp Cell Res* **128**(2): 485-490.

Bauters, M., Van Esch, H., Friez, M.J., Boespflug-Tanguy, O., Zenker, M., Vianna-Morgante, A.M., Rosenberg, C., Ignatius, J., Raynaud, M., Hollanders, K., Govaerts, K., Vandenreijt, K., Niel, F., Blanc, P., Stevenson, R.E., Fryns, J.P., Marynen, P., Schwartz, C.E., and Froyen, G. 2008. Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res* **18**(6): 847-858.

Bertelsen, A. 2002. Schizophrenia and related disorders: experience with current diagnostic systems. *Psychopathology* **35**(2-3): 89-93.

Biederer, T. and Sudhof, T.C. 2000. Mints as adaptors. Direct binding to neurexins and recruitment of munc18. *J Biol Chem* **275**(51): 39803-39806.

Birchler, J.A., Riddle, N.C., Auger, D.L., and Veitia, R.A. 2005. Dosage balance in gene regulation: biological implications. *Trends Genet* **21**(4): 219-226.

Blackwood, D.H., Fordyce, A., Walker, M.T., St Clair, D.M., Porteous, D.J., and Muir, W.J. 2001. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet* **69**(2): 428-433.

Bleuler, E. 1911. *Dementia praecox or the group of schizophrenias*. International Universities Press, New York.

Borroni, B., Yancopoulou, D., Tsutsui, M., Padovani, A., Sawcer, S.J., Hodges, J.R., and Spillantini, M.G. 2005. Association between tau H2 haplotype and age at onset in frontotemporal dementia. *Arch Neurol* **62**(9): 1419-1422.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., and Cavalli-Sforza, L.L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**(6470): 455-457.

Bradshaw, N.J., Ogawa, F., Antolin-Fontes, B., Chubb, J.E., Carlyle, B.C., Christie, S., Claessens, A., Porteous, D.J., and Millar, J.K. 2008. DISC1, PDE4B, and NDE1 at the centrosome and synapse. *Biochem Biophys Res Commun*.

Braff, D.L., Greenwood, T.A., Swerdlow, N.R., Light, G.A., and Schork, N.J. 2008. Advances in endophenotyping schizophrenia. *World Psychiatry* **7**(1): 11-18.

Brambilla, R. 2003. Targeting Ras/ERK signalling in the striatum: will it help? *Mol Psychiatry* **8**(4): 366-368.

Brandon, N.J. 2007. Dissecting DISC1 function through protein-protein interactions. *Biochem Soc Trans* **35**(Pt 5): 1283-1286.

Bray, N.J. 2008. Gene expression in the etiology of schizophrenia. *Schizophr Bull* **34**(3): 412-418.

Bridges, C.B. 1936. The Bar "Gene" a Duplication. *Science* **83**(2148): 210-211.

Brindley, M. 2008. Shock therapy's still being used in Wales. In *WalesOnline*. Media Wales Ltd., Wales.

Buckley, P.G., Mantripragada, K.K., Benetkiewicz, M., Tapia-Paez, I., Diaz De Stahl, T., Rosenquist, M., Ali, H., Jarbo, C., De Bustos, C., Hirvela, C., Sinder Wilen, B., Fransson, I., Thyr, C., Johnsson, B.I., Bruder, C.E., Menzel, U., Hergersberg, M., Mandahl, N., Blennow, E., Wedell, A., Beare, D.M., Collins, J.E., Dunham, I., Albertson, D., Pinkel, D., Bastian, B.C., Faruqi, A.F., Lasken, R.S., Ichimura, K., Collins, V.P., and Dumanski, J.P. 2002. A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum Mol Genet* **11**(25): 3221-3229.

Buhusi, M., Midkiff, B.R., Gates, A.M., Richter, M., Schachner, M., and Maness, P.F. 2003. Close homolog of L1 is an enhancer of integrin-mediated cell migration. *J Biol Chem* **278**(27): 25024-25031.

Burmeister, M., McInnis, M.G., and Zollner, S. 2008. Psychiatric genetics: progress amid controversy. *Nat Rev Genet* **9**(7): 527-540.

Bustillo, J., Lauriello, J., Horan, W., and Keith, S. 2001. The psychosocial treatment of schizophrenia: an update. *Am J Psychiatry* **158**(2): 163-175.

Caffrey, T.M., Joachim, C., Paracchini, S., Esiri, M.M., and Wade-Martins, R. 2006. Haplotype-specific expression of exon 10 at the human MAPT locus. *Hum Mol Genet* **15**(24): 3529-3537.

Caffrey, T.M., Joachim, C., and Wade-Martins, R. 2007. Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus. *Neurobiol Aging*.

Caldwell, H.K., Stephens, S.L., and Young, W.S., 3rd. 2008. Oxytocin as a natural antipsychotic: a study using oxytocin knockout mice. *Mol Psychiatry*.

Cantor, R.M. and Geschwind, D.H. 2008. Schizophrenia: genome, interrupted. *Neuron* **58**(2): 165-167.

Cardno, A.G. and Gottesman, II. 2000. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* **97**(1): 12-17.

Carlson, C.S., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. 2004. Mapping complex disease loci in whole-genome association studies. *Nature* **429**(6990): 446-452.

Carlsson, A., Waters, N., and Carlsson, M.L. 1999. Neurotransmitter interactions in schizophrenia--therapeutic implications. *Biol Psychiatry* **46**(10): 1388-1395.

Carson, R., Craig, D., McGuinness, B., Johnston, J.A., O'Neill, F.A., Passmore, A.P., and Ritchie, C.W. 2008. Alpha7 nicotinic acetylcholine receptor gene and reduced risk of Alzheimer's disease. *J Med Genet* **45**(4): 244-248.

Carter, N.P. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* **39**(7 Suppl): S16-21.

Cascorbi, I. 2006. Role of pharmacogenetics of ATP-binding cassette transporters in the pharmacokinetics of drugs. *Pharmacol Ther* **112**(2): 457-473.

Cassidy, S.B., Dykens, E., and Williams, C.A. 2000. Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am J Med Genet* **97**(2): 136-146.

Cerletti, U. 1954. Electroshock therapy. *J Clin Exp Psychopathol* **15**(3): 191-217.

Chaisuksunt, V., Zhang, Y., Anderson, P.N., Campbell, G., Vaudano, E., Schachner, M., and Lieberman, A.R. 2000. Axonal regeneration from CNS neurons in the cerebellum and brainstem of adult rats: correlation with the patterns of expression and distribution of messenger RNAs for L1, CHL1, c-jun and growth-associated protein-43. *Neuroscience* **100**(1): 87-108.

Chan, S.L., Kim, W.S., Kwok, J.B., Hill, A.F., Cappai, R., Rye, K.A., and Garner, B. 2008. ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. *J Neurochem* **106**(2): 793-804.

Chance, P.F. and Fischbeck, K.H. 1994. Molecular genetics of Charcot-Marie-Tooth disease and related neuropathies. *Hum Mol Genet* **3 Spec No**: 1503-1507.

Chartier-Harlin, M.C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M., Waucquier, N., Defebvre, L., Amouyel, P., Farrer, M., and Destee, A. 2004. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**(9440): 1167-1169.

Chen, Q.Y., Chen, Q., Feng, G.Y., Lindpaintner, K., Chen, Y., Sun, X., Chen, Z., Gao, Z., Tang, J., and He, L. 2005. Case-control association study of the close homologue of L1 (CHL1) gene and schizophrenia in the Chinese population. *Schizophr Res* **73**(2-3): 269-274.

Chen, S., Mantei, N., Dong, L., and Schachner, M. 1999. Prevention of neuronal cell death by neural adhesion molecules L1 and CHL1. *J Neurobiol* **38**(3): 428-439.

Chin, J., Palop, J.J., Puolivali, J., Massaro, C., Bien-Ly, N., Gerstein, H., Scearce-Levie, K., Masliah, E., and Mucke, L. 2005. Fyn kinase induces synaptic and cognitive impairments in a transgenic mouse model of Alzheimer's disease. *J Neurosci* **25**(42): 9694-9703.

Christian, S.L., Robinson, W.P., Huang, B., Mutirangura, A., Line, M.R., Nakao, M., Surti, U., Chakravarti, A., and Ledbetter, D.H. 1995. Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. *Am J Hum Genet* **57**(1): 40-48.

Chubb, J.E., Bradshaw, N.J., Soares, D.C., Porteous, D.J., and Millar, J.K. 2008. The DISC locus in psychiatric illness. *Mol Psychiatry* **13**(1): 36-64.

Clapcote, S.J., Lipina, T.V., Millar, J.K., Mackie, S., Christie, S., Ogawa, F., Lerch, J.P., Trimble, K., Uchiyama, M., Sakuraba, Y., Kaneda, H., Shiroishi, T., Houslay, M.D., Henkelman, R.M., Sled, J.G., Gondo, Y., Porteous, D.J., and Roder, J.C. 2007. Behavioral phenotypes of Disc1 missense mutations in mice. *Neuron* **54**(3): 387-402.

Collins, F.S., Guyer, M.S., and Charkravarti, A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**(5343): 1580-1581.

Collins, M.O., Husi, H., Yu, L., Brandon, J.M., Anderson, C.N., Blackstock, W.P., Choudhary, J.S., and Grant, S.G. 2006. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J Neurochem* **97 Suppl 1**: 16-23.

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., and Pritchard, J.K. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**(1): 75-81.

Conrad, D.F. and Hurles, M.E. 2007. The population genetics of structural variation. *Nat Genet* **39**(7 Suppl): S30-36.

Cook, E.H., Jr. and Scherer, S.W. 2008. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455**(7215): 919-923.

Cooper, G.M., Nickerson, D.A., and Eichler, E.E. 2007. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* **39**(7 Suppl): S22-29.

Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E., and Nickerson, D.A. 2008. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* **40**(10): 1199-1203.

Couturier, S., Bertrand, D., Matter, J.M., Hernandez, M.C., Bertrand, S., Millar, N., Valera, S., Barkas, T., and Ballivet, M. 1990. A neuronal nicotinic acetylcholine receptor subunit (alpha 7) is developmentally regulated and forms a homo-oligomeric channel blocked by alpha-BTX. *Neuron* **5**(6): 847-856.

Craddock, N., O'Donovan, M.C., and Owen, M.J. 2006. Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophr Bull* **32**(1): 9-16.

-. 2007. Phenotypic and genetic complexity of psychosis. Invited commentary on ... Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* **190**: 200-203.

Craig, A.M. and Kang, Y. 2007. Neurexin-neuroligin signaling in synapse development. *Curr Opin Neurobiol* **17**(1): 43-52.

Craig, J.M. and Bickmore, W.A. 1993. Chromosome bands--flavours to savour. *Bioessays* **15**(5): 349-354.

Creese, I., Burt, D.R., and Snyder, S.H. 1976. Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science* **192**(4238): 481-483.

Croning, M.D., Marshall, M.C., McLaren, P., Armstrong, J.D., and Grant, S.G. 2008. G2Cdb: the Genes to Cognition database. *Nucleic Acids Res*.

Cruts, M., Rademakers, R., Gijselinck, I., van der Zee, J., Dermaut, B., de Pooter, T., de Rijk, P., Del-Favero, J., and van Broeckhoven, C. 2005. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. *Hum Mol Genet* **14**(13): 1753-1762.

Curzon, P., Anderson, D.J., Nikkel, A.L., Fox, G.B., Gopalakrishnan, M., Decker, M.W., and Bitner, R.S. 2006. Antisense knockdown of the rat alpha7 nicotinic acetylcholine receptor produces spatial memory impairment. *Neurosci Lett* **410**(1): 15-19.

Dani, J.A. and Bertrand, D. 2007. Nicotinic acetylcholine receptors and nicotinic cholinergic mechanisms of the central nervous system. *Annu Rev Pharmacol Toxicol* **47**: 699-729.

Davey, F., Hill, M., Falk, J., Sans, N., and Gunn-Moore, F.J. 2005. Synapse associated protein 102 is a novel binding partner to the cytoplasmic terminus of neurone-glial related cell adhesion molecule. *J Neurochem* **94**(5): 1243-1253.

Davis, R.L. and Dauwalder, B. 1991. The Drosophila dunce locus: learning and memory genes in the fly. *Trends Genet* **7**(7): 224-229.

Dazzan, P. and Murray, R.M. 1999. Schizophrenia is (not simply) a neurodevelopmental disorder. *Epidemiol Psichiatr Soc* **8**(4): 235-241.

De Luca, V., Likhodi, O., Van Tol, H.H., Kennedy, J.L., and Wong, A.H. 2006. Regulation of alpha7-nicotinic receptor subunit and alpha7-like gene expression in the prefrontal cortex of patients with bipolar disorder and schizophrenia. *Acta Psychiatr Scand* **114**(3): 211-215.

de Silva, R., Weiler, M., Morris, H.R., Martin, E.R., Wood, N.W., and Lees, A.J. 2001. Strong association of a novel Tau promoter haplotype in progressive supranuclear palsy. *Neurosci Lett* **311**(3): 145-148.

de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben-Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., Laderman, S., Froguel, P., and Blakemore, A.I. 2007. Array CGH analysis of

copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum Mol Genet* **16**(23): 2783-2794.

de Smith, A.J., Walters, R.G., Coin, L.J., Steinfeld, I., Yakhini, Z., Sladek, R., Froguel, P., and Blakemore, A.I. 2008. Small deletion variants have stable breakpoints commonly associated with alu elements. *PLoS ONE* **3**(8): e3104.

Dean, C., Scholl, F.G., Choih, J., DeMaria, S., Berger, J., Isacoff, E., and Scheiffele, P. 2003. Neurexin mediates the assembly of presynaptic terminals. *Nat Neurosci* **6**(7): 708-716.

Dean, M., Hamon, Y., and Chimini, G. 2001. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* **42**(7): 1007-1017.

Deary, I.J., Gow, A.J., Taylor, M.D., Corley, J., Brett, C., Wilson, V., Campbell, H., Whalley, L.J., Visscher, P.M., Porteous, D.J., and Starr, J.M. 2007. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**: 28.

Deeb, S.S. 2005. The molecular basis of variation in human color vision. *Clin Genet* **67**(5): 369-377.

Delabar, J.M., Goldgaber, D., Lamour, Y., Nicole, A., Huret, J.L., de Grouchy, J., Brown, P., Gajdusek, D.C., and Sinet, P.M. 1987. Beta amyloid gene duplication in Alzheimer's disease and karyotypically normal Down syndrome. *Science* **235**(4794): 1390-1392.

Dempster, E.L., Toulopoulou, T., McDonald, C., Bramon, E., Walshe, M., Wickham, H., Sham, P.C., Murray, R.M., and Collier, D.A. 2006. Episodic memory performance predicted by the 2bp deletion in exon 6 of the "alpha 7-like" nicotinic receptor subunit gene. *Am J Psychiatry* **163**(10): 1832-1834.

Demyanenko, G.P., Schachner, M., Anton, E., Schmid, R., Feng, G., Sanes, J., and Maness, P.F. 2004. Close homolog of L1 modulates area-specific neuronal positioning and dendrite orientation in the cerebral cortex. *Neuron* **44**(3): 423-437.

Dermitzakis, E.T. and Clark, A.G. 2001. Differential selection after duplication in mammalian developmental genes. *Mol Biol Evol* **18**(4): 557-562.

Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. 2005. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* **76**(5): 750-762.

Dimitropoulos, A. and Schultz, R.T. 2007. Autistic-like symptomatology in Prader-Willi syndrome: a review of recent findings. *Curr Psychiatry Rep* **9**(2): 159-164.

Doheny, D.O., Brin, M.F., Morrison, C.E., Smith, C.J., Walker, R.H., Abbasi, S., Muller, B., Garrels, J., Liu, L., De Carvalho Aguiar, P., Schilling, K., Kramer, P., De Leon, D., Raymond, D., Saunders-Pullman, R., Klein, C., Bressman, S.B., Schmand, B., Tijssen, M.A., Ozelius, L.J., and Silverman, J.M. 2002. Phenotypic features of myoclonus-dystonia in three kindreds. *Neurology* **59**(8): 1187-1196.

Dopman, E.B. and Hartl, D.L. 2007. A portrait of copy-number polymorphism in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **104**(50): 19920-19925.

Duan, S., Gao, R., Xing, Q., Du, J., Liu, Z., Chen, Q., Wang, H., Feng, G., and He, L. 2005. A family-based association study of schizophrenia with polymorphisms at three candidate genes. *Neurosci Lett* **379**(1): 32-36.

Duan, X., Chang, J.H., Ge, S., Faulkner, R.L., Kim, J.Y., Kitabatake, Y., Liu, X.B., Yang, C.H., Jordan, J.D., Ma, D.K., Liu, C.Y., Ganesan, S., Cheng, H.J., Ming, G.L., Lu, B., and Song, H. 2007. Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. *Cell* **130**(6): 1146-1158.

Duncan, G., Miyamoto, S., Gu, H., Lieberman, J., Koller, B., and Snouwaert, J. 2002. Alterations in regional brain metabolism in genetic and pharmacological models of reduced NMDA receptor function. *Brain Res* **951**(2): 166-176.

Duncan, G.E., Moy, S.S., Perez, A., Eddy, D.M., Zinzow, W.M., Lieberman, J.A., Snouwaert, J.N., and Koller, B.H. 2004. Deficits in sensorimotor gating and tests of social behavior in a genetic model of reduced NMDA receptor function. *Behav Brain Res* **153**(2): 507-519.

Eastwood, S.L., Burnet, P.W., and Harrison, P.J. 2005. Decreased hippocampal expression of the susceptibility gene PPP3CC and other calcineurin subunits in schizophrenia. *Biol Psychiatry* **57**(7): 702-710.

Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S., Magenis, E., Shprintzen, R.J., and Morrow, B.E. 1999. A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet* **8**(7): 1157-1167.

Edenberg, H.J., Foroud, T., Conneally, P.M., Sorbel, J.J., Carr, K., Crose, C., Willig, C., Zhao, J., Miller, M., Bowman, E., Mayeda, A., Rau, N.L., Smiley, C., Rice, J.P., Goate, A., Reich, T., Stine, O.C., McMahon, F., DePaulo, J.R., Meyers, D., Detera-Wadleigh, S.D., Goldin, L.R., Gershon, E.S., Blehar, M.C., and Nurnberger, J.I., Jr. 1997. Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22. *Am J Med Genet* **74**(3): 238-246.

Egan, C.M., Sridhar, S., Wigler, M., and Hall, I.M. 2007. Recurrent DNA copy number variation in the laboratory mouse. *Nat Genet* **39**(11): 1384-1389.

Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., Goldman, D., and Weinberger, D.R. 2001. Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci U S A* **98**(12): 6917-6922.

Eichler, E.E., Nickerson, D.A., Altshuler, D., Bowcock, A.M., Brooks, L.D., Carter, N.P., Church, D.M., Felsenfeld, A., Guyer, M., Lee, C., Lupski, J.R., Mullikin, J.C., Pritchard, J.K., Sebat, J., Sherry, S.T., Smith, D., Valle, D., and Waterston, R.H. 2007. Completing the map of human genetic variation. *Nature* **447**(7141): 161-165.

Ekelund, J., Hovatta, I., Parker, A., Paunio, T., Varilo, T., Martin, R., Suhonen, J., Ellonen, P., Chan, G., Sinsheimer, J.S., Sobel, E., Juvonen, H., Arajarvi, R., Partonen, T., Suvisaari, J., Lonnqvist, J., Meyer, J., and Peltonen, L. 2001. Chromosome 1 loci in Finnish schizophrenia families. *Hum Mol Genet* **10**(15): 1611-1617.

Elmslie, F.V., Rees, M., Williamson, M.P., Kerr, M., Kjeldsen, M.J., Pang, K.A., Sundqvist, A., Friis, M.L., Chadwick, D., Richens, A., Covanis, A., Santos, M., Arzimanoglou, A., Panayiotopoulos, C.P., Curtis, D., Whitehouse, W.P., and Gardiner, R.M. 1997. Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. *Hum Mol Genet* **6**(8): 1329-1334.

Emanuel, B.S. and Shaikh, T.H. 2001. Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet* **2**(10): 791-800.

Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O., and Long, M. 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. *Science* **320**(5883): 1629-1631.

Ermak, G., Harris, C.D., Battocchio, D., and Davies, K.J. 2006. RCAN1 (DSCR1 or Adapt78) stimulates expression of GSK-3beta. *FEBS J* **273**(10): 2100-2109.

Estivill, X. and Armengol, L. 2007. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* **3**(10): 1787-1799.

Evans, W., Fung, H.C., Steele, J., Eerola, J., Tienari, P., Pittman, A., Silva, R., Myers, A., Vrieze, F.W., Singleton, A., and Hardy, J. 2004. The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci Lett* **369**(3): 183-185.

Evins, A.E., Amico, E., Posever, T.A., Toker, R., and Goff, D.C. 2002. D-Cycloserine added to risperidone in patients with primary negative symptoms of schizophrenia. *Schizophr Res* **56**(1-2): 19-23.

Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M., Gough, S.C., de Smith, A., Blakemore, A.I., Froguel, P., Owen, C.J., Pearce, S.H., Teixeira, L., Guillevin, L., Graham, D.S., Pusey, C.D., Cook, H.T., Vyse, T.J., and Aitman, T.J. 2007. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* **39**(6): 721-723.

Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W., Teml, A., Schwab, M., Lichter, P., Radlwimmer, B., and Stange, E.F. 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* **79**(3): 439-448.

Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., Cook, E.H., Jr., Skinner, C., Schwartz, C.E., and Sommer, S.S. 2006. High frequency of neurexin 1beta signal peptide structural variants in patients with autism. *Neurosci Lett* **409**(1): 10-13.

Fenton, W.S., McGlashan, T.H., Victor, B.J., and Blyler, C.R. 1997. Symptoms, subtype, and suicidality in patients with schizophrenia spectrum disorders. *Am J Psychiatry* **154**(2): 199-204.

Fernandes, C.C., Pinto-Duarte, A., Ribeiro, J.A., and Sebastiao, A.M. 2008. Postsynaptic action of brain-derived neurotrophic factor attenuates alpha7 nicotinic acetylcholine receptor-mediated responses in hippocampal interneurons. *J Neurosci* **28**(21): 5611-5618.

Feuk, L., Carson, A.R., and Scherer, S.W. 2006a. Structural variation in the human genome. *Nat Rev Genet* **7**(2): 85-97.

Feuk, L., Marshall, C.R., Wintle, R.F., and Scherer, S.W. 2006b. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* **15 Spec No 1**: R57-66.

Fiegler, H., Carr, P., Douglas, E.J., Burford, D.C., Hunt, S., Scott, C.E., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I.P., and Carter, N.P. 2003. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* **36**(4): 361-374.

Fiegler, H., Redon, R., Andrews, D., Scott, C., Andrews, R., Carder, C., Clark, R., Dovey, O., Ellis, P., Feuk, L., French, L., Hunt, P., Kalaitzopoulos, D., Larkin, J., Montgomery, L., Perry, G.H., Plumb, B.W., Porter, K., Rigby, R.E., Rigler, D., Valsesia, A., Langford, C., Humphray, S.J., Scherer, S.W., Lee, C., Hurles, M.E., and Carter, N.P. 2006. Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res* **16**(12): 1566-1574.

Fiegler, H., Redon, R., and Carter, N.P. 2007. Construction and use of spotted large-insert clone DNA microarrays for the detection of genomic copy number changes. *Nat Protoc* **2**(3): 577-587.

Fine, S.E., Weissman, A., Gerdes, M., Pinto-Martin, J., Zackai, E.H., McDonald-McGinn, D.M., and Emanuel, B.S. 2005. Autism spectrum disorders and symptoms in children with molecularly confirmed 22q11.2 deletion syndrome. *J Autism Dev Disord* **35**(4): 461-470.

Flomen, R.H., Collier, D.A., Osborne, S., Munro, J., Breen, G., St Clair, D., and Makoff, A.J. 2006. Association study of CHRFAM7A copy number and 2 bp deletion polymorphisms with schizophrenia and bipolar affective disorder. *Am J Med Genet B Neuropsychiatr Genet* **141B**(6): 571-575.

Flomen, R.H., Davies, A.F., Di Forti, M., Cascia, C.L., Mackie-Ogilvie, C., Murray, R., and Makoff, A.J. 2008. The copy number variant involving part of the alpha7 nicotinic receptor gene contains a polymorphic inversion. *Eur J Hum Genet*.

Florijn, R.J., Bonden, L.A., Vrolijk, H., Wiegant, J., Vaandrager, J.W., Baas, F., den Dunnen, J.T., Tanke, H.J., van Ommen, G.J., and Raap, A.K. 1995. High-resolution DNA Fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. *Hum Mol Genet* **4**(5): 831-836.

Forozan, F., Karhu, R., Kononen, J., Kallioniemi, A., and Kallioniemi, O.P. 1997. Genome screening by comparative genomic hybridization. *Trends Genet* **13**(10): 405-409.

Fradley, R.L., O'Meara, G.F., Newman, R.J., Andrieux, A., Job, D., and Reynolds, D.S. 2005. STOP knockout and NMDA NR1 hypomorphic mice exhibit deficits in sensorimotor gating. *Behav Brain Res* **163**(2): 257-264.

Fredman, D., White, S.J., Potter, S., Eichler, E.E., Den Dunnen, J.T., and Brookes, A.J. 2004. Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet* **36**(8): 861-866.

Freedman, R., Coon, H., Myles-Worsley, M., Orr-Urtreger, A., Olincy, A., Davis, A., Polymeropoulos, M., Holik, J., Hopkins, J., Hoff, M., Rosenthal, J., Waldo, M.C., Reimherr, F., Wender, P., Yaw, J., Young, D.A., Breese, C.R., Adams, C., Patterson, D., Adler, L.E., Kruglyak, L., Leonard, S., and Byerley, W. 1997. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc Natl Acad Sci U S A* **94**(2): 587-592.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., Carter, N.P., Scherer, S.W., and Lee, C. 2006. Copy number variation: new insights in genome diversity. *Genome Res* **16**(8): 949-961.

Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelmann, L., Davis, K.L., Silverman, J.M., Brunner, H.G., van Kessel, A.G., Wijmenga, C., Ophoff, R.A., and Veltman, J.A. 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* **13**(3): 261-266.

Friedman, J.M., Baross, A., Delaney, A.D., Ally, A., Arbour, L., Armstrong, L., Asano, J., Bailey, D.K., Barber, S., Birch, P., Brown-John, M., Cao, M., Chan, S., Charest, D.L., Farnoud, N., Fernandes, N., Flibotte, S., Go, A., Gibson, W.T., Holt, R.A., Jones, S.J., Kennedy, G.C., Krzywinski, M., Langlois,

S., Li, H.I., McGillivray, B.C., Nayar, T., Pugh, T.J., Rajcan-Separovic, E., Schein, J.E., Schnerch, A., Siddiqui, A., Van Allen, M.I., Wilson, G., Yong, S.L., Zahir, F., Eydoux, P., and Marra, M.A. 2006. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* **79**(3): 500-513.

Frints, S.G., Marynen, P., Hartmann, D., Fryns, J.P., Steyaert, J., Schachner, M., Rolf, B., Craessaerts, K., Snellinx, A., Hollanders, K., D'Hooge, R., De Deyn, P.P., and Froyen, G. 2003. CALL interrupted in a patient with non-specific mental retardation: gene dosage-dependent alteration of murine brain development and behavior. *Hum Mol Genet* **12**(13): 1463-1474.

Garcia, R.A., Vasudevan, K., and Buonanno, A. 2000. The neuregulin receptor ErbB-4 interacts with PDZ-containing proteins at neuronal synapses. *Proc Natl Acad Sci U S A* **97**(7): 3596-3601.

Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., Moore, T., Jacobs, S., Meriwether, J., Choi, M.J., Kim, E.J., Walton, K., Buiting, K., Davis, A., Breese, C., Freedman, R., and Leonard, S. 1998. Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7). *Genomics* **52**(2): 173-185.

Gelder, M.G. 1996. *Oxford textbook of psychiatry*. Oxford University Press, Oxford ; New York.

Genome Structural Variation Consortium. 2008. High-resolution Copy Number Variation (CNV) Discovery. In. Wellcome Trust Sanger Institue.

Ghidoni, R., Signorini, S., Barbiero, L., Sina, E., Cominelli, P., Villa, A., Benussi, L., and Binetti, G. 2006. The H2 MAPT haplotype is associated with familial frontotemporal dementia. *Neurobiol Dis* **22**(2): 357-362.

Goff, D.C., Tsai, G., Manoach, D.S., and Coyle, J.T. 1995. Dose-finding trial of D-cycloserine added to neuroleptics for negative symptoms in schizophrenia. *Am J Psychiatry* **152**(8): 1213-1215.

Gogos, J.A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., and Karayiorgou, M. 1998. Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc Natl Acad Sci U S A* **95**(17): 9991-9996.

Gogos, J.A., Santha, M., Takacs, Z., Beck, K.D., Luine, V., Lucas, L.R., Nadler, J.V., and Karayiorgou, M. 1999. The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. *Nat Genet* **21**(4): 434-439.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W., Clark, R.A., Anderson, S.A., O'Connell R, J., Agan, B.K., Ahuja, S.S., Bologna, R., Sen, L., Dolan, M.J., and Ahuja, S.K. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**(5714): 1434-1440.

Goris, A., Williams-Gray, C.H., Clark, G.R., Foltynie, T., Lewis, S.J., Brown, J., Ban, M., Spillantini, M.G., Compston, A., Burn, D.J., Chinnery, P.F., Barker, R.A., and Sawcer, S.J. 2007. Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson's disease. *Ann Neurol* **62**(2): 145-153.

Gottesman, II and Gould, T.D. 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**(4): 636-645.

Gottesman, I. 1991. *Schizophrenia Genesis*. W.H. Freeman and Company, New York.

Grant, S.G. 2003. Systems biology in neuroscience: bridging genes to cognition. *Curr Opin Neurobiol* **13**(5): 577-582.

Grant, S.G., Marshall, M.C., Page, K.L., Cumiskey, M.A., and Armstrong, J.D. 2005. Synapse proteomics of multiprotein complexes: en route from genes to nervous system diseases. *Hum Mol Genet* **14 Spec No. 2**: R225-234.

Gribble, S.M., Kalaitzopoulos, D., Burford, D.C., Prigmore, E., Selzer, R.R., Ng, B.L., Matthews, N.S., Porter, K.M., Curley, R., Lindsay, S.J., Baptista, J., Richmond, T.A., and Carter, N.P. 2007. Ultra-high resolution array painting facilitates breakpoint sequencing. *J Med Genet* **44**(1): 51-58.

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes, J.A., Swinkels, D.W., den Heijer, M., Franke, B., Verbeek, A.L., Becker, D.M., Yanek, L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J.,

Kiemeney, L.A., Kong, A., Thorsteinsdottir, U., and Stefansson, K. 2008. Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**(5): 609-615.

Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S.A., Cook, S., Pravenec, M., Aitman, T., Jacob, H., Shull, J.D., Hubner, N., and Cuppen, E. 2008. Distribution and functional impact of DNA copy number variation in the rat. *Nat Genet* **40**(5): 538-545.

Haas, A. 1998. NSF--fusion and beyond. *Trends Cell Biol* **8**(12): 471-473.

Hafner, H. and an der Heiden, W. 1999. The course of schizophrenia in the light of modern follow-up studies: the ABC and WHO studies. *Eur Arch Psychiatry Clin Neurosci* **249 Suppl 4**: 14-26.

Hafner, H., Hambrecht, M., Loffler, W., Munk-Jorgensen, P., and Riecher-Rossler, A. 1998. Is schizophrenia a disorder of all ages? A comparison of first episodes and early course across the life-cycle. *Psychol Med* **28**(2): 351-365.

Hancock, M.L., Canetta, S.E., Role, L.W., and Talmage, D.A. 2008. Presynaptic type III neuregulin1-ErbB signaling targets {alpha}7 nicotinic acetylcholine receptors to axons. *J Cell Biol* **181**(3): 511-521.

Hardy, J., Pittman, A., Myers, A., Gwinn-Hardy, K., Fung, H.C., de Silva, R., Hutton, M., and Duckworth, J. 2005. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. *Biochem Soc Trans* **33**(Pt 4): 582-585.

Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M.F., and Weinberger, D.R. 2002. Serotonin transporter genetic variation and the response of the human amygdala. *Science* **297**(5580): 400-403.

Harris, E.C. and Barraclough, B. 1998. Excess mortality of mental disorder. *Br J Psychiatry* **173**: 11-53.

Harrison, P.J. 1999. The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain* **122 (Pt 4)**: 593-624.

Harrison, P.J. and Weinberger, D.R. 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* **10**(1): 40-68; image 45.

Hayesmoore, J.B., Bray, N.J., Cross, W.C., Owen, M.J., O'Donovan, M.C., and Morris, H.R. 2008. The effect of age and the H1c MAPT haplotype on MAPT expression in human brain. *Neurobiol Aging*.

Hebb, A.L., Robertson, H.A., and Denovan-Wright, E.M. 2008. Phosphodiesterase 10A inhibition is associated with locomotor and cognitive deficits and increased anxiety in mice. *Eur Neuropsychopharmacol* **18**(5): 339-363.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. 1996. Real time quantitative PCR. *Genome Res* **6**(10): 986-994.

Hermann, D.M. and Bassetti, C.L. 2007. Implications of ATP-binding cassette transporters for brain pharmacotherapies. *Trends Pharmacol Sci* **28**(3): 128-134.

Herr, A., Grutzmann, R., Matthaei, A., Artelt, J., Schrock, E., Rump, A., and Pilarsky, C. 2005. High-resolution analysis of chromosomal imbalances using the Affymetrix 10K SNP genotyping chip. *Genomics* **85**(3): 392-400.

Herschleb, J., Ananiev, G., and Schwartz, D.C. 2007. Pulsed-field gel electrophoresis. *Nat Protoc* **2**(3): 677-684.

Heston, L.L. 1966. Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br J Psychiatry* **112**(489): 819-825.

Higuchi, R., Dollinger, G., Walsh, P.S., and Griffith, R. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y)* **10**(4): 413-417.

Hillenbrand, R., Molthagen, M., Montag, D., and Schachner, M. 1999. The close homologue of the neural adhesion molecule L1 (CHL1): patterns of expression and promotion of neurite outgrowth by heterophilic interactions. *Eur J Neurosci* **11**(3): 813-826.

Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**(5712): 1072-1079.

Hirsch-Reinshagen, V., Zhou, S., Burgess, B.L., Bernier, L., McIsaac, S.A., Chan, J.Y., Tansley, G.H., Cohn, J.S., Hayden, M.R., and Wellington, C.L. 2004. Deficiency of ABCA1 impairs apolipoprotein E metabolism in brain. *J Biol Chem* **279**(39): 41197-41207.

Hodgkinson, C.A., Goldman, D., Jaeger, J., Persaud, S., Kane, J.M., Lipsky, R.H., and Malhotra, A.K. 2004. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet* **75**(5): 862-872.

Hoebeeck, J., Speleman, F., and Vandesompele, J. 2007. Real-time quantitative PCR as an alternative to Southern blot or fluorescence in situ hybridization for detection of gene copy number changes. *Methods Mol Biol* **353**: 205-226.

Hoeffer, C.A., Dey, A., Sachan, N., Wong, H., Patterson, R.J., Shelton, J.M., Richardson, J.A., Klann, E., and Rothermel, B.A. 2007. The Down syndrome critical region protein RCAN1 regulates long-term potentiation and memory via inhibition of phosphatase signaling. *J Neurosci* **27**(48): 13161-13172.

Hogart, A., Wu, D., Lasalle, J.M., and Schanen, N.C. 2008. The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13. *Neurobiol Dis*.

Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M., Reis, A., Armour, J.A., and Schalkwijk, J. 2008. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet* **40**(1): 23-25.

Holm, J., Hillenbrand, R., Steuber, V., Bartsch, U., Moos, M., Lubbert, H., Montag, D., and Schachner, M. 1996. Structural features of a close homologue of L1 (CHL1) in the mouse: a new member of the L1 family of neural recognition molecules. *Eur J Neurosci* **8**(8): 1613-1629.

Hong, C.J., Lai, I.C., Liou, L.L., and Tsai, S.J. 2004. Association study of the human partially duplicated alpha7 nicotinic acetylcholine receptor genetic variant with bipolar disorder. *Neurosci Lett* **355**(1-2): 69-72.

Hurles, M. 2005. How homologous recombination generates a mutable genome. *Hum Genomics* **2**(3): 179-186.

Husi, H. and Grant, S.G. 2001. Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. *J Neurochem* **77**(1): 281-291.

Husi, H., Ward, M.A., Choudhary, J.S., Blackstock, W.P., and Grant, S.G. 2000. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat Neurosci* **3**(7): 661-669.

Hutcheson, H.B., Olson, L.M., Bradford, Y., Folstein, S.E., Santangelo, S.L., Sutcliffe, J.S., and Haines, J.L. 2004. Examination of NRCAM, LRRN3, KIAA0716, and LAMB1 as autism candidate genes. *BMC Med Genet* **5**: 12.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. 2004. Detection of large-scale variation in the human genome. *Nat Genet* **36**(9): 949-951.

IHGSC. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011): 931-945.

IHMC. 2005. A haplotype map of the human genome. *Nature* **437**(7063): 1299-1320.

Inoue, K. and Lupski, J.R. 2003. Genetics and genomics of behavioral and psychiatric disorders. *Curr Opin Genet Dev* **13**(3): 303-309.

Inoue, K., Osaka, H., Thurston, V.C., Clarke, J.T., Yoneyama, A., Rosenbarker, L., Bird, T.D., Hodes, M.E., Shaffer, L.G., and Lupski, J.R. 2002. Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. *Am J Hum Genet* **71**(4): 838-853.

Insel, T.R., O'Brien, D.J., and Leckman, J.F. 1999. Oxytocin, vasopressin, and autism: is there a connection? *Biol Psychiatry* **45**(2): 145-157.

Ionita-Laza, I., Perry, G.H., Raby, B.A., Klanderman, B., Lee, C., Laird, N.M., Weiss, S.T., and Lange, C. 2008a. On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. *Genet Epidemiol* **32**(3): 273-284.

Ionita-Laza, I., Rogers, A.J., Lange, C., Raby, B.A., and Lee, C. 2008b. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics*.

Irintchev, A., Koch, M., Needham, L.K., Maness, P., and Schachner, M. 2004. Impairment of sensorimotor gating in mice deficient in the cell adhesion molecule L1 or its close homologue, CHL1. *Brain Res* **1029**(1): 131-134.

ISC. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**(7210): 237-241.

Ishiguro, H., Liu, Q.R., Gong, J.P., Hall, F.S., Ujike, H., Morales, M., Sakurai, T., Grumet, M., and Uhl, G.R. 2006. NrCAM in addiction vulnerability: positional cloning, drug-regulation, haplotype-specific expression, and altered drug reward in knockout mice. *Neuropsychopharmacology* **31**(3): 572-584.

Ishkanian, A.S., Malloff, C.A., Watson, S.K., DeLeeuw, R.J., Chi, B., Coe, B.P., Snijders, A., Albertson, D.G., Pinkel, D., Marra, M.A., Ling, V., MacAulay, C., and Lam, W.L. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* **36**(3): 299-303.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Bachner-Melman, R., Granot, R., Bornstein, G., Knafo, A., Yirmiya, N., and Ebstein, R.P. 2008. Molecular genetic studies of the arginine vasopressin 1a receptor (AVPR1a) and the oxytocin receptor (OXTR) in human behaviour: from autism to altruism with some notes in between. *Prog Brain Res* **170**: 435-449.

Iwata, Y., Nakajima, M., Yamada, K., Nakamura, K., Sekine, Y., Tsuchiya, K.J., Sugihara, G., Matsuzaki, H., Suda, S., Suzuki, K., Takei, N., Mori, N., Iwayama, Y., Takao, H., Yoshikawa, T., Riley, B., Makoff, A., Sham, P., Chen, R., and Collier, D. 2007. Linkage disequilibrium analysis of the CHRNA7 gene and its partially duplicated region in schizophrenia. *Neurosci Res* **57**(2): 194-202.

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J.E., Day, R., and Bertelsen, A. 1992. Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol Med Monogr Suppl* **20**: 1-97.

Jakobsson, J., Ekstrom, L., Inotsume, N., Garle, M., Lorentzon, M., Ohlsson, C., Roh, H.K., Carlstrom, K., and Rane, A. 2006. Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J Clin Endocrinol Metab* **91**(2): 687-693.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A., and Singleton, A.B. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**(7181): 998-1003.

Jakovcevski, I., Wu, J., Karl, N., Leshchyns'ka, I., Sytnyk, V., Chen, J., Irintchev, A., and Schachner, M. 2007. Glial scar expression of CHL1, the close homolog of the adhesion molecule L1, limits recovery after spinal cord injury. *J Neurosci* **27**(27): 7222-7233.

Jenkins, R.L., Holsopple, J.Q., and Lorr, M. 1954. Effects of prefrontal lobotomy on patients with severe chronic schizophrenia. *Am J Psychiatry* **111**(2): 84-90.

Jiang, Z., Tang, H., Ventura, M., Cardone, M.F., Marques-Bonet, T., She, X., Pevzner, P.A., and Eichler, E.E. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat Genet* **39**(11): 1361-1368.

Kahn, E. and Pokorny, A.D. 1964. Concerning the Concept of Schizophrenia. *Am J Psychiatry* **120**: 856-860.

Kalinderi, K., Fidani, L., and Bostantjopoulou, S. 2008. From 1997 to 2007: A decade journey through the H1 haplotype on 17q21 chromosome. *Parkinsonism Relat Disord*.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., and Pinkel, D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**(5083): 818-821.

Kanakry, C.G., Li, Z., Nakai, Y., Sei, Y., and Weinberger, D.R. 2007. Neuregulin-1 regulates cell adhesion via an ErbB2/phosphoinositide-3 kinase/Akt-dependent pathway: potential implications for schizophrenia and cancer. *PLoS ONE* **2**(12): e1369.

Kandel, E.R., Schwartz, J. H., Jessell, T. M. . 2000. *Principles of neural science*. McGraw-Hill Professional, New York.

Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., Lasseter, V.K., and et al. 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A* **92**(17): 7612-7616.

Keating, D.J., Dubach, D., Zanin, M.P., Yu, Y., Martin, K., Zhao, Y.F., Chen, C., Porta, S., Arbones, M.L., Mittaz, L., and Pritchard, M.A. 2008. DSCR1/RCAN1 regulates vesicle exocytosis and fusion pore kinetics: implications for Down syndrome and Alzheimer's disease. *Hum Mol Genet* **17**(7): 1020-1030.

Keith, D. and El-Husseini, A. 2008. Excitation Control: Balancing PSD-95 Function at the Synapse. *Front Mol Neurosci* **1**: 4.

Kety, S.S., Wender, P.H., Jacobsen, B., Ingraham, L.J., Jansson, L., Faber, B., and Kinney, D.K. 1994. Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. *Arch Gen Psychiatry* **51**(6): 442-455.

Khaja, R., Zhang, J., MacDonald, J.R., He, Y., Joseph-George, A.M., Wei, J., Rafiq, M.A., Qian, C., Shago, M., Pantano, L., Aburatani, H., Jones, K., Redon, R., Hurles, M., Armengol, L., Estivill, X., Mural, R.J., Lee, C., Scherer, S.W., and Feuk, L. 2006. Genome assembly comparison identifies structural variants in the human genome. *Nat Genet* **38**(12): 1413-1418.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tuzun, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., Gillett, W., Phelps, K.A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J.D., Korn, J.M., McCarroll, S.A., Altshuler, D.A., Peiffer, D.A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D.A., Mullikin, J.C., Wilson, R.K., Bruhn, L., Olson, M.V., Kaul, R., Smith, D.R., and Eichler, E.E. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**(7191): 56-64.

Kim, E. and Sheng, M. 2004. PDZ domain proteins of synapses. *Nat Rev Neurosci* **5**(10): 771-781.

Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., Descartes, M., Holt, L., Braddock, S., Troxell, R., Kaplan, L., Volkmar, F., Klin, A., Tsatsanis, K., Harris, D.J., Noens, I., Pauls, D.L., Daly, M.J., MacDonald, M.E., Morton, C.C., Quade, B.J., and Gusella, J.F. 2008a. Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet* **82**(1): 199-207.

Kim, P.M., Lam, H.Y., Urban, A.E., Korbel, J.O., Chen, X., Snyder, M., and Gerstein, M.B. 2008b. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res*.

Kim, W.S., Weickert, C.S., and Garner, B. 2008c. Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. *J Neurochem* **104**(5): 1145-1166.

Kirchhoff, M., Bisgaard, A.M., Duno, M., Hansen, F.J., and Schwartz, M. 2007. A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. *Eur J Med Genet* **50**(4): 256-263.

Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H., and Ullmann, R. 2008. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* **17**(3): 458-465.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmuller, G. 1999. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A* **96**(8): 4494-4499.

Kleinjan, D.A. and van Heyningen, V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* **76**(1): 8-32.

Kockelkorn, T.T., Arai, M., Matsumoto, H., Fukuda, N., Yamada, K., Minabe, Y., Toyota, T., Ujike, H., Sora, I., Mori, N., Yoshikawa, T., and Itokawa, M. 2004. Association study of polymorphisms in the 5' upstream region of human DISC1 gene with schizophrenia. *Neurosci Lett* **368**(1): 41-45.

Koike, H., Arguello, P.A., Kvajo, M., Karayiorgou, M., and Gogos, J.A. 2006. Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice. *Proc Natl Acad Sci U S A* **103**(10): 3693-3697.

Koldamova, R.P., Lefterov, I.M., Ikonomovic, M.D., Skoko, J., Lefterov, P.I., Isanski, B.A., DeKosky, S.T., and Lazo, J.S. 2003. 22R-hydroxycholesterol and 9-cis-retinoic acid induce ATP-binding cassette transporter A1 expression and cholesterol efflux in brain cells and decrease amyloid beta secretion. *J Biol Chem* **278**(15): 13244-13256.

Komiyama, N.H., Watabe, A.M., Carlisle, H.J., Porter, K., Charlesworth, P., Monti, J., Strathdee, D.J., O'Carroll, C.M., Martin, S.J., Morris, R.G., O'Dell, T.J., and Grant, S.G. 2002. SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J Neurosci* **22**(22): 9721-9732.

Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E., Lee, C., Scherer, S.W., Jones, K.W., Shapero, M.H., Huang, J., and Aburatani, H. 2006. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* **16**(12): 1575-1584.

Koolen, D.A., Sharp, A.J., Hurst, J.A., Firth, H.V., Knight, S.J., Goldenberg, A., Saugier-Veber, P., Pfundt, R., Vissers, L.E., Destree, A., Grisart, B., Rooms, L., Van der Aa, N., Field, M., Hackett, A., Bell, K., Nowaczyk, M.J., Mancini, G.M., Poddighe, P.J., Schwartz, C.E., Rossi, E., De Gregori, M., Antonacci-Fulton, L.L., McLellan, M.D., 2nd, Garrett, J.M., Wiechert, M.A., Miner, T.L., Crosby, S., Ciccone, R., Willatt, L., Rauch, A., Zenker, M., Aradhya, S., Manning, M.A., Strom, T.M., Wagenstaller, J., Krepischi-Santos, A.C., Vianna-Morgante, A.M., Rosenberg, C., Price, S.M., Stewart, H., Shaw-Smith, C., Brunner, H.G., Wilkie, A.O., Veltman, J.A., Zuffardi, O., Eichler, E.E., and de Vries, B.B. 2008. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J Med Genet* **45**(11): 710-720.

Koolen, D.A., Vissers, L.E., Pfundt, R., de Leeuw, N., Knight, S.J., Regan, R., Kooy, R.F., Reyniers, E., Romano, C., Fichera, M., Schinzel, A., Baumer, A., Anderlid, B.M., Schoumans, J., Knoers, N.V., van Kessel, A.G., Sistermans, E.A., Veltman, J.A., Brunner, H.G., and de Vries, B.B. 2006. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* **38**(9): 999-1001.

Korbel, J.O., Kim, P.M., Chen, X., Urban, A.E., Weissman, S., Snyder, M., and Gerstein, M.B. 2008. The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Curr Opin Struct Biol* **18**(3): 366-374.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, A.C., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., and Snyder, M. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**(5849): 420-426.

Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., Lee, C., Nizzari, M.M., Gabriel, S.B., Purcell, S., Daly, M.J., and Altshuler, D. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* **40**(10): 1253-1260.

Koros, E. and Dorner-Ciossek, C. 2007. The role of glycogen synthase kinase-3beta in schizophrenia. *Drug News Perspect* **20**(7): 437-445.

Kraepelin, E. 1919. *Dementia Praecox and Paraphrenia*. Krieger Publishing Co., Huntington, NY.

Krystal, J.H., Karper, L.P., Seibyl, J.P., Freeman, G.K., Delaney, R., Bremner, J.D., Heninger, G.R., Bowers, M.B., Jr., and Charney, D.S. 1994. Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch Gen Psychiatry* **51**(3): 199-214.

Kurotaki, N., Stankiewicz, P., Wakui, K., Niikawa, N., and Lupski, J.R. 2005. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum Mol Genet* **14**(4): 535-542.

Kurumaji, A., Nomoto, H., Okano, T., and Toru, M. 2001. An association study between polymorphism of L1CAM gene and schizophrenia in a Japanese sample. *Am J Med Genet* **105**(1): 99-104.

Lachman, H.M., Kelsoe, J.R., Remick, R.A., Sadovnick, A.D., Rapaport, M.H., Lin, M., Pazur, B.A., Roe, A.M., Saito, T., and Papolos, D.F. 1997. Linkage studies suggest a possible locus for bipolar disorder near the velo-cardio-facial syndrome region on chromosome 22. *Am J Med Genet* **74**(2): 121-128.

Lachman, H.M., Papolos, D.F., Saito, T., Yu, Y.M., Szumlanski, C.L., and Weinshilboum, R.M. 1996. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics* **6**(3): 243-250.

Lahti, A.C., Holcomb, H.H., Medoff, D.R., and Tamminga, C.A. 1995. Ketamine activates psychosis and alters limbic blood flow in schizophrenia. *Neuroreport* **6**(6): 869-872.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. 2001a. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**(6855): 519-523.

Lai, I.C., Hong, C.J., and Tsai, S.J. 2001b. Association study of a nicotinic receptor variant with schizophrenic disorders. *Neuropsychobiology* **43**(1): 15-18.

-. 2001c. Association study of nicotinic-receptor variants and major depressive disorder. *J Affect Disord* **66**(1): 79-82.

Lander, E.S. Linton, L.M. Birren, B. Nusbaum, C. Zody, M.C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. Funke, R. Gage, D. Harris, K. Heaford, A. Howland, J. Kann, L. Lehoczky, J. LeVine, R. McEwan, P. McKernan, K. Meldrim, J. Mesirov, J.P. Miranda, C. Morris, W. Naylor, J. Raymond, C. Rosetti, M. Santos, R. Sheridan, A. Sougnez, C. Stange-Thomann, N. Stojanovic, N. Subramanian, A. Wyman, D. Rogers, J. Sulston, J. Ainscough, R. Beck, S. Bentley, D. Burton, J. Clee, C. Carter, N. Coulson, A. Deadman, R. Deloukas, P. Dunham, A. Dunham, I. Durbin, R. French, L. Grafham, D. Gregory, S. Hubbard, T. Humphray, S. Hunt, A. Jones, M. Lloyd, C. McMurray, A. Matthews, L. Mercer, S. Milne, S. Mullikin, J.C. Mungall, A. Plumb, R. Ross, M. Shownkeen, R. Sims, S. Waterston, R.H. Wilson, R.K. Hillier, L.W. McPherson, J.D. Marra, M.A. Mardis, E.R. Fulton, L.A. Chinwalla, A.T. Pepin, K.H. Gish, W.R. Chissoe, S.L. Wendl, M.C. Delehaunty, K.D. Miner, T.L. Delehaunty, A. Kramer, J.B. Cook, L.L. Fulton, R.S. Johnson, D.L. Minx, P.J. Clifton, S.W. Hawkins, T. Branscomb, E. Predki, P. Richardson, P. Wenning, S. Slezak, T. Doggett, N. Cheng, J.F. Olsen, A. Lucas, S. Elkin, C. Uberbacher, E. Frazier, M. Gibbs, R.A. Muzny, D.M. Scherer, S.E. Bouck, J.B. Sodergren, E.J. Worley, K.C. Rives, C.M. Gorrell, J.H. Metzker, M.L. Naylor, S.L. Kucherlapati, R.S. Nelson, D.L. Weinstock, G.M. Sakaki, Y. Fujiyama, A. Hattori, M. Yada, T. Toyoda, A. Itoh, T. Kawagoe, C. Watanabe, H. Totoki, Y. Taylor, T. Weissenbach, J. Heilig, R. Saurin, W. Artiguenave, F. Brottier, P. Bruls, T. Pelletier, E. Robert, C. Wincker, P. Smith, D.R. Doucette-Stamm, L. Rubenfield, M. Weinstock, K. Lee, H.M. Dubois, J. Rosenthal, A. Platzer, M. Nyakatura, G. Taudien, S. Rump, A. Yang, H. Yu, J. Wang, J. Huang, G. Gu, J. Hood, L. Rowen, L. Madan, A. Qin, S. Davis, R.W. Federspiel, N.A. Abola, A.P. Proctor, M.J. Myers, R.M. Schmutz, J. Dickson, M. Grimwood, J. Cox, D.R. Olson, M.V. Kaul, R. Shimizu, N. Kawasaki, K. Minoshima, S. Evans, G.A. Athanasiou, M. Schultz, R. Roe, B.A. Chen, F. Pan, H. Ramser, J. Lehrach, H. Reinhardt, R. McCombie, W.R. de la Bastide, M. Dedhia, N. Blocker, H. Hornischer, K. Nordsiek, G. Agarwala, R. Aravind, L. Bailey, J.A. Bateman, A. Batzoglou, S. Birney, E. Bork, P. Brown, D.G. Burge, C.B. Cerutti, L. Chen, H.C. Church, D. Clamp, M. Copley, R.R. Doerks, T. Eddy, S.R. Eichler, E.E. Furey, T.S. Galagan, J. Gilbert, J.G. Harmon, C. Hayashizaki, Y. Haussler, D. Hermjakob, H. Hokamp, K. Jang, W. Johnson, L.S. Jones, T.A. Kasif, S. Kaspryzk, A. Kennedy, S. Kent, W.J. Kitts, P. Koonin, E.V. Korf, I. Kulp, D. Lancet, D. Lowe, T.M. McLysaght, A. Mikkelsen, T. Moran, J.V. Mulder, N. Pollara, V.J. Ponting, C.P. Schuler, G. Schultz, J. Slater, G. Smit, A.F. Stupka, E. Szustakowski, J. Thierry-Mieg, D. Thierry-Mieg, J. Wagner, L. Wallis, J. Wheeler, R. Williams, A. Wolf, Y.I. Wolfe, K.H. Yang, S.P. Yeh, R.F. Collins, F. Guyer, M.S. Peterson, J. Felsenfeld, A. Wetterstrand, K.A. Patrinos, A. Morgan, M.J. de Jong, P. Catanese, J.J. Osoegawa, K. Shizuya, H. Choi, S. and Chen, Y.J. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Laruelle, M., Kegeles, L.S., and Abi-Dargham, A. 2003. Glutamate, dopamine, and schizophrenia: from pathophysiology to treatment. *Ann N Y Acad Sci* **1003**: 138-158.

Laumonnier, F., Cuthbert, P.C., and Grant, S.G. 2007. The role of neuronal complexes in human X-linked brain diseases. *Am J Hum Genet* **80**(2): 205-220.

Laurendeau, I., Bahuau, M., Vodovar, N., Larramendy, C., Olivi, M., Bieche, I., Vidaud, M., and Vidaud, D. 1999. TaqMan PCR-based gene dosage assay for predictive testing in individuals from a cancer family with INK4 locus haploinsufficiency. *Clin Chem* **45**(7): 982-986.

Lavery, G.G., Walker, E.A., Turan, N., Rogoff, D., Ryder, J.W., Shelton, J.M., Richardson, J.A., Falciani, F., White, P.C., Stewart, P.M., Parker, K.L., and McMillan, D.R. 2008. Deletion of hexose-6-phosphate dehydrogenase activates the unfolded protein response pathway and induces skeletal myopathy. *J Biol Chem* **283**(13): 8453-8461.

Lee, C., Iafrate, A.J., and Brothman, A.R. 2007a. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* **39**(7 Suppl): S48-54.

Lee, J.A., Carvalho, C.M., and Lupski, J.R. 2007b. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**(7): 1235-1247.

Lee, J.A. and Lupski, J.R. 2006. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52**(1): 103-121.

Lee, J.A., Madrid, R.E., Sperle, K., Ritterson, C.M., Hobson, G.M., Garbern, J., Lupski, J.R., and Inoue, K. 2006. Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann Neurol* **59**(2): 398-403.

Lee, P.R., Brady, D.L., Shapiro, R.A., Dorsa, D.M., and Koenig, J.I. 2005. Social interaction deficits caused by chronic phencyclidine administration are reversed by oxytocin. *Neuropsychopharmacology* **30**(10): 1883-1894.

Lencz, T., Morgan, T.V., Athanasiou, M., Dain, B., Reed, C.R., Kane, J.M., Kucherlapati, R., and Malhotra, A.K. 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* **12**(6): 572-580.

Leonard, S. and Bertrand, D. 2001. Neuronal nicotinic receptors: from structure to function. *Nicotine Tob Res* **3**(3): 203-223.

Leonard, S., Gault, J., Moore, T., Hopkins, J., Robinson, M., Olincy, A., Adler, L.E., Cloninger, C.R., Kaufmann, C.A., Tsuang, M.T., Faraone, S.V., Malaspina, D., Svrakic, D.M., and Freedman, R. 1998. Further investigation of a chromosome 15 locus in schizophrenia: analysis of affected sibpairs from the NIMH Genetics Initiative. *Am J Med Genet* **81**(4): 308-312.

Leshchyns'ka, I., Sytnyk, V., Richter, M., Andreyeva, A., Puchkov, D., and Schachner, M. 2006. The adhesion molecule CHL1 regulates uncoating of clathrin-coated synaptic vesicles. *Neuron* **52**(6): 1011-1025.

Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guiducci, C., Illig, T., Hackett, R., Heid, I.M., Jacobs, K.B., Lyssenko, V., Uda, M., Boehnke, M., Chanock, S.J., Groop, L.C., Hu, F.B., Isomaa, B., Kraft, P., Peltonen, L., Salomaa, V., Schlessinger, D., Hunter, D.J., Hayes, R.B., Abecasis, G.R., Wichmann, H.E., Mohlke, K.L., and Hirschhorn, J.N. 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* **40**(5): 584-591.

Leucht, S., Kane, J.M., Kissling, W., Hamann, J., Etschel, E., and Engel, R. 2005. Clinical implications of Brief Psychiatric Rating Scale scores. *Br J Psychiatry* **187**: 366-371.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., and Venter, J.C. 2007. The diploid genome sequence of an individual human. *PLoS Biol* **5**(10): e254.

Lewis, C.M., Levinson, D.F., Wise, L.H., DeLisi, L.E., Straub, R.E., Hovatta, I., Williams, N.M., Schwab, S.G., Pulver, A.E., Faraone, S.V., Brzustowicz, L.M., Kaufmann, C.A., Garver, D.L., Gurling, H.M., Lindholm, E., Coon, H., Moises, H.W., Byerley, W., Shaw, S.H., Mesen, A., Sherrington, R., O'Neill, F.A., Walsh, D., Kendler, K.S., Ekelund, J., Paunio, T., Lonnqvist, J., Peltonen, L., O'Donovan, M.C., Owen, M.J., Wildenauer, D.B., Maier, W., Nestadt, G., Blouin, J.L., Antonarakis, S.E., Mowry, B.J., Silverman, J.M., Crowe, R.R., Cloninger, C.R., Tsuang, M.T., Malaspina, D., Harkavy-Friedman, J.M., Svrakic, D.M., Bassett, A.S., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoega, T., and Helgason, T. 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**(1): 34-48.

Li, J., Jiang, T., Mao, J.H., Balmain, A., Peterson, L., Harris, C., Rao, P.H., Havlak, P., Gibbs, R., and Cai, W.W. 2004. Genomic segmental polymorphisms in inbred mouse strains. *Nat Genet* **36**(9): 952-954.

Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**(9): 712-720.

Lin, W., Sanchez, H.B., Deerinck, T., Morris, J.K., Ellisman, M., and Lee, K.F. 2000. Aberrant development of motor axons and neuromuscular synapses in erbB2-deficient mice. *Proc Natl Acad Sci U S A* **97**(3): 1299-1304.

Lindstrom, J. 1996. Neuronal nicotinic acetylcholine receptors. *Ion Channels* **4**: 377-450.

Liu, C.M., Hwu, H.G., Lin, M.W., Ou-Yang, W.C., Lee, S.F., Fann, C.S., Wong, S.H., and Hsieh, S.H. 2001. Suggestive evidence for linkage of schizophrenia to markers at chromosome 15q13-14 in Taiwanese families. *Am J Med Genet* **105**(8): 658-661.

Liu, H., Heath, S.C., Sobin, C., Roos, J.L., Galke, B.L., Blundell, M.L., Lenane, M., Robertson, B., Wijsman, E.M., Rapoport, J.L., Gogos, J.A., and Karayiorgou, M. 2002. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proc Natl Acad Sci U S A* **99**(6): 3717-3722.

Locke, D.P., Archidiacono, N., Misceo, D., Cardone, M.F., Deschamps, S., Roe, B., Rocchi, M., and Eichler, E.E. 2003a. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol* **4**(8): R50.

Locke, D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D., and Eichler, E.E. 2003b. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13**(3): 347-357.

Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M., and Eichler, E.E. 2006. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* **79**(2): 275-290.

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S., and Hirschhorn, J.N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**(2): 177-182.

Loscher, W. and Potschka, H. 2005. Blood-brain barrier active efflux transporters: ATP-binding cassette gene family. *NeuroRx* **2**(1): 86-98.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., West, J.A., Rostan, S., Nguyen, K.C., Powers, S., Ye, K.Q., Olshen, A., Venkatraman, E., Norton, L., and Wigler, M. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**(10): 2291-2305.

Lupski, J.R. 2007. Genomic rearrangements and sporadic disease. *Nat Genet* **39**(7 Suppl): S43-47.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., Chakravarti, A., and Patel, P.I. 1991. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**(2): 219-232.

Lupski, J.R. and Stankiewicz, P. 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* **1**(6): e49.

Lupski, J.R., Stankiewicz, P.T. , ed. 2006. *Genomic disorders: the genomic basis of disease*. Springer.

Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**(5494): 1151-1155.

MacIntyre, D.J., Blackwood, D.H., Porteous, D.J., Pickard, B.S., and Muir, W.J. 2003. Chromosomal abnormalities and mental illness. *Mol Psychiatry* **8**(3): 275-287.

Malakho, S.G., Korshunov, A., Stroganova, A.M., and Poltaraus, A.B. 2008. Fast detection of MYCN copy number alterations in brain neuronal tumors by real-time PCR. *J Clin Lab Anal* **22**(2): 123-130.

Malhotra, A.K., Kestler, L.J., Mazzanti, C., Bates, J.A., Goldberg, T., and Goldman, D. 2002. A functional polymorphism in the COMT gene and performance on a test of prefrontal cognition. *Am J Psychiatry* **159**(4): 652-654.

Maness, P.F. and Schachner, M. 2007. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat Neurosci* **10**(1): 19-26.

Marazziti, D. and Catena Dell'osso, M. 2008. The role of oxytocin in neuropsychiatric disorders. *Curr Med Chem* **15**(7): 698-704.

Marchionni, M.A., Kirk, C.J., Isaacs, I.J., Hoban, C.J., Mahanthappa, N.K., Anton, E.S., Chen, C., Wason, F., Lawson, D., Hamers, F.P., Canoll, P.D., Reynolds, R., Cannella, B., Meun, D., Holt, W.F., Matthew, W.D., Chen, L.E., Gispen, W.H., Raine, C.S., Salzer, J.L., and Gwynne, D.I. 1996. Neuregulins as potential drugs for neurological disorders. *Cold Spring Harb Symp Quant Biol* **61**: 459-472.

Marechal, L., Raux, G., Dumanchin, C., Lefebvre, G., Deslandre, E., Girard, C., Campion, D., Parain, D., Frebourg, T., and Hannequin, D. 2003. Severe myoclonus-dystonia syndrome associated with a

novel epsilon-sarcoglycan gene truncating mutation. *Am J Med Genet B Neuropsychiatr Genet* **119B**(1): 114-117.

Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C.E., Vos, Y.J., Ficicioglu, C., Kirkpatrick, S., Nicolson, R., Sloman, L., Summers, A., Gibbons, C.A., Teebi, A., Chitayat, D., Weksberg, R., Thompson, A., Vardy, C., Crosbie, V., Luscombe, S., Baatjes, R., Zwaigenbaum, L., Roberts, W., Fernandez, B., Szatmari, P., and Scherer, S.W. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**(2): 477-488.

Marui, T., Funatogawa, I., Koishi, S., Yamamoto, K., Matsumoto, H., Hashimoto, O., Nanba, E., Nishida, H., Sugiyama, T., Kasai, K., Watanabe, K., Kano, Y., and Kato, N. 2008. Association of the neuronal cell adhesion molecule (NRCAM) gene variants with autism. *Int J Neuropsychopharmacol*: 1-10.

McCarroll, S.A. and Altshuler, D.M. 2007. Copy-number variation and association studies of human disease. *Nat Genet* **39**(7 Suppl): S37-42.

McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J., and Altshuler, D.M. 2006. Common deletion polymorphisms in the human genome. *Nat Genet* **38**(1): 86-92.

McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., and Altshuler, D. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* **40**(10): 1166-1174.

McClellan, J.M., Susser, E., and King, M.C. 2007. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* **190**: 194-199.

McGuffin, P., Owen, M.J., and Farmer, A.E. 1995. Genetic basis of schizophrenia. *Lancet* **346**(8976): 678-682.

Mei, L. and Xiong, W.C. 2008. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* **9**(6): 437-452.

Mejat, A., Ramond, F., Bassel-Duby, R., Khochbin, S., Olson, E.N., and Schaeffer, L. 2005. Histone deacetylase 9 couples neuronal activity to muscle chromatin acetylation and gene expression. *Nat Neurosci* **8**(3): 313-321.

Menniti, F.S., Chappie, T.A., Humphrey, J.M., and Schmidt, C.J. 2007. Phosphodiesterase 10A inhibitors: a novel approach to the treatment of the symptoms of schizophrenia. *Curr Opin Investig Drugs* **8**(1): 54-59.

Meyer-Lindenberg, A., Kohn, P.D., Kolachana, B., Kippenhan, S., McInerney-Leo, A., Nussbaum, R., Weinberger, D.R., and Berman, K.F. 2005. Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nat Neurosci* **8**(5): 594-596.

Mileyko, Y., Joh, R.I., and Weitz, J.S. 2008. Small-scale copy number variation and large-scale changes in gene expression. *Proc Natl Acad Sci U S A*.

Millar, J.K., Pickard, B.S., Mackie, S., James, R., Christie, S., Buchanan, S.R., Malloy, M.P., Chubb, J.E., Huston, E., Baillie, G.S., Thomson, P.A., Hill, E.V., Brandon, N.J., Rain, J.C., Camargo, L.M., Whiting, P.J., Houslay, M.D., Blackwood, D.H., Muir, W.J., and Porteous, D.J. 2005. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**(5751): 1187-1191.

Millar, J.K., Thomson, P.A., Wray, N.R., Muir, W.J., Blackwood, D.H., and Porteous, D.J. 2003. Response to Amar J. Klar: The chromosome 1;11 translocation provides the best evidence supporting genetic etiology for schizophrenia and bipolar affective disorders. *Genetics* **163**(2): 833-835; author reply 837-838.

Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., Devon, R.S., Clair, D.M., Muir, W.J., Blackwood, D.H., and Porteous, D.J. 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet* **9**(9): 1415-1423.

Miller, D.W., Hague, S.M., Clarimon, J., Baptista, M., Gwinn-Hardy, K., Cookson, M.R., and Singleton, A.B. 2004. Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication. *Neurology* **62**(10): 1835-1838.

Mills, R.E., Luttig, C.T., Larkins, C.E., Beauchamp, A., Tsui, C., Pittard, W.S., and Devine, S.E. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* **16**(9): 1182-1190.

Misbahuddin, A., Placzek, M., Lennox, G., Taanman, J.W., and Warner, T.T. 2007. Myoclonus-dystonia syndrome with severe depression is caused by an exon-skipping mutation in the epsilon-sarcoglycan gene. *Mov Disord* **22**(8): 1173-1175.

Mohn, A.R., Gainetdinov, R.R., Caron, M.G., and Koller, B.H. 1999. Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. *Cell* **98**(4): 427-436.

Montag-Sallaz, M., Schachner, M., and Montag, D. 2002. Misguided axonal projections, neural cell adhesion molecule 180 mRNA upregulation, and altered behavior in mice deficient for the close homolog of L1. *Mol Cell Biol* **22**(22): 7967-7981.

Moon, H.J., Yim, S.V., Lee, W.K., Jeon, Y.W., Kim, Y.H., Ko, Y.J., Lee, K.S., Lee, K.H., Han, S.I., and Rha, H.K. 2006. Identification of DNA copy-number aberrations by array-comparative genomic hybridization in patients with schizophrenia. *Biochem Biophys Res Commun* **344**(2): 531-539.

Morellini, F., Lepsveridze, E., Kahler, B., Dityatev, A., and Schachner, M. 2007. Reduced reactivity to novelty, impaired social behavior, and enhanced basal synaptic excitatory activity in perforant path projections to the dentate gyrus in young adult mice deficient in the neural cell adhesion molecule CHL1. *Mol Cell Neurosci* **34**(2): 121-136.

Morris, D.W., Rodgers, A., McGhee, K.A., Schwaiger, S., Scully, P., Quinn, J., Meagher, D., Waddington, J.L., Gill, M., and Corvin, A.P. 2004. Confirming RGS4 as a susceptibility gene for schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **125**(1): 50-53.

Murphy, K.C. 2002. Schizophrenia and velo-cardio-facial syndrome. *Lancet* **359**(9304): 426-430.

Murphy, K.C., Jones, L.A., and Owen, M.J. 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* **56**(10): 940-945.

Murphy, K.C. and Owen, M.J. 2001. Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. *Br J Psychiatry* **179**: 397-402.

Murray, R.M. and Lewis, S.W. 1987. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed)* **295**(6600): 681-682.

-. 1988. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed)* **296**(6614): 63.

Murthy, S.K., Nygren, A.O., El Shakankiry, H.M., Schouten, J.P., Al Khayat, A.I., Ridha, A., and Al Ali, M.T. 2007. Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment. *Cytogenet Genome Res* **116**(1-2): 135-140.

Myers, A.J., Pittman, A.M., Zhao, A.S., Rohrer, K., Kaleem, M., Marlowe, L., Lees, A., Leung, D., McKeith, I.G., Perry, R.H., Morris, C.M., Trojanowski, J.Q., Clark, C., Karlawish, J., Arnold, S., Forman, M.S., Van Deerlin, V., de Silva, R., and Hardy, J. 2007. The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. *Neurobiol Dis* **25**(3): 561-570.

Nakanishi, S., Nakajima, Y., Masu, M., Ueda, Y., Nakahara, K., Watanabe, D., Yamaguchi, S., Kawabata, S., and Okada, M. 1998. Glutamate receptors: brain function and signal transduction. *Brain Res Brain Res Rev* **26**(2-3): 230-235.

Napoli, I., Mercaldo, V., Boyl, P.P., Eleuteri, B., Zalfa, F., De Rubeis, S., Di Marino, D., Mohr, E., Massimi, M., Falconi, M., Witke, W., Costa-Mattioli, M., Sonenberg, N., Achsel, T., and Bagni, C. 2008. The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* **134**(6): 1042-1054.

Neitz, M. and Neitz, J. 1995. Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* **267**(5200): 1013-1016.

Nguyen, D.Q., Webber, C., Hehir-Kwa, J., Pfundt, R., Veltman, J., and Ponting, C.P. 2008. Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Res*.

Nguyen, D.Q., Webber, C., and Ponting, C.P. 2006. Bias of selection on human copy-number variants. *PLoS Genet* **2**(2): e20.

Ni, X., Valente, J., Azevedo, M.H., Pato, M.T., Pato, C.N., and Kennedy, J.L. 2007. Connexin 50 gene on human chromosome 1q21 is associated with schizophrenia in matched case control and family-based studies. *J Med Genet* **44**(8): 532-536.

Nikonenko, A.G., Sun, M., Lepsveridze, E., Apostolova, I., Petrova, I., Irintchev, A., Dityatev, A., and Schachner, M. 2006. Enhanced perisomatic inhibition and impaired long-term potentiation in the CA1 region of juvenile CHL1-deficient mice. *Eur J Neurosci* **23**(7): 1839-1852.

Nishimune, H., Bernreuther, C., Carroll, P., Chen, S., Schachner, M., and Henderson, C.E. 2005. Neural adhesion molecules L1 and CHL1 are survival factors for motoneurons. *J Neurosci Res* **80**(5): 593-599.

Nishimura, Y., Martin, C.L., Vazquez-Lopez, A., Spence, S.J., Alvarez-Retuerto, A.I., Sigman, M., Steindler, C., Pellegrini, S., Schanen, N.C., Warren, S.T., and Geschwind, D.H. 2007. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet* **16**(14): 1682-1698.

North, K.N., Yang, N., Wattanasirichaigoon, D., Mills, M., Easteal, S., and Beggs, A.H. 1999. A common nonsense mutation results in alpha-actinin-3 deficiency in the general population. *Nat Genet* **21**(4): 353-354.

Norton, N., Williams, H.J., and Owen, M.J. 2006. An update on the genetics of schizophrenia. *Curr Opin Psychiatry* **19**(2): 158-164.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F., and Green, R.D. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* **12**(11): 1749-1755.

O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., Dwyer, S., Holmans, P., Marchini, J.L., Spencer, C.C., Howie, B., Leung, H.T., Hartmann, A.M., Moller, H.J., Morris, D.W., Shi, Y., Feng, G., Hoffmann, P., Propping, P., Vasilescu, C., Maier, W., Rietschel, M., Zammit, S., Schumacher, J., Quinn, E.M., Schulze, T.G., Williams, N.M., Giegling, I., Iwata, N., Ikeda, M., Darvasi, A., Shifman, S., He, L., Duan, J., Sanders, A.R., Levinson, D.F., Gejman, P.V., Buccola, N.G., Mowry, B.J., Freedman, R., Amin, F., Black, D.W., Silverman, J.M., Byerley, W.F., Cloninger, C.R., Cichon, S., Nothen, M.M., Gill, M., Corvin, A., Rujescu, D., Kirov, G., and Owen, M.J. 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*.

Oliveira, S.A., Scott, W.K., Zhang, F., Stajich, J.M., Fujiwara, K., Hauser, M., Scott, B.L., Pericak-Vance, M.A., Vance, J.M., and Martin, E.R. 2004. Linkage disequilibrium and haplotype tagging polymorphisms in the Tau H1 haplotype. *Neurogenetics* **5**(3): 147-155.

Owen, M.J., Craddock, N., and O'Donovan, M.C. 2005a. Schizophrenia: genes at last? *Trends Genet* **21**(9): 518-525.

Owen, M.J., O'Donovan, M.C., and Harrison, P.J. 2005b. Schizophrenia: a genetic disorder of the synapse? *BMJ* **330**(7484): 158-159.

Ozeki, Y., Tomoda, T., Kleiderlein, J., Kamiya, A., Bord, L., Fujii, K., Okawa, M., Yamada, N., Hatten, M.E., Snyder, S.H., Ross, C.A., and Sawa, A. 2003. Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proc Natl Acad Sci U S A* **100**(1): 289-294.

Pastor, P., Ezquerra, M., Perez, J.C., Chakraverty, S., Norton, J., Racette, B.A., McKeel, D., Perlmutter, J.S., Tolosa, E., and Goate, A.M. 2004. Novel haplotypes in 17q21 are associated with progressive supranuclear palsy. *Ann Neurol* **56**(2): 249-258.

Patel, P.I., Roa, B.B., Welcher, A.A., Schoener-Scott, R., Trask, B.J., Pentao, L., Snipes, G.J., Garcia, C.A., Francke, U., Shooter, E.M., Lupski, J.R., and Suter, U. 1992. The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. *Nat Genet* **1**(3): 159-165.

Paylor, R. and Lindsay, E. 2006. Mouse models of 22q11 deletion syndrome. *Biol Psychiatry* **59**(12): 1172-1179.

Paylor, R., McIlwain, K.L., McAninch, R., Nellis, A., Yuva-Paylor, L.A., Baldini, A., and Lindsay, E.A. 2001. Mice deleted for the DiGeorge/velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments. *Hum Mol Genet* **10**(23): 2645-2650.

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C.A., Belmont, J., Cheung, S.W., Shen, R.M., Barker, D.L., and Gunderson, K.L. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* **16**(9): 1136-1148.

Perry, G.H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C.W., Scheffer, A., Steinfeld, I., Tsang, P., Yamada, N.A., Park, H.S., Kim, J.I., Seo, J.S., Yakhini, Z., Laderman, S., Bruhn, L., and Lee, C. 2008a. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* **82**(3): 685-695.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., Carter, N.P., Lee, C., and Stone, A.C. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39**(10): 1256-1260.

Perry, G.H., Tchinda, J., McGrath, S.D., Zhang, J., Picker, S.R., Caceres, A.M., Iafrate, A.J., Tyler-Smith, C., Scherer, S.W., Eichler, E.E., Stone, A.C., and Lee, C. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci U S A* **103**(21): 8006-8011.

Perry, G.H., Yang, F., Marques-Bonet, T., Murphy, C., Fitzgerald, T., Lee, A.S., Hyland, C., Stone, A.C., Hurles, M.E., Tyler-Smith, C., Eichler, E.E., Carter, N.P., Lee, C., and Redon, R. 2008b. Copy number variation and evolution in humans and chimpanzees. *Genome Res*.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**(9): e45.

Pinkel, D. and Albertson, D.G. 2005. Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* **6**: 331-354.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., and Albertson, D.G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**(2): 207-211.

Pinto, D., Marshall, C., Feuk, L., and Scherer, S.W. 2007. Copy-number variation in control population cohorts. *Hum Mol Genet* **16 Spec No. 2**: R168-173.

Piotrowski, A., Bruder, C.E., Andersson, R., de Stahl, T.D., Menzel, U., Sandgren, J., Poplawski, A., von Tell, D., Crasto, C., Bogdan, A., Bartoszewski, R., Bebok, Z., Krzyzanowski, M., Jankowski, Z., Partridge, E.C., Komorowski, J., and Dumanski, J.P. 2008. Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* **29**(9): 1118-1124.

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L., Thomas, N., Morris, C.M., Dickson, D., Wood, N.W., Hardy, J., Lees, A.J., and de Silva, R. 2005. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet* **42**(11): 837-846.

Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., and de Silva, R. 2004. The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum Mol Genet* **13**(12): 1267-1274.

Pletnikov, M.V., Ayhan, Y., Nikolskaia, O., Xu, Y., Ovanesov, M.V., Huang, H., Mori, S., Moran, T.H., and Ross, C.A. 2008. Inducible expression of mutant human DISC1 in mice is associated with brain and behavioral abnormalities reminiscent of schizophrenia. *Mol Psychiatry* **13**(2): 173-186, 115.

Pocklington, A.J., Cumiskey, M., Armstrong, J.D., and Grant, S.G. 2006. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol Syst Biol* **2**: 2006 0023.

Poliak, S., Gollan, L., Martinez, R., Custer, A., Einheber, S., Salzer, J.L., Trimmer, J.S., Shrager, P., and Peles, E. 1999. Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels. *Neuron* **24**(4): 1037-1047.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**(1): 41-46.

Porta, S., Serra, S.A., Huch, M., Valverde, M.A., Llorens, F., Estivill, X., Arbones, M.L., and Marti, E. 2007. RCAN1 (DSCR1) increases neuronal susceptibility to oxidative stress: a potential pathogenic process in neurodegeneration. *Hum Mol Genet* **16**(9): 1039-1050.

Prades, C., Arnould, I., Annilo, T., Shulenin, S., Chen, Z.Q., Orosco, L., Triunfol, M., Devaud, C., Maintoux-Larois, C., Lafargue, C., Lemoine, C., Denefle, P., Rosier, M., and Dean, M. 2002. The human ATP binding cassette gene ABCA13, located on chromosome 7p12.3, encodes a 5058 amino acid protein with an extracellular domain encoded in part by a 4.8-kb conserved exon. *Cytogenet Genome Res* **98**(2-3): 160-168.

Pritchard, J.K. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69**(1): 124-137.

Pritchard, J.K. and Cox, N.J. 2002. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* **11**(20): 2417-2423.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**(3): 559-575.

Qiao, Y., Liu, X., Harvard, C., Nolin, S.L., Brown, W.T., Koochek, M., Holden, J.J., Lewis, M.E., and Rajcan-Separovic, E. 2007. Large-scale copy number variants (CNVs): distribution in normal subjects and FISH/real-time qPCR analysis. *BMC Genomics* **8**: 167.

Rademakers, R., Cruts, M., and van Broeckhoven, C. 2004. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum Mutat* **24**(4): 277-295.

Rademakers, R., Melquist, S., Cruts, M., Theuns, J., Del-Favero, J., Poorkaj, P., Baker, M., Sleegers, K., Crook, R., De Pooter, T., Bel Kacem, S., Adamson, J., Van den Bossche, D., Van den Broeck, M., Gass, J., Corsmit, E., De Rijk, P., Thomas, N., Engelborghs, S., Heckman, M., Litvan, I., Crook, J., De Deyn, P.P., Dickson, D., Schellenberg, G.D., Van Broeckhoven, C., and Hutton, M.L. 2005. High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. *Hum Mol Genet* **14**(21): 3281-3292.

Ramocki, M.B. and Zoghbi, H.Y. 2008. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* **455**(7215): 912-918.

Rao, Y., Hoffmann, E., Zia, M., Bodin, L., Zeman, M., Sellers, E.M., and Tyndale, R.F. 2000. Duplications and defects in the CYP2A6 gene: identification, genotyping, and in vivo effects on smoking. *Mol Pharmacol* **58**(4): 747-755.

Raux, G., Bonnet-Brilhault, F., Louchart, S., Houy, E., Gantier, R., Levillain, D., Allio, G., Haouzir, S., Petit, M., Martinez, M., Frebourg, T., Thibaut, F., and Campion, D. 2002. The -2 bp deletion in exon 6 of the 'alpha 7-like' nicotinic receptor subunit gene is a risk factor for the P50 sensory gating deficit. *Mol Psychiatry* **7**(9): 1006-1011.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. 2006. Global variation in copy number in the human genome. *Nature* **444**(7118): 444-454.

Reich, D.E., Gabriel, S.B., and Altshuler, D. 2003. Quality and completeness of SNP databases. *Nat Genet* **33**(4): 457-458.

Reissner, C., Klose, M., Fairless, R., and Missler, M. 2008. Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. *Proc Natl Acad Sci U S A* **105**(39): 15124-15129.

Ren, K., Thinschmidt, J., Liu, J., Ai, L., Papke, R.L., King, M.A., Hughes, J.A., and Meyer, E.M. 2007. alpha7 Nicotinic receptor gene delivery into mouse hippocampal neurons leads to functional receptor

expression, improved spatial memory-related performance, and tau hyperphosphorylation. *Neuroscience* **145**(1): 314-322.

Riley, B., Williamson, M., Collier, D., Wilkie, H., and Makoff, A. 2002. A 3-Mb map of a large Segmental duplication overlapping the alpha7-nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14. *Genomics* **79**(2): 197-209.

Riley, B.P., Makoff, A., Mogudi-Carter, M., Jenkins, T., Williamson, R., Collier, D., and Murray, R. 2000. Haplotype transmission disequilibrium and evidence for linkage of the CHRNA7 gene region to schizophrenia in Southern African Bantu families. *Am J Med Genet* **96**(2): 196-201.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* **273**(5281): 1516-1517.

Rogers, S.J., Wehner, D.E., and Hagerman, R. 2001. The behavioral phenotype in fragile X: symptoms of autism in very young children with fragile X syndrome, idiopathic autism, and other developmental disorders. *J Dev Behav Pediatr* **22**(6): 409-417.

Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., Dubas, F., Frebourg, T., and Campion, D. 2006. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* **38**(1): 24-26.

Roy, A.M., Carroll, M.L., Nguyen, S.V., Salem, A.H., Oldridge, M., Wilkie, A.O., Batzer, M.A., and Deininger, P.L. 2000. Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res* **10**(10): 1485-1495.

Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-386.

Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., Murray, R., Ruggeri, M., Tosato, S., Bonetto, C., Steinberg, S., Sigurdsson, E., Sigmundsson, T., Petursson, H., Gylfason, A., Olason, P.I., Hardarsson, G., Jonsdottir, G.A., Gustafsson, O., Fossdal, R., Giegling, I., Moller, H.J., Hartmann, A., Hoffmann, P., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Andreassen, O.A., Djurovic, S., Hansen, T., Werge, T., Melle, I., Kiemeney, L.A., Franke, B., Buizer-Voskamp, J.E., Ophoff, R.A., Rietschel, M., Nothen, M.M., Stefansson, K., Peltonen, L., St Clair, D., Stefansson, H., and Collier, D.A. 2008. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*.

Saito, T., Aghalar, M.R., and Lachman, H.M. 2005. Analysis of PIK3C3 promoter variant in African-Americans with schizophrenia. *Schizophr Res* **76**(2-3): 361-362.

Sakel, M. 1954. The classical Sakel shock treatment: a reappraisal. *J Clin Exp Psychopathol* **15**(3): 255-316.

Sakurai, K., Migita, O., Toru, M., and Arinami, T. 2002. An association between a missense polymorphism in the close homologue of L1 (CHL1, CALL) gene and schizophrenia. *Mol Psychiatry* **7**(4): 412-415.

Sakurai, T., Ramoz, N., Reichert, J.G., Corwin, T.E., Kryzak, L., Smith, C.J., Silverman, J.M., Hollander, E., and Buxbaum, J.D. 2006. Association analysis of the NrCAM gene in autism and in subsets of families with severe obsessive-compulsive or self-stimulatory behaviors. *Psychiatr Genet* **16**(6): 251-257.

Sanders, A.R., Duan, J., Levinson, D.F., Shi, J., He, D., Hou, C., Burrell, G.J., Rice, J.P., Nertney, D.A., Olincy, A., Rozic, P., Vinogradov, S., Buccola, N.G., Mowry, B.J., Freedman, R., Amin, F., Black, D.W., Silverman, J.M., Byerley, W.F., Crowe, R.R., Cloninger, C.R., Martinez, M., and Gejman, P.V. 2008. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry* **165**(4): 497-506.

Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science* **214**(4526): 1205-1210.

Sass, H. 1989. The historical evolution of the concept of negative symptoms in schizophrenia. *Br J Psychiatry Suppl*(7): 26-31; discussion 37-40.

Scannevin, R.H. and Huganir, R.L. 2000. Postsynaptic organization and regulation of excitatory synapses. *Nat Rev Neurosci* **1**(2): 133-141.

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E., and Feuk, L. 2007. Challenges and standards in integrating surveys of structural variation. *Nat Genet* **39**(7 Suppl): S7-15.

Schmidt, C.J., Chapin, D.S., Cianfrogna, J., Corman, M.L., Hajos, M., Harms, J.F., Hoffman, W.E., Lebel, L.A., McCarthy, S.A., Nelson, F.R., Proulx-LaFrance, C., Majchrzak, M.J., Ramirez, A.D., Schmidt, K., Seymour, P.A., Siuciak, J.A., Tingley, F.D., 3rd, Williams, R.D., Verhoest, P.R., and Menniti, F.S. 2008. Preclinical characterization of selective phosphodiesterase 10A inhibitors: a new therapeutic approach to the treatment of schizophrenia. *J Pharmacol Exp Ther* **325**(2): 681-690.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* **30**(12): e57.

Schwartz, D.C., Saffran, W., Welsh, J., Haas, R., Goldenberg, M., and Cantor, C.R. 1983. New techniques for purifying large DNAs and studying their properties and packaging. *Cold Spring Harb Symp Quant Biol* **47 Pt 1**: 189-195.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimaki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., and Wigler, M. 2007. Strong association of de novo copy number mutations with autism. *Science* **316**(5823): 445-449.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., and Wigler, M. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305**(5683): 525-528.

Seeman, P., Lee, T., Chau-Wong, M., and Wong, K. 1976. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature* **261**(5562): 717-719.

Service, R.F. 2006. Gene sequencing. The race for the $1000 genome. *Science* **311**(5767): 1544-1546.

Shaikh, T.H., Kurahashi, H., and Emanuel, B.S. 2001. Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. *Genet Med* **3**(1): 6-13.

Sharp, A.J., Hansen, S., Selzer, R.R., Cheng, Z., Regan, R., Hurst, J.A., Stewart, H., Price, S.M., Blair, E., Hennekam, R.C., Fitzpatrick, C.A., Segraves, R., Richmond, T.A., Guiver, C., Albertson, D.G., Pinkel, D., Eis, P.S., Schwartz, S., Knight, S.J., and Eichler, E.E. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* **38**(9): 1038-1042.

Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., Oseroff, V.V., Albertson, D.G., Pinkel, D., and Eichler, E.E. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**(1): 78-88.

Sharp, A.J., Mefford, H.C., Li, K., Baker, C., Skinner, C., Stevenson, R.E., Schroer, R.J., Novara, F., De Gregori, M., Ciccone, R., Broomer, A., Casuga, I., Wang, Y., Xiao, C., Barbacioru, C., Gimelli, G., Bernardina, B.D., Torniero, C., Giorda, R., Regan, R., Murday, V., Mansour, S., Fichera, M., Castiglia, L., Failla, P., Ventura, M., Jiang, Z., Cooper, G.M., Knight, S.J., Romano, C., Zuffardi, O., Chen, C., Schwartz, C.E., and Eichler, E.E. 2008. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* **40**(3): 322-328.

Shaw-Smith, C., Pittman, A.M., Willatt, L., Martin, H., Rickman, L., Gribble, S., Curley, R., Cumming, S., Dunn, C., Kalaitzopoulos, D., Porter, K., Prigmore, E., Krepischi-Santos, A.C., Varela, M.C., Koiffmann, C.P., Lees, A.J., Rosenberg, C., Firth, H.V., de Silva, R., and Carter, N.P. 2006. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* **38**(9): 1032-1037.

Shaw, C.J. and Lupski, J.R. 2005. Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum Genet* **116**(1-2): 1-7.

She, X., Cheng, Z., Zollner, S., Church, D.M., and Eichler, E.E. 2008. Mouse segmental duplication and copy number variation. *Nat Genet* **40**(7): 909-914.

Shelling, A.N. and Ferguson, L.R. 2007. Genetic variation in human disease and a new role for copy number variants. *Mutat Res* **622**(1-2): 33-41.

Shen, S., Lang, B., Nakamoto, C., Zhang, F., Pu, J., Kuan, S.L., Chatzi, C., He, S., Mackie, I., Brandon, N.J., Marquis, K.L., Day, M., Hurko, O., McCaig, C.D., Riedel, G., and St Clair, D. 2008. Schizophrenia-related neural and behavioral phenotypes in transgenic mice expressing truncated Disc1. *J Neurosci* **28**(43): 10893-10904.

Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26**(10): 1135-1145.

Simon-Sanchez, J., Scholz, S., Fung, H.C., Matarin, M., Hernandez, D., Gibbs, J.R., Britton, A., de Vrieze, F.W., Peckham, E., Gwinn-Hardy, K., Crawley, A., Keen, J.C., Nash, J., Borgaonkar, D., Hardy, J., and Singleton, A. 2007. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet* **16**(1): 1-14.

Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., Lincoln, S., Crawley, A., Hanson, M., Maraganore, D., Adler, C., Cookson, M.R., Muenter, M., Baptista, M., Miller, D., Blancato, J., Hardy, J., and Gwinn-Hardy, K. 2003. alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**(5646): 841.

Sklar, P. 2002. Linkage analysis in psychiatric disorders: the emerging picture. *Annu Rev Genomics Hum Genet* **3**: 371-413.

Sleegers, K., Brouwers, N., Gijselinck, I., Theuns, J., Goossens, D., Wauters, J., Del-Favero, J., Cruts, M., van Duijn, C.M., and Van Broeckhoven, C. 2006. APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* **129**(Pt 11): 2977-2983.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., and Albertson, D.G. 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* **29**(3): 263-264.

Snijders, A.M., Nowak, N.J., Huey, B., Fridlyand, J., Law, S., Conroy, J., Tokuyasu, T., Demir, K., Chiu, R., Mao, J.H., Jain, A.N., Jones, S.J., Balmain, A., Pinkel, D., and Albertson, D.G. 2005. Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res* **15**(2): 302-311.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. 1997. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**(4): 399-407.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**(3): 503-517.

St Clair, D. 2008. Large Recurrent deletions are associated with schizophrenia. In *XVIth World Congress on Psychiatric Genetics*, pp. 19, Osaka, Japan.

St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Gosden, C., and Evans, H.J. 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* **336**(8706): 13-16.

Stankiewicz, P., Inoue, K., Bi, W., Walz, K., Park, S.S., Kurotaki, N., Shaw, C.J., Fonseca, P., Yan, J., Lee, J.A., Khajavi, M., and Lupski, J.R. 2003. Genomic disorders: genome architecture results in susceptibility to DNA rearrangements causing common human traits. *Cold Spring Harb Symp Quant Biol* **68**: 445-454.

Stankiewicz, P. and Lupski, J.R. 2002a. Genome architecture, rearrangements and genomic disorders. *Trends Genet* **18**(2): 74-82.

-. 2002b. Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* **12**(3): 312-319.

Stark, K.L., Xu, B., Bagchi, A., Lai, W.S., Liu, H., Hsu, R., Wan, X., Pavlidis, P., Mills, A.A., Karayiorgou, M., and Gogos, J.A. 2008. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet* **40**(6): 751-760.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G., Desnica, N., Hicks, A., Gylfason, A., Gudbjartsson, D.F., Jonsdottir, G.M., Sainz, J., Agnarsson, K., Birgisdottir, B., Ghosh, S., Olafsdottir, A., Cazier, J.B., Kristjansson, K., Frigge, M.L., Thorgeirsson, T.E., Gulcher, J.R., Kong, A., and Stefansson, K. 2005. A common inversion under selection in Europeans. *Nat Genet* **37**(2): 129-137.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., Hansen, T., Jakobsen, K.D., Muglia, P.,

Francks, C., Matthews, P.M., Gylfason, A., Halldorsson, B.V., Gudbjartsson, D., Thorgeirsson, T.E., Sigurdsson, A., Jonasdottir, A., Bjornsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B.B., Giegling, I., Moller, H.J., Hartmann, A., Shianna, K.V., Ge, D., Need, A.C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Toulopoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Muhleisen, T.W., Wang, A.G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemeney, L.A., Franke, B., Sabatti, C., Freimer, N.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Andreassen, O.A., Ophoff, R.A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D.B., Nothen, M.M., Peltonen, L., Collier, D.A., St Clair, D., Stefansson, K., Kahn, R.S., Linszen, D.H., van Os, J., Wiersma, D., Bruggeman, R., Cahn, W., de Haan, L., Krabbendam, L., and Myin-Germeys, I. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**(7210): 232-236.

Stopkova, P., Saito, T., Papolos, D.F., Vevera, J., Paclt, I., Zukov, I., Bersson, Y.B., Margolis, B.A., Strous, R.D., and Lachman, H.M. 2004. Identification of PIK3C3 promoter variant associated with bipolar disorder and schizophrenia. *Biol Psychiatry* **55**(10): 981-988.

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S.W., Tavare, S., Deloukas, P., Hurles, M.E., and Dermitzakis, E.T. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**(5813): 848-853.

Stromgren, E. 1987. Changes in the incidence of schizophrenia? *Br J Psychiatry* **150**: 1-7.

Sudhof, T.C. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455**(7215): 903-911.

Sullivan, P.F. 2005. The genetics of schizophrenia. *PLoS Med* **2**(7): e212.

Sullivan, P.F., Kendler, K.S., and Neale, M.C. 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* **60**(12): 1187-1192.

Sullivan, P.F., Lin, D., Tzeng, J.Y., van den Oord, E., Perkins, D., Stroup, T.S., Wagner, M., Lee, S., Wright, F.A., Zou, F., Liu, W., Downing, A.M., Lieberman, J., and Close, S.L. 2008. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* **13**(6): 570-584.

Sutrala, S.R., Goossens, D., Williams, N.M., Heyrman, L., Adolfsson, R., Norton, N., Buckland, P.R., and Del-Favero, J. 2007. Gene copy number variation in schizophrenia. *Schizophr Res* **96**(1-3): 93-99.

Sutrala, S.R., Norton, N., Williams, N.M., and Buckland, P.R. 2008. Gene copy number variation in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **147B**(5): 606-611.

Szatmari, P. Paterson, A.D. Zwaigenbaum, L. Roberts, W. Brian, J. Liu, X.Q. Vincent, J.B. Skaug, J.L. Thompson, A.P. Senman, L. Feuk, L. Qian, C. Bryson, S.E. Jones, M.B. Marshall, C.R. Scherer, S.W. Vieland, V.J. Bartlett, C. Mangin, L.V. Goedken, R. Segre, A. Pericak-Vance, M.A. Cuccaro, M.L. Gilbert, J.R. Wright, H.H. Abramson, R.K. Betancur, C. Bourgeron, T. Gillberg, C. Leboyer, M. Buxbaum, J.D. Davis, K.L. Hollander, E. Silverman, J.M. Hallmayer, J. Lotspeich, L. Sutcliffe, J.S. Haines, J.L. Folstein, S.E. Piven, J. Wassink, T.H. Sheffield, V. Geschwind, D.H. Bucan, M. Brown, W.T. Cantor, R.M. Constantino, J.N. Gilliam, T.C. Herbert, M. Lajonchere, C. Ledbetter, D.H. Lese-Martin, C. Miller, J. Nelson, S. Samango-Sprouse, C.A. Spence, S. State, M. Tanzi, R.E. Coon, H. Dawson, G. Devlin, B. Estes, A. Flodman, P. Klei, L. McMahon, W.M. Minshew, N. Munson, J. Korvatska, E. Rodier, P.M. Schellenberg, G.D. Smith, M. Spence, M.A. Stodgell, C. Tepper, P.G. Wijsman, E.M. Yu, C.E. Roge, B. Mantoulan, C. Wittemeyer, K. Poustka, A. Felder, B. Klauck, S.M. Schuster, C. Poustka, F. Bolte, S. Feineis-Matthews, S. Herbrecht, E. Schmotzer, G. Tsiantis, J. Papanikolaou, K. Maestrini, E. Bacchelli, E. Blasi, F. Carone, S. Toma, C. Van Engeland, H. de Jonge, M. Kemner, C. Koop, F. Langemeijer, M. Hijmans, C. Staal, W.G. Baird, G. Bolton, P.F. Rutter, M.L. Weisblatt, E. Green, J. Aldred, C. Wilkinson, J.A. Pickles, A. Le Couteur, A. Berney, T. McConachie, H. Bailey, A.J. Francis, K. Honeyman, G. Hutchinson, A. Parr, J.R. Wallace, S. Monaco, A.P. Barnby, G. Kobayashi, K. Lamb, J.A. Sousa, I. Sykes, N. Cook, E.H. Guter, S.J. Leventhal, B.L. Salt, J. Lord, C. Corsello, C. Hus, V. Weeks, D.E. Volkmar, F. Tauber, M. Fombonne, E. Shih, A. and Meyer, K.J. 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39**(3): 319-328.

Takao, T., Tachikawa, H., Kawanishi, Y., Katano, T., Sen, B., Homma, M., Kohda, Y., Mizukami, K., and Asada, T. 2006. Association of treatment-resistant schizophrenia with the G2677A/T and C3435T polymorphisms in the ATP-binding cassette subfamily B member 1 gene. *Psychiatr Genet* **16**(2): 47-48.

Tang, R., Zhao, X., Fang, C., Tang, W., Huang, K., Wang, L., Li, H., Feng, G., Zhu, S., Liu, H., He, L., and Shi, Y. 2008. Investigation of variants in the promoter region of PIK3C3 in schizophrenia. *Neurosci Lett* **437**(1): 42-44.

Taske, N.L., Williamson, M.P., Makoff, A., Bate, L., Curtis, D., Kerr, M., Kjeldsen, M.J., Pang, K.A., Sundqvist, A., Friis, M.L., Chadwick, D., Richens, A., Covanis, A., Santos, M., Arzimanoglou, A., Panayiotopoulos, C.P., Whitehouse, W.P., Rees, M., and Gardiner, R.M. 2002. Evaluation of the positional candidate gene CHRNA7 at the juvenile myoclonic epilepsy locus (EJM2) on chromosome 15q13-14. *Epilepsy Res* **49**(2): 157-172.

Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A., and Tunnacliffe, A. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**(3): 718-725.

Tills, D. 1977. The use of the FST statistic of Wright for estimating the effects of genetic drift, selection and migration populations, with special reference to Ireland. *Hum Hered* **27**(2): 153-159.

Todd, J.A. 2006. Statistical false positive or true disease pathway? *Nat Genet* **38**(7): 731-733.

Todd, P.K., Mack, K.J., and Malter, J.S. 2003. The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95. *Proc Natl Acad Sci U S A* **100**(24): 14374-14378.

Townson, J.R., Barcellos, L.F., and Nibbs, R.J. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* **32**(10): 3016-3026.

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M.V., and Eichler, E.E. 2005. Fine-scale structural variation of the human genome. *Nat Genet* **37**(7): 727-732.

Uehara, Y., Yamada, T., Baba, Y., Miura, S., Abe, S., Kitajima, K., Higuchi, M.A., Iwamoto, T., and Saku, K. 2008. ATP-binding cassette transporter G4 is highly expressed in microglia in Alzheimer's brain. *Brain Res* **1217**: 239-246.

Uhr, M., Tontsch, A., Namendorf, C., Ripke, S., Lucae, S., Ising, M., Dose, T., Ebinger, M., Rosenhagen, M., Kohli, M., Kloiber, S., Salyakina, D., Bettecken, T., Specht, M., Putz, B., Binder, E.B., Muller-Myhsok, B., and Holsboer, F. 2008. Polymorphisms in the drug transporter gene ABCB1 predict antidepressant treatment response in depression. *Neuron* **57**(2): 203-209.

Urban, A.E., Korbel, J.O., Selzer, R., Richmond, T., Hacker, A., Popescu, G.V., Cubells, J.F., Green, R., Emanuel, B.S., Gerstein, M.B., Weissman, S.M., and Snyder, M. 2006. High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays. *Proc Natl Acad Sci U S A* **103**(12): 4534-4539.

Valente, E.M., Edwards, M.J., Mir, P., DiGiorgio, A., Salvi, S., Davis, M., Russo, N., Bozi, M., Kim, H.T., Pennisi, G., Quinn, N., Dallapiccola, B., and Bhatia, K.P. 2005. The epsilon-sarcoglycan gene in myoclonic syndromes. *Neurology* **64**(4): 737-739.

van Rossum, J.M. 1966. The significance of dopamine-receptor blockade for the mechanism of action of neuroleptic drugs. *Arch Int Pharmacodyn Ther* **160**(2): 492-494.

Veltman, M.W., Craig, E.E., and Bolton, P.F. 2005. Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review. *Psychiatr Genet* **15**(4): 243-254.

Venter, J.C. Adams, M.D. Myers, E.W. Li, P.W. Mural, R.J. Sutton, G.G. Smith, H.O. Yandell, M. Evans, C.A. Holt, R.A. Gocayne, J.D. Amanatides, P. Ballew, R.M. Huson, D.H. Wortman, J.R. Zhang, Q. Kodira, C.D. Zheng, X.H. Chen, L. Skupski, M. Subramanian, G. Thomas, P.D. Zhang, J. Gabor Miklos, G.L. Nelson, C. Broder, S. Clark, A.G. Nadeau, J. McKusick, V.A. Zinder, N. Levine, A.J. Roberts, R.J. Simon, M. Slayman, C. Hunkapiller, M. Bolanos, R. Delcher, A. Dew, I. Fasulo, D. Flanigan, M. Florea, L. Halpern, A. Hannenhalli, S. Kravitz, S. Levy, S. Mobarry, C. Reinert, K. Remington, K. Abu-Threideh, J. Beasley, E. Biddick, K. Bonazzi, V. Brandon, R. Cargill, M. Chandramouliswaran, I. Charlab, R. Chaturvedi, K. Deng, Z. Di Francesco, V. Dunn, P. Eilbeck, K. Evangelista, C. Gabrielian, A.E. Gan, W. Ge, W. Gong, F. Gu, Z. Guan, P. Heiman, T.J. Higgins,

M.E. Ji, R.R. Ke, Z. Ketchum, K.A. Lai, Z. Lei, Y. Li, Z. Li, J. Liang, Y. Lin, X. Lu, F. Merkulov, G.V. Milshina, N. Moore, H.M. Naik, A.K. Narayan, V.A. Neelam, B. Nusskern, D. Rusch, D.B. Salzberg, S. Shao, W. Shue, B. Sun, J. Wang, Z. Wang, A. Wang, X. Wang, J. Wei, M. Wides, R. Xiao, C. Yan, C. Yao, A. Ye, J. Zhan, M. Zhang, W. Zhang, H. Zhao, Q. Zheng, L. Zhong, F. Zhong, W. Zhu, S. Zhao, S. Gilbert, D. Baumhueter, S. Spier, G. Carter, C. Cravchik, A. Woodage, T. Ali, F. An, H. Awe, A. Baldwin, D. Baden, H. Barnstead, M. Barrow, I. Beeson, K. Busam, D. Carver, A. Center, A. Cheng, M.L. Curry, L. Danaher, S. Davenport, L. Desilets, R. Dietz, S. Dodson, K. Doup, L. Ferriera, S. Garg, N. Gluecksmann, A. Hart, B. Haynes, J. Haynes, C. Heiner, C. Hladun, S. Hostin, D. Houck, J. Howland, T. Ibegwam, C. Johnson, J. Kalush, F. Kline, L. Koduru, S. Love, A. Mann, F. May, D. McCawley, S. McIntosh, T. McMullen, I. Moy, M. Moy, L. Murphy, B. Nelson, K. Pfannkoch, C. Pratts, E. Puri, V. Qureshi, H. Reardon, M. Rodriguez, R. Rogers, Y.H. Romblad, D. Ruhfel, B. Scott, R. Sitter, C. Smallwood, M. Stewart, E. Strong, R. Suh, E. Thomas, R. Tint, N.N. Tse, S. Vech, C. Wang, G. Wetter, J. Williams, S. Williams, M. Windsor, S. Winn-Deen, E. Wolfe, K. Zaveri, J. Zaveri, K. Abril, J.F. Guigo, R. Campbell, M.J. Sjolander, K.V. Karlak, B. Kejariwal, A. Mi, H. Lazareva, B. Hatton, T. Narechania, A. Diemer, K. Muruganujan, A. Guo, N. Sato, S. Bafna, V. Istrail, S. Lippert, R. Schwartz, R. Walenz, B. Yooseph, S. Allen, D. Basu, A. Baxendale, J. Blick, L. Caminha, M. Carnes-Stine, J. Caulk, P. Chiang, Y.H. Coyne, M. Dahlke, C. Mays, A. Dombroski, M. Donnelly, M. Ely, D. Esparham, S. Fosler, C. Gire, H. Glanowski, S. Glasser, K. Glodek, A. Gorokhov, M. Graham, K. Gropman, B. Harris, M. Heil, J. Henderson, S. Hoover, J. Jennings, D. Jordan, C. Jordan, J. Kasha, J. Kagan, L. Kraft, C. Levitsky, A. Lewis, M. Liu, X. Lopez, J. Ma, D. Majoros, W. McDaniel, J. Murphy, S. Newman, M. Nguyen, T. Nguyen, N. Nodell, M. Pan, S. Peck, J. Peterson, M. Rowe, W. Sanders, R. Scott, J. Simpson, M. Smith, T. Sprague, A. Stockwell, T. Turner, R. Venter, E. Wang, M. Wen, M. Wu, D. Wu, M. Xia, A. Zandieh, A. and Zhu, X. 2001. The sequence of the human genome. *Science* **291**(5507): 1304-1351.

Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A., and Veltman, J.A. 2008. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* **83**(4): 504-510.

Wahrle, S.E., Jiang, H., Parsadanian, M., Hartman, R.E., Bales, K.R., Paul, S.M., and Holtzman, D.M. 2005. Deletion of Abca1 increases Abeta deposition in the PDAPP transgenic mouse model of Alzheimer disease. *J Biol Chem* **280**(52): 43236-43242.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S.M., Rippey, C.F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R.L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E.E., Meltzer, P.S., Nelson, S.F., Singleton, A.B., Lee, M.K., Rapoport, J.L., King, M.C., and Sebat, J. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**(5875): 539-543.

Wang, D., Li, Z., Messing, E.M., and Wu, G. 2005. The SPRY domain-containing SOCS box protein 1 (SSB-1) interacts with MET and enhances the hepatocyte growth factor-induced Erk-Elk-1-serum response element pathway. *J Biol Chem* **280**(16): 16393-16401.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**(11): 1665-1674.

Wang, X.D., Su, Y.A., Guo, C.M., Yang, Y., and Si, T.M. 2008. Chronic antipsychotic drug administration alters the expression of neuregulin 1beta, ErbB2, ErbB3, and ErbB4 in the rat prefrontal cortex and hippocampus. *Int J Neuropsychopharmacol* **11**(4): 553-561.

Weksberg, R., Hughes, S., Moldovan, L., Bassett, A.S., Chow, E.W., and Squire, J.A. 2005. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. *BMC Genomics* **6**: 180.

White, S., Kalf, M., Liu, Q., Villerius, M., Engelsma, D., Kriek, M., Vollebregt, E., Bakker, B., van Ommen, G.J., Breuning, M.H., and den Dunnen, J.T. 2002. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* **71**(2): 365-374.

WHO. 1992. *The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic guidelines*. World Health Organisation, Geneva.

Wilke, K., Duman, B., and Horst, J. 2000. Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR. *Hum Mutat* **16**(5): 431-436.

Williams, N.M., Williams, H., Majounie, E., Norton, N., Glaser, B., Morris, H.R., Owen, M.J., and O'Donovan, M.C. 2008. Analysis of copy number variation using quantitative interspecies competitive PCR. *Nucleic Acids Res* **36**(17): e112.

Wilson, G.M., Flibotte, S., Chopra, V., Melnyk, B.L., Honer, W.G., and Holt, R.A. 2006a. DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. *Hum Mol Genet* **15**(5): 743-749.

Wilson, G.M., Flibotte, S., Missirlis, P.I., Marra, M.A., Jones, S., Thornton, K., Clark, A.G., and Holt, R.A. 2006b. Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Res* **16**(2): 173-181.

Winslow, J.T. and Insel, T.R. 2002. The social deficits of the oxytocin knockout mouse. *Neuropeptides* **36**(2-3): 221-229.

Wollstein, A., Herrmann, A., Wittig, M., Nothnagel, M., Franke, A., Nurnberg, P., Schreiber, S., Krawczak, M., and Hampe, J. 2007. Efficacy assessment of SNP sets for genome-wide disease association studies. *Nucleic Acids Res* **35**(17): e113.

Wong, K.K., deLeeuw, R.J., Dosanjh, N.S., Kimm, L.R., Cheng, Z., Horsman, D.E., MacAulay, C., Ng, R.T., Brown, C.J., Eichler, E.E., and Lam, W.L. 2007. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* **80**(1): 91-104.

Wu, E.Q., Birnbaum, H.G., Shi, L., Ball, D.E., Kessler, R.C., Moulis, M., and Aggarwal, J. 2005. The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry* **66**(9): 1122-1129.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* **40**(7): 880-885.

Xue, Y., Sun, D., Daly, A., Yang, F., Zhou, X., Zhao, M., Huang, N., Zerjal, T., Lee, C., Carter, N.P., Hurles, M.E., and Tyler-Smith, C. 2008. Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet* **83**(3): 337-346.

Yamasaki, N., Maekawa, M., Kobayashi, K., Kajii, Y., Maeda, J., Soma, M., Takao, K., Tanda, K., Ohira, K., Toyama, K., Kanzaki, K., Fukunaga, K., Sudo, Y., Ichinose, H., Ikeda, M., Iwata, N., Ozaki, N., Suzuki, H., Higuchi, M., Suhara, T., Yuasa, S., and Miyakawa, T. 2008. Alpha-CaMKII deficiency causes immature dentate gyrus, a novel candidate endophenotype of psychiatric disorders. *Mol Brain* **1**(1): 6.

Yang, N., MacArthur, D.G., Gulbin, J.P., Hahn, A.G., Beggs, A.H., Easteal, S., and North, K. 2003. ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* **73**(3): 627-631.

Yang, Y., Chung, E.K., Wu, Y.L., Savelli, S.L., Nagaraja, H.N., Zhou, B., Hebert, M., Jones, K.N., Shu, Y., Kitzmiller, K., Blanchong, C.A., McBride, K.L., Higgins, G.C., Rennebohm, R.M., Rice, R.R., Hackshaw, K.V., Roubey, R.A., Grossman, J.M., Tsao, B.P., Birmingham, D.J., Rovin, B.H., Hebert, L.A., and Yu, C.Y. 2007. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* **80**(6): 1037-1054.

Ylstra, B., van den Ijssel, P., Carvalho, B., Brakenhoff, R.H., and Meijer, G.A. 2006. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res* **34**(2): 445-450.

Yokoi, F., Dang, M.T., Li, J., and Li, Y. 2006. Myoclonus, motor deficits, alterations in emotional responses and monoamine metabolism in epsilon-sarcoglycan deficient mice. *J Biochem* **140**(1): 141-146.

Young, J.W., Crawford, N., Kelly, J.S., Kerr, L.E., Marston, H.M., Spratt, C., Finlayson, K., and Sharkey, J. 2007. Impaired attention is central to the cognitive deficits observed in alpha 7 deficient mice. *Eur Neuropsychopharmacol* **17**(2): 145-155.

Zabetian, C.P., Hutter, C.M., Factor, S.A., Nutt, J.G., Higgins, D.S., Griffith, A., Roberts, J.W., Leis, B.C., Kay, D.M., Yearout, D., Montimurro, J.S., Edwards, K.L., Samii, A., and Payami, H. 2007.

Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann Neurol* **62**(2): 137-144.

Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.A., and Friedman, J.M. 2008. A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. *J Med Genet* **45**(4): 239-243.

Zalfa, F., Eleuteri, B., Dickson, K.S., Mercaldo, V., De Rubeis, S., di Penta, A., Tabolacci, E., Chiurazzi, P., Neri, G., Grant, S.G., and Bagni, C. 2007. A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat Neurosci* **10**(5): 578-587.

Zhang, Y., Roslan, R., Lang, D., Schachner, M., Lieberman, A.R., and Anderson, P.N. 2000. Expression of CHL1 and L1 by neurons and glia following sciatic nerve and dorsal root injury. *Mol Cell Neurosci* **16**(1): 71-86.

Zheng, B., Tang, T., Tang, N., Kudlicka, K., Ohtsubo, K., Ma, P., Marth, J.D., Farquhar, M.G., and Lehtonen, E. 2006. Essential role of RGS-PX1/sorting nexin 13 in mouse development and regulation of endocytosis dynamics. *Proc Natl Acad Sci U S A* **103**(45): 16776-16781.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. 2008. On the origin of new genes in Drosophila. *Genome Res* **18**(9): 1446-1455.

Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C., Scherer, S.W., Zanke, B., Hudson, T.J., Bader, G.D., and Gallinger, S. 2007. Germ-line DNA copy number variation frequencies in a large North American population. *Hum Genet* **122**(3-4): 345-353.

Zondervan, K.T. and Cardon, L.R. 2004. The complex interplay among factors that influence allelic association. *Nat Rev Genet* **5**(2): 89-100.

Appendix A

List of Oligonucleotides

Primer/Probe Name	Sequence
Long Range PCR for H6PD/SPSB1 Puplication	
SPSB1-LR-f1	CTGTGTATCCAGTAGGTGCTTAATACTTT
SPSB1-LR-f2	CTTGGGAGAGATGTTTTCTAGCAGTT
SPSB1-LR-f3	CTTGAGAAGCTCATTCTAACCTGTCTTAG
SPSB1-LR-f4	AGAAACTCCATCTACAGGTGCTGTC
SPSB1-LR-f5	TCTCTGTAAGTTCTGTGCCTCTCTCT
SPSB1-LR-f6	GTATCATGCACTAGTACTTTATGCCTTTTT
SPSB1-LR-f7	AGTGTAATCCTGATGGTTTTCATTTAATAG
SPSB1-LR-f8	CTTTATTAGGCAATTCTATGGTGACATC
SPSB1-LR-f9	CTATGTCCAAGAGACTGCCTTGTT
SPSB1-LR-r1	GTATGAGAGGCATCTATAAAGCACCTG
SPSB1-LR-r2	CACCTGCTCACTGTAAGACTGGTACT
SPSB1-LR-r3	CCTCTGTCAATGAGAACTACTGTAAGAGAA
SPSB1-LR-r4	AGTGGGAAGAAGAGAAAAGGAAGTAAAC
SPSB1-LR-r5	GTAACTGTGACGATAAAAGGGAAGAAC
SPSB1-LR-r6	TAACATAACAACATCAGAATAACAGCAGTG
SPSB1-LR-r7	ATAACACCAGCTTAACAACAGCAATATAAC
SPSB1-LR-r8	GCTTAATAACAGCAATATAACACCATTGTA
qPCR Validation Primers for H6PD/SPSB1 Duplication	
SPSB1-I-f	GGGGGTCATCGTGAATAATG
SPSB1-I-r	CCACTTGGACCACTTCCACT
SPSB1-II-f	ATGGGTTTGGTGTGGTTTGT
SPSB1-II-r	GAGTGGAGTGCAGGAAGGAG
qPCR Validation Primers for 10 CNV Loci	
ED1176-loss-17-A-f	CGAGAGTGACTCCTGTTTTTCC
ED1176-loss-17-A-r	AGCCCCCTTTCAAACCTAAA
ED1176-loss-17-B-f	TGGAGCCCATCTCACCTTTA
ED1176-loss-17-B-r	TGCTTCTGGATTTGTTGCTG
1085-loss-13-C-f	AGGCCAAAATGAATCTGGAG
1085-loss-13-C-r	CTCGCTGAACCACTTGAACA
3409-loss-23-B-f	CAGACTGCTGAAGCCATGAA
3409-loss-23-B-r	CAAACCACGAGCAAGCTGTA
4203-loss-28-B-f	TTGACCCTAACCAAGGATGC
4203-loss-28-B-r	TTCTTGGCTGGGATGTTTTC
5324-loss-21-B-f	GCATATTGCGATCCTGGAAT
5324-loss-21-B-r	CGTTTGGTTTGCTCCTCAAT
850-loss-22-B-f	AGCAGCCACATTGCCTTACT
850-loss-22-B-r	GGCAGCCTTAGTTTTCATGG
3584-gain-11-A-f	AATTGGGGGTGTGATGATGT
3584-gain-11-A-r	CAAATAAAACGTCTGGTTGCAG
3584-gain-11-B-f	AAACTATGCCCTGGCCTTTT
3584-gain-11-B-r	CGAAGCCACTCCTCTGACTC

Primer/Probe Name	Sequence
3766-gain-3-B-f	GCGCTGATGTTCTTTTCTCC
3766-gain-3-B-r	CAGCCATCAGCACTTTCAGA
3789-loss-42-A-f	ACCGCGTGCCTTAATTGTAT
3789-loss-42-A-r	TCTACCCACCCCTCTTCCTT
3789-loss-42-C-f	AAAGCGATGCTCCACAAACT
3789-loss-42-C-r	CATCGAAGCACGTGAACATT
4179-gain-19-A-f	CTGAGCCCCTGAGATCAGT
4179-gain-19-A-r	GAGAGTTCCCCAAAGCAG
PCR genotyping primers for *CHL1* 5' deletion	
CHL-fwd-5	TGATGATAATCCTCTTAGTTTTCATTT
wt-rev-2	GGGTTGGGAGTGGAATTGTA
del-rev-1	AGCCACAGGTTGTTCTCCAC
qPCR genotyping primers for *CHRFAM7A* deletion	
CHRFAM7A-MGB-fwd1	AGTAATAGTGTAATACTGTAACTTTAAAATGTGTTACTTGT
CHRFAM7A-MGB-rev1	AGCCGGGATGGTCTCGAT
B2M-MGB-fwd1	TGGGTTTCATCCATCCGACATT
B2M-MGB-rev1	AGACAAGTCTGAATGCTCCACTTTT
B2M_probe1	VIC-ATTCTTCAGTAAGTCAACTTC
CHRFAM7A_probe1	6FAM-TCCTGACTGTACACATAAAA

Appendix B

ABCA13 Custom Designed Oligonucleotide Array Validation

To identify copy number changes at the *ABCA13* region in families with schizophrenia, we designed a custom oligo array (Agilent Technologies) with 34,036 probes covering a 1.5-Mb interval at chromosome 7p12 spanning *ABCA13*. Validation experiments were performed to assess the probe dose response on the custom array by hybridising a series of 1x, 2x, 3x and 4x reference DNA (labelled in Cy5) against 1x reference DNA (labelled in Cy3). The expected log2ratio of each probe would be at 0, 1, 1.58 and 2 for the 4 experiments (corresponding to the 1x: 2x: 3x: 4x dosage). Analysis shows that majority of the oligo probes fall on the dosage response curves for the expected labelled DNA amount (Figure AI.1)

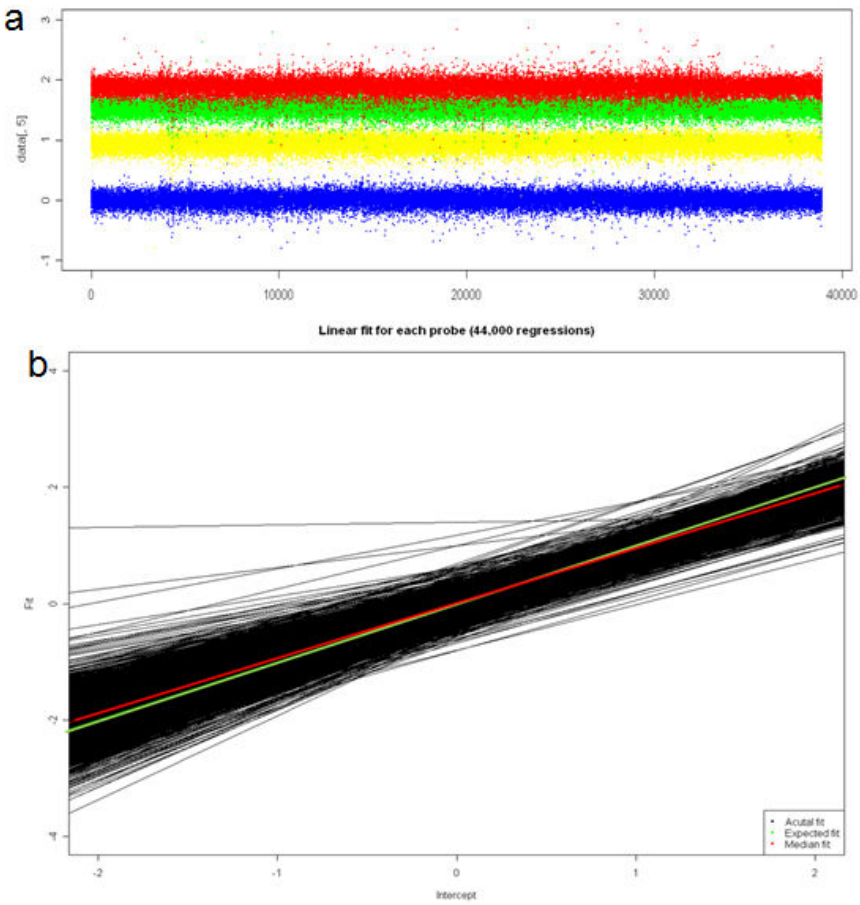


Figure B-1 Dose response of the custom *ABCA13* oligo array. a) A series of experiment were performed with 1x (blue), 2x (yellow), 3x (green) and 4x (red) reference DNA hybridised against 1x reference DNA. The expected log2ratio of each probe for the 4 experiments would be 0, 1, 1.58 and 2 respectively. The dose response for the majority of probes falls on the expected values. (x-axis: log2ratio, y-axis: genomic coordinates of the probes) **b)** Best-fit linear regression lines were generated for the expected and actual dosage response. The red line represents the expected fit. The actual fit for each probe is shown in black, and the median best-fit linear regression line for all probes is shown in green. The actual median fit (green) is close to the expected fit (red).

Appendix C

Whole Genome Tiling Path (WGTP) Array Quality Control

C1 Data Normalization

Data normalization, which includes background subtraction and spatial normalization, were performed as previously described (Fiegler et al. 2006, Redon et al. 2006). Some array profiles in our dataset demonstrated genomic "waves" in the signal intensities, a phenomenon which appeared to be a general feature of array CGH datasets (Marioni et al. 2007). One characteristic of these "waves" were their correlation to GC content of the array probes. All array profiles were therefore GC-normalized: the log2ratio of each clone in every sample is normalized against the GC content of the BAC clone using a custom PERL script (developed by Dr Richard Redon and Armand Valsesia).

C2 Clone Filtering

Each WGTP array constructed at the Sanger Institute is spotted with 29043 large insert clones from the "Golden Path" library (Fiegler et al., 2006). To eliminate possible artefacts due to individual clones, the following inspection and filtering steps were taken.

C2.1 Inconsistent and frequently excluded clones

Clones that showed poor morphology, spot intensities or low signal to background noise in individual experiment, or behaved inconsistently between dye swap experiments of a given sample, would be flagged as being excluded from the analysis of that DNA sample. Clones that were frequently excluded (in >75% DNA samples in both case and control cohorts) were subsequently filtered from the CNV dataset. 93 clones were filtered based on this criterion.

C2.2 Autosomal clones with homology to X/Y chromosome

Autosomal clones that have homology to either sex chromosome would give variable log2 signal ratios depending on whether hybridisations were performed with a male or female test DNA. These “pseudo-X/Y clones” were detected using a custom R script based on the difference between the average log2ratio of all female versus male test DNA hybridisations (see Fig C-1). 100 clones with a difference of >0.1 between average male and female log2ratio were excluded using this criterion. Examples of the histograms of two removed clones are shown in Fig C-1b.

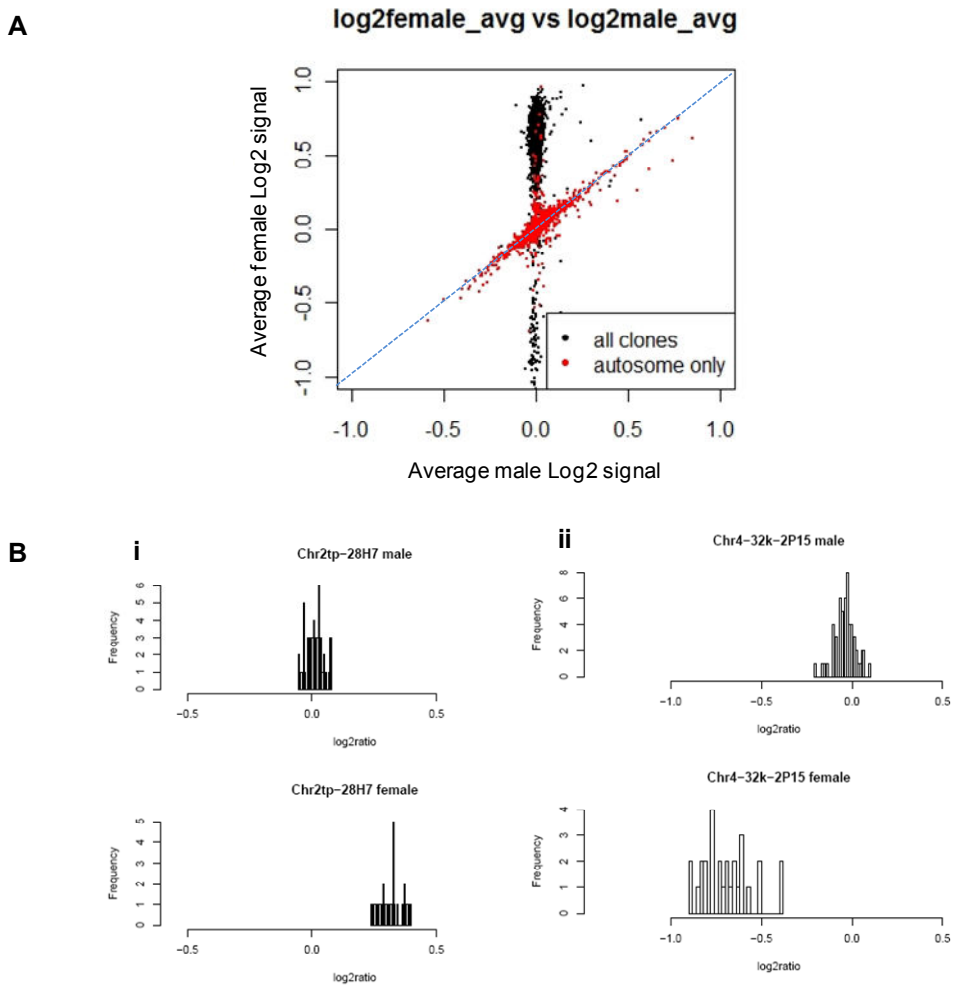


Fig C-1 Autosomal clones with homology to X/Y chromosomes. Autosomal clones with homology to X/Y chromosomes were detected using the average log2ratio values of all female versus all male test DNA experiments. (a) A plot of the average female log2ratio against the average male log2ratio for all clones. Majority of the autosomal clones lie on the x=y line (blue line). Clones above this line are X-chromosome clones or pseudo-X autosomal clones (filtered). Clones below this line are Y-chromosome clones or pseudo-Y autosomal clones (filtered). (b) Histograms of two filtered autosomal clones. i) a pseudo-X clone Chr2tp-28H7 and ii) a pseudo-Y clone Chr4-32K-2P15 (top panel: experiments from male DNAs, bottom panel: experiments from female DNAs)

C2.3 Artefacts due to hybridisation date

Clones were inspected with respect to the dates of hybridisation to detect artefacts due to environmental factors (e.g. variations in array batch used, ozone level and humidity level). When a set of clones with such artefact pattern was detected, the remaining 29,043 clones were screened by means of Pearson correlation coefficients to check for similar artefact patterns. A list of 110 clones showing two systematic date-specific artefact patterns were removed from further analysis (Figure C-2).

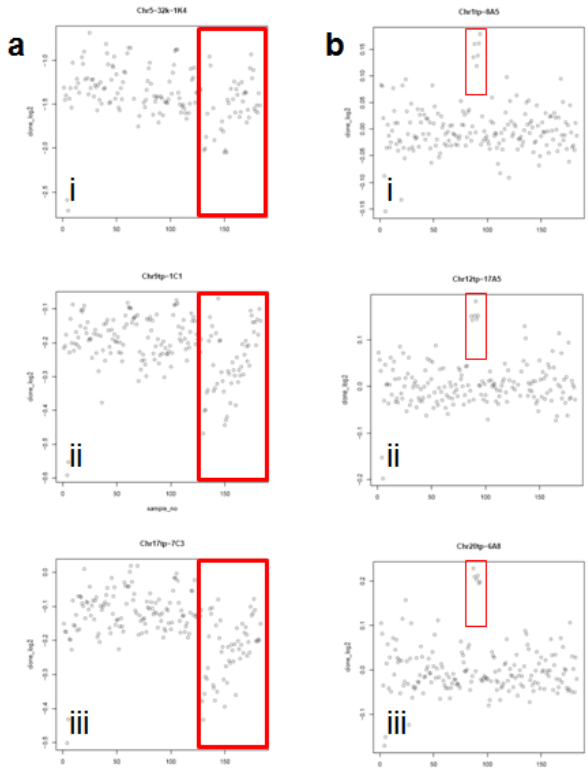


Fig C-2 BAC clones showing systematic date-specific artefact patterns. Two systematic date-specific artefact patterns were detected. Each plot shows clone log2ratio fluorescence signal against the DNA sample sorted by dates of hybridisation. (a) Hybridisations carried out prior to 05-Jul-2007 compared to those after 05-Jul-2007 i) Chr5-32K-1K4, ii) Chr5tp-1C1 and iii) Chr17tp-7C3). (b) Hybridisations carried out on Jun-20-2007 & June-21-2007 compared to the rest of hybridisations i) Chr1tp-8A5, ii) Chr12tp-17A5 and iii) Chr20tp-6A8).

C2.4 Sex chromosome artefacts in male Lothian Birth Control Samples

23 samples from the Lothian Birth Control (LBC) cohort demonstrated a mosaic Y chromosome pattern, such that clones on Y chromosome exhibited lower log2ratio fluorescence signal than in a typical male DNA sample hybridisation. These chromosomal artefacts may be due to Y chromosome loss from the LBC subjects, who were at relatively advanced age (~90 years old) at the time of DNA extraction. All non-autosomal clones were removed from these 23 LBC male samples from further analysis.

Appendix D

WGTP Array Supplementary Tables and Figures

Table D-1 Number and types of CNVs (Gain/ Loss) per DNA sample.

(a) SCZ samples (n=91) and **(b)** LBC samples (n=92)

a)

SCZ Sample	Gain	Loss	Total
211	29	22	51
220	27	17	44
241	20	13	33
297	42	21	63
302	56	24	80
323	31	37	68
358	34	17	51
385	55	45	100
513	52	31	83
621	36	39	75
665	27	14	41
850	33	16	49
899	36	26	62
1085	33	19	52
1278	54	24	78
1295	26	19	45
1784	28	26	54
3071	58	41	99
3138	41	30	71
3141	48	44	92
3198	33	16	49
3199	47	24	71
3333	36	37	73
3341	27	18	45
3377	34	23	57
3409	24	20	44
3443	44	33	77
3503	45	26	71
3559	53	41	94
3572	34	23	57
3581	51	37	88
3584	45	19	64
3594	28	15	43
3652	39	33	72
3661	33	26	59
3703	30	13	43
3751	54	28	82
3766	31	19	50
3776	24	43	67
3789	54	35	89
3802	41	21	62
3812	57	23	80
3815	47	28	75
3857	46	38	84
3931	54	17	71
3945	24	25	49
3975	36	12	48
3976	27	13	40
4042	44	25	69
4100	40	36	76
4101	36	22	58
4111	48	31	79
4160	46	31	77
4179	27	14	41
4203	43	27	70
4355	43	30	73
4398	37	38	75
4495	37	36	73
4710	42	28	70
4716	51	33	84
4748	54	22	76
4773	47	19	66
5307	34	26	60
5324	57	24	81
5379	41	20	61
5386	55	40	95
5390	65	30	95
5446	46	24	70
5467	48	41	89
5541	24	20	44
5562	20	18	38
5587	41	39	80
5660	61	22	83
5758	50	30	80
5792	29	15	44
6273	59	32	91
6276	26	14	40
6289	73	34	107
6596	29	17	46
6626	53	46	99
6634	41	32	73
6638	32	25	57
6667	33	16	49
6744	39	20	59
7019	32	22	54
7103	38	16	54
7132	35	20	55
7180	10	21	31
7183	24	18	42
7287	56	43	99
7294	40	38	78
Grand Total	**3650**	**2386**	**6036**

b)

LBC Sample	Gain	Loss	Total
D003	51	28	79
D007	39	29	68
D008	61	26	87
D011	56	41	97
D012	34	22	56
D015	40	12	52
D019	46	27	73
D020	41	25	66
D026	42	30	72
D027	24	27	51
D032	29	22	51
D033	58	34	92
D035	46	26	72
D045	37	36	73
D049	31	18	49
D052	31	23	54
D053	33	32	65
D055	37	16	53
D058	58	34	92
D059	47	59	106
D060	47	32	79
D062	52	33	85
D075	45	39	84
D078	56	34	90
D086	37	37	74
D088	44	20	64
D098	44	31	75
D107	56	30	86
D127	52	30	82
D158	29	17	46
D161	26	6	32
D162	28	16	44
D164	37	18	55
D168	29	18	47
D176	31	8	39
D209	34	25	59
D211	35	32	67
D220	35	39	74
D223	46	38	84
D224	45	38	83
D226	43	34	77
D234	45	28	73
D235	50	41	91
D237	54	31	85
D242	44	32	76
D244	41	15	56
D245	25	15	40
D246	32	26	58
D255	59	45	104
D263	54	41	95
D269	43	25	68
D272	41	35	76
D283	38	33	71
D287	52	45	97
D301	41	27	68
D328	34	37	71
D331	53	22	75
D332	55	38	93
D337	51	41	92
D341	69	34	103
D357	42	27	69
D365	44	33	77
D371	35	34	69
D379	49	32	81
D390	41	29	70
D391	45	39	84
D392	47	31	78
D399	42	19	61
D402	86	34	120
D435	50	38	88
D455	66	50	116
D477	43	22	65
D485	46	47	93
D489	72	36	108
D508	44	51	95
D514	52	33	85
D515	28	24	52
D518	26	11	37
D520	18	17	35
D521	24	26	50
D522	31	30	61
D531	36	20	56
D538	37	26	63
D548	44	29	73
LBC1	23	20	43
LBC10	76	16	92
LBC2	52	34	86
LBC3	32	31	63
LBC4	28	33	61
LBC6	31	22	53
LBC8	57	17	74
LBC9	53	25	78
Grand Total	**3973**	**2689**	**6662**

Table D-2 113 Variants Specific to the SCZ Cohort

	Chr Coordinates	Clones	Genes
1	chr1:2062380-2546655	Chr1tp-25A3, Chr1tp-25A4, Chr1tp-1A1, Chr1tp-1A2, Chr1tp-25A2, Chr1tp-25A1	PEX10, C1orf93, SKI, C1orf86, HES5, RER1, PRKCZ, TNFRSF14, MORN1, MMEL1, PLCH2, PANK4
2	chr1:2847618-3355092	Chr1tp-25B4, Chr1tp-1A8, Chr1tp-1A9, Chr1tp-25B3, Chr1tp-1A7	ACTRT2, PRDM16
3	chr1:9177728-9373336	Chr1tp-23A3, Chr1tp-25D10, Chr1tp-1C7	H6PD, SPSB1
4	chr1:68325266-68897278	Chr1tp-4A8, Chr1tp-20H2, Chr1tp-28G5, Chr1tp-23D1, Chr1tp-28G4, Chr1tp-28G3, Chr1tp-36B4	DEPDC1, RPE65, GPR177
5	chr1:71346768-71893785	Chr1tp-4C1, Chr1tp-4B12, Chr1tp-38E10, Chr1tp-18C6	NEGR1
6	chr1:105728511-105980971	Chr1tp-6E3, Chr1tp-6E4	
7	chr1:111515826-111802379	Chr1tp-30E4, Chr1tp-6H4	CHIA, C1orf88, DENND2D, OVGP1, WDR77, CHI3L2, ATP5F1, CEPT1
8	chr1:205624502-205991690	Chr1tp-21E1, Chr1tp-11G7	CR1L, CR2, CR1
9	chr1:232939949-233194586	Chr1tp-13G11, Chr1tp-13H1	
10	chr1:246013817-246295055	Chr1tp-15A5, Chr1tp-15A4	OR2AK2, OR2AJ1, OR2L13, OR2L3, OR14A16, OR2L8, OR2L2, OR2W3, TRIM58, OR2L5, OR11L1
11	chr2:731429-1026826	Chr2tp-3D6, Chr2tp-29G2, Chr2tp-31F10	SNTG2
12	chr2:37650158-37992443	Chr2tp-16B5, Chr2tp-5F10, Chr2tp-22C3	CDC42EP3, SLC30A6
13	chr2:57138035-57341900	Chr2tp-24G11, Chr2tp-1D9	
14	chr2:86130488-86470772	Chr2tp-5F6, Chr2tp-22D2	REEP1, PTCD3, POLR1A, IMMT, MRPL35
15	chr2:127333466-127814482	Chr2-32k-1B7, Chr2tp-7B12, Chr2tp-12G11, Chr2tp-7E6	ERCC3, BIN1, MAP3K2, CYP27C1
16	chr2:132668884-132859586	Chr2tp-16E12, Chr2tp-11B5	
17	chr2:188772739-189096411	Chr18tp-1A6, Chr2tp-7E12, Chr2tp-4H8	GULP1
18	chr2:213735598-214074098	Chr2-32k-4O4, Chr2tp-29F9	SPAG16
19	chr3:8686094-8880344	Chr3tp-27B5, Chr3tp-4D7	OXTR, CAV3, C3orf32
20	chr3:76248984-77652531	Chr3tp-11A8, Chr3tp-11H6, Chr3tp-4H9, Chr3-32k-6M10, Chr3tp-27A6	ROBO2
21	chr3:86998287-87362355	Chr3-32k-2L14, Chr3tp-20B2, Chr3tp-18G12	CHMP2B, VGLL3
22	chr3:139094479-139416525	Chr3tp-10F10, Chr3tp-18G6	ARMC8, DZIP1L, CLDN18, A4GNT, DBR1
23	chr3:143326112-143612526	Chr3tp-11B1, Chr3tp-1G2	XRN1, TFDP2, GK5

	Chr Coordinates	Clones	Genes
24	chr3:189546565-189761158	Chr3tp-9H12, Chr3tp-18F1	LPP
25	chr4:7039049-7320802	Chr4tp-21E8, Chr4tp-7G10	SORCS2, TBC1D14, GRPEL1, CCDC96
26	chr4:11668862-11990799	Chr4-32k-3N20, Chr4tp-9C5, Chr4tp-1B4	
27	chr4:28258974-28570510	Chr4tp-6E1, Chr4tp-16E1, Chr4tp-6C4	
28	chr4:115564534-115809774	Chr4tp-13A7, Chr4tp-14D9, Chr4-32k-4B10	UGT8
29	chr4:162138593-162324657	Chr4tp-17A12, Chr4tp-3E1	
30	chr4:162734833-163023593	Chr4tp-3D7, Chr4tp-25C4, Chr4tp-21C3	FSTL5
31	chr4:168885634-169288268	Chr4tp-10G12, Chr4tp-9F9	ANXA10
32	chr4:189077762-189851429	Chr4tp-17E1, Chr4tp-17D1, Chr4tp-7D1, Chr7tp-16B7, Chr4tp-5C8, Chr4tp-24A9	TRIML2, ZFP42, TRIML1
33	chr5:9850992-10044066	Chr5tp-18A3, Chr5tp-5H11	
34	chr5:25051484-25333181	Chr5tp-6H2, Chr5tp-3E11	
35	chr5:61438895-61723574	Chr5tp-19D4, Chr5tp-17G10	DIMT1L, KIF2A
36	chr5:84073606-84367471	Chr5tp-24B1, Chr5tp-6E5	
37	chr5:101134363-101393863	Chr5tp-25F12, Chr5tp-2A1	
38	chr5:104334472-104660077	Chr5tp-23B3, Chr5tp-1B3, Chr5tp-15B7, Chr5tp-23E12	
39	chr5:110212743-110577198	Chr5tp-15G1, Chr5tp-24F9, Chr5-32k-5C18	WDR36, TSLP
40	chr6:1994541-2304998	Chr6tp-4B5, Chr6tp-4D5, Chr6tp-5A4	GMDS
41	chr6:7088879-7351773	Chr6tp-1C9, Chr6tp-3B7	RREB1, CAGE1, RIOK1, SSR1
42	chr6:11754000-11930337	Chr6tp-5E7, Chr6tp-6A12	C6orf105
43	chr6:22212251-22607763	Chr6tp-3C9, Chr6tp-6A8, Chr6tp-3F3, Chr6tp-3B3	PRL
44	chr6:34358245-34603317	Chr6tp-2C4, Chr6tp-4H8	NUDT3, PACSIN1, RPS10
45	chr6:97469365-98475130	Chr6tp-18H7, Chr6tp-15B2, Chr6tp-8A7, Chr6tp-16H7, Chr6tp-17D11, Chr6tp-17A3, Chr6tp-11E8, Chr6tp-16D3, Chr6tp-13C2, Chr6tp-12G11	KLHL32, C6orf167
46	chr6:143958314-144219762	Chr6tp-11C7, Chr6tp-18B6	PHACTR2, LTV1
47	chr6:160857306-161154655	Chr6tp-14D11, Chr6tp-17C6	LPA, PLG
48	chr6:165840859-166673790	Chr6tp-8B6, Chr6tp-19H8, Chr6tp-14C5, Chr6tp-15E7, Chr6tp-8C3, Chr6tp-19A1, Chr6tp-10C12, Chr6tp-14F5	SFT2D1, T, PDE10A, PRR18
49	chr7:885103-1354746	Chr7tp-14F1, Chr7tp-4D8	CYP2W1, CENTA1, UNCX, C7orf50, GPR146, COX19, GPER, C7orf20, ZFAND2A
50	chr7:17320942-17609921	Chr7tp-4C1, Chr7tp-5D12	AHR

	Chr Coordinates	Clones	Genes
51	chr7:17795286-18873121	Chr7tp-23G6, Chr7tp-18H11, Chr7tp-10E12, Chr7tp-3E3, Chr7tp-22H9, Chr7tp-21E3, Chr7tp-1H3, Chr7tp-17H9, Chr7tp-5E2, Chr7-32k-2I12	SNX13, HDAC9
52	chr7:21281130-21501769	Chr7tp-21E8, Chr7tp-21A5	SP4
53	chr7:23686581-23846847	Chr7tp-14F2, Chr7tp-19E4	C7orf46, STK31
54	chr7:75289669-75520758	Chr7-32k-6E9, Chr7tp-17H6	TMEM120A, RHBDD2, POR, STYXL1, MDH2
55	chr7:88344724-88962788	Chr7tp-19G12, Chr7tp-10D4, Chr7tp-18A2, Chr7tp-18C10, Chr7tp-20H7, Chr7tp-22H12, Chr7tp-16C8	ZNF804B
56	chr7:93658283-94064936	Chr7tp-1E6, Chr7tp-17D3, Chr7tp-21F8	SGCE, CASD1, COL1A2
57	chr7:123672713-123975901	Chr7tp-3C10, Chr7tp-9B6, Chr7-32k-3I18	CENTG3
58	chr7:124338397-124505938	Chr7tp-3G9, Chr7tp-3F4	CENTG3, POT1
59	chr8:9670799-9988807	Chr8tp-20B12, Chr8tp-6F6	TNKS, MSRA
60	chr8:13581607-13896999	Chr8tp-7A2, Chr8tp-7D2	
61	chr8:47038753-47556339	Chr8tp-9B4, Chr8tp-19B9, Chr8tp-13H11, Chr8tp-19C4	
62	chr8:47728697-47982207	Chr8tp-5G11, Chr8tp-3C12	
63	chr8:48950846-49170567	Chr8tp-17F4, Chr8tp-16B12, Chr8tp-3E4	PRKDC, UBE2V2, MCM4
64	chr8:145496837-145824934	Chr8-32k-5I13, Chr8tp-16F9	GPT, CYHR1, GPR172A, PPP1R16A, KIFC2, FOXH1, HSF1, SCRT1, NFKBIL2, RECQL4, MFSD3, CPSF1, SLC39A4, DGAT1, ADCK5, LRRC14, VPS28, FBXL6, LRRC24
65	chr9:76652583-77315436	Chr9tp-4G5, Chr9tp-5C11, Chr9tp-6D10, Chr9tp-8C4, Chr9tp-5C12	C9orf41, C9orf95, TRPM6, OSTF1, C9orf40
66	chr9:137112070-137464520	Chr9tp-10A3, Chr9tp-7H9	C9orf62, OLFM1
67	chr9:137590464-137867305	Chr9tp-3C2, Chr9tp-3B10	LCN9, CAMSAP1, SOHLH1, KCNT1, PAEP, GLT6D1
68	chr10:3722438-4002432	Chr10tp-9H1, Chr18tp-10B5, Chr10tp-4H7	KLF6
69	chr10:42524238-42830197	Chr10tp-10B2, Chr10tp-3D8, Chr10tp-12D4	BMS1
70	chr10:76904943-77122607	Chr10tp-13C7, Chr10tp-6G6	C10orf11
71	chr11:31339431-31549584	Chr11tp-13E3, Chr11tp-13E8	DPH4, DCDC1, IMMP1L, ELP4
72	chr11:50013000-50737095	Chr11tp-7H3, Chr11tp-11A11, Chr11tp-11E7, Chr11tp-16H3, Chr11tp-3E2, Chr11tp-2B4, Chr11tp-5F6	OR4A5

	Chr Coordinates	Clones	Genes
73	chr11:58456668-58635415	Chr11tp-2G5, Chr11tp-13E10	FAM111B, GLYATL1
74	chr12:17761730-18154145	Chr12tp-12A7, Chr12tp-5D12	RERGL
75	chr12:51437221-51449660	Chr12tp-10A4, Chr12tp-11A11	KRT76
76	chr12:131115602-131701464	Chr0-32k-1C11, Chr12tp-7A9	EP400NL, GALNT9, EP400, MUC8, NOC4L, DDX51
77	chr13:18881229-19386914	Chr13tp-5D8, Chr13tp-1B2, Chr13tp-8F2, Chr13tp-10A5	MPHOSPH8, PSPC1, ZMYM5, TPTE2
78	chr13:51626371-51966133	Chr13tp-11D8, Chr13tp-11E5	THSD1, VPS36, CKAP2, NEK3
79	chr13:82035430-82280653	Chr13tp-8E10, Chr13tp-1B10	
80	chr13:86846920-87112550	Chr13tp-1F2, Chr13tp-5B8	
81	chr13:112568275-112939870	Chr13tp-1D12, Chr13tp-4E6, Chr13tp-3B4, Chr13tp-10G6	PROZ, ATP11A, MCF2L, PCID2, F10, CUL4A, F7
82	chr13:113166605-113658050	Chr13tp-11B6, Chr13tp-11A8, Chr13tp-3D4	TMCO3, TFDP1, GAS6, FAM70B, DCUN1D2, GRK1, ATP4B
83	chr14:74945402-75396561	Chr14tp-8C7, Chr14tp-10B10, Chr14tp-3B9	JDP2, C14orf1, TTLL5, FLVCR2, BATF
84	chr14:103592318-103986703	Chr14-32k-3E3, Chr14-32k-2C12, Chr14tp-2C5	KIF26A
85	chr15:22894873-23111579	Chr15tp-7D1, Chr15tp-11C3	
86	chr15:25060215-25335769	Chr15tp-7C1, Chr15tp-4A1	GABRG3
87	chr15:41610450-41935933	Chr15tp-3H4, Chr15tp-3G7	STRC, CATSPER2, SERF2, CKMT1B, MAP1A, ELL3, SERINC4, HISPPD2A, MFAP1, CKMT1A, PDIA3, WDR76
88	chr15:83451942-84109983	Chr15tp-10F9, Chr15tp-10H5, Chr15tp-5E3, Chr15tp-7G3, Chr15tp-11G11	PDE8A, AKAP13, KLHL25
89	chr15:99707205-100036184	Chr15tp-1C4, Chr15tp-5B5	TM2D3, PCSK6, TARSL2
90	chr16:2870892-3098404	Chr16tp-13C6, Chr16-32k-2L14	KREMEN2, HCFC1R1, CLDN9, IL32, MMP25, CCDC64B, TNFRSF12A, FLYWCH1, ZSCAN10, PAQR4, FLYWCH2, THOC6, PKMYT1, CLDN6
91	chr16:57136506-57433538	Chr16tp-12E8, Chr16tp-3F2	GOT2, SLC38A7, CNOT1
92	chr17:61780311-61961765	Chr17tp-6A10, Chr17tp-11D10	PRKCA
93	chr17:74849624-75362772	Chr17-32k-2I12, Chr17-32k-2J21, Chr17tp-6B11	ENPP7
94	chr17:76741026-77562577	Chr17-32k-2P9, Chr17tp-3H2, Chr17-32k-2B20	C17orf55, NPB, SLC38A10, MRPL12, NOTUM, C17orf70, BAHCC1, THOC4, AATK, CCDC137, SLC25A10, HGS, FSCN2, ARL16, C17orf90, ANAPC11, TMEM105, AZI1, P4HB, C17orf56, PDE6G, PYCR1, TSPAN10, MAFG, ASPSCR1, ACTG1, ARHGDIA, PCYT2, DYSFIP1, NPLOC4, SIRT7

	Chr Coordinates	Clones	Genes
95	chr18:37767209-38078089	Chr18tp-5C12, Chr18-32k-1L20	PIK3C3
96	chr18:64349737-64656600	Chr18tp-1H10, Chr18tp-5H6	CCDC102B, TXNDC10
97	chr19:20265108-20637787	Chr19tp-6G2, Chr19tp-2B11, Chr19tp-6G2, Chr19tp-8A10	ZNF826
98	chr19:50553125-51160225	Chr19-32k-2E17, Chr19-32k-3I5	GPR4, DMPK, CD3EAP, IRF2BP1, VASP, QPCTL, EML2, NOVA2, NANOS2, DMWD, PPP1R13L, GIPR, RSHL1, SYMPK, FOXA3, FOSB, SIX5, ERCC1, OPA3, SNRPD2, ERCC2, RTN2
99	chr20:14561835-14936485	Chr20tp-5G6, Chr20tp-5G3, Chr20tp-6C12, Chr20tp-1H11	MACROD2
100	chr20:29267570-29480559	Chr20tp-6F10, Chr20tp-5G2	DEFB121, DEFB117, DEFB115, DEFB119, DEFB116, DEFB118
101	chr20:61940840-62284575	Chr20tp-1F12, Chr20tp-4B8, Chr20tp-2H9, Chr20tp-2H7	RGS19, MYT1, SAMD10, TCEA2, PRPF6, UCKL1, PRR17, DNAJC5, SOX18, OPRL1, ZNF512B, UCKL1OS, C20orf135, TPD52L2, NPBWR2
102	chr20:62108666-62376639	Chr20tp-2H9, Chr20tp-2H7, Chr20tp-6G1,	OPRL1, SOX18, RGS19, MYT1, TCEA2, PRPF6, ZNF512B, PRR17, NPBWR2, PCMTD2
103	chr21:34576662-34852873	Chr21tp-3F8, Chr21tp-3A5	KCNE2, RCAN1, C21orf51, KCNE1
104	chr21:44384865-44825939	Chr21-32k-2C5, Chr1-32k-7C17, Chr21-32k-2E17, Chr21-32k-2G11	TRPM2, C21orf33, ICOSLG, C21orf2, KRTAP10-2, PFKL, KRTAP10-5, KRTAP10-1, AIRE, C21orf29, DNMT3L, C21orf90, KRTAP10-3, LRRC3, KRTAP10-4
105	chr21:45508487-45829731	Chr21-32k-1N9, Chr21-32k-1B12, Chr21-32k-1L4	C21orf93, SLC19A1, C21orf123, COL18A1, POFUT2
106	chr22:39890708-40040762	Chr22tp-9A1, Chr22tp-8B6	CHADL, RANGAP1, L3MBTL2, EP300, ZC3H7B
107	chr22:46219057-46441810	Chr22tp-6F7, Chr22tp-4A6, Chr22-32k-1H11, Chr22tp-4H5	
108	chr22:46903974-46908346	Chr22tp-7A3, Chr22tp-7A10	
109	chrX:51220257-51533184	ChrXtp-21H12, ChrXtp-20B10, ChrXtp-4E12	NUDT11, GSPT2
110	chrX:56037801-56325064	ChrXtp-11B12, ChrXtp-6A3, ChrXtp-21B4	KLF8
111	chrX:79042086-79447858	ChrXtp-9E7, ChrXtp-23D8, ChrXtp-10A5, ChrXtp-4H6, ChrXtp-11A8	TBX22
112	chrX:105332666-105446815	ChrXtp-11H2, ChrXtp-7E1	MUM1L1
113	chrY:26625200-26913359	ChrYtp-1G9, ChrYtp-2D10	

Table D-3 CNV Genotyping of WGTP Data With Bivariate Clustering

31 regions demonstrate significant difference (p-value < 0.05) in the CNV genotype distributions between SCZ and LBC following bivariate cluster of log2ratio. For each region the p-value of chi-squared test was shown.

	clones	SCZ class	LBC class	p-val	chrom coordinates	Genes
1	Chr3-32k-5N10,**Chr3tp-9C3**,Chr3tp-7F8	4_25_42	1_16_62	0.02711	chr3:46141-343409	**CHL1***
2	Chr17tp-8C2,Chr17tp-12B1,Chr17tp-2F6,Chr17tp-14B12,**Chr17tp-14B6**,Chr17tp-2F9,Chr17tp-1D11,Chr17tp-11E1	31_29_19_6_3	22_26_23_17_0	0.03517	chr17:41747597-42632332	ARL17P1, LRRC37A, LRRC37A2, AC139026.2, **NSF***, AC004098.1, WNT3, WNT9B, AC005670.1, GOSR2, RPRML, AC068152.11, LRRC37A4., CDC27
3	Chr15-32k-1I14,**Chr15tp-7E12**,Chr15tp-7C6,Chr15tp-9C6,Chr15tp-1G5,Chr15tp-1G2,Chr15tp-4F3,Chr15tp-12A3,Chr15tp-7H3	17_72	7_83	0.04511	chr15:26714864-28805754	AC055876.16, AC055876.16, AC055876.16, AC055876.1, APBA2, AC024474.8, NDNL2, AC102941.9, AC102941.9, TJP1, AC022613.13, AC111152.2, AC120045.19, AC120045.19, AC120045.19, AC120045.19, AC120045.19, AC120045.19, AC135731.6, AC135731.6, AC019322.8, **CHRFAM7A***, AC019322.8, AC019322.8, AC019322.8, AC026150.7, AC026150.7, AC026150.7, AC026150.7, AC026150.7, ARHGAP11B, AC091057.6, AC091057.6
4	**Chr13-32k-3D13**,Chr13tp-7D5	50_15_9	71_7_0	0.00044	chr13:44704567-48370518	GTF2F2, AL138963.20, TPT1, AL138963.20, AL627107.32, AL627107.32, SLC25A30, AL606514.3, COG3, AL139326.15, AL139326.15, SPERT, AL139320.18, AL157819.15, ZC3H13, CPB2, LCP1, AL137141.10, AL139801.17, AL137141.10, AL139801.17, C13orf18, AL139801.17, N/A, AL138686.14, RP11-189B4.3, N/A, LRCH1, ESD, HTR2A, AL160397.17, AL160397.17, AL359180.17, SUCLA2, NUDT15, MED4, N/A, N/A, AL158196.24, ITM2B, RB1, AL392048.9, P2RY5, AL136960.4, RCBTB2, CYSLTR2, AL137118.20, GTF2F2, AL138963.20, TPT1, AL138963.20, AL627107.32, AL627107.32, SLC25A30, AL606514.3, COG3, AL139326.15, AL139326.15, SPERT, AL139320.18, AL157819.15, ZC3H13, CPB2, LCP1, AL137141.10, AL139801.17, AL137141.10, AL139801.17, C13orf18, AL139801.17, N/A, AL138686.14, RP11-189B4.3, N/A, LRCH1, ESD, HTR2A, AL160397.17, AL160397.17, AL359180.17, SUCLA2, NUDT15, MED4, N/A, N/A, AL158196.24, ITM2B, RB1, AL392048.9, P2RY5, AL136960.4, RCBTB2, CYSLTR2, AL137118.20
5	**Chr8-32k-4F15**,Chr8tp-21E9,Chr8tp-16F9,Chr8-32k-5I13	63_16	75_1	0.00044	chr8:144483554-145824934	TOP1MT, AC087793.17, N/A, RHPN1, MAFA, ZC3H3, AC067930.7, GSDMDC1, N/A, C8orf73, N/APRT1, EEF1D, TIGD5, PYCRL, TSTA3, ZNF623, ZNF707, AC105219.6, MAPK15, FAM83H, N/A, AC105219.6, SCRIB, hsa-mir-937, PUF60, NRBP2, PLEC1, hsa-mir-661, PARP10, GRIN/A, SPATC1, AC109322.16, OPLAH, EXOSC4, GPAA1, CYC1, SHARPIN, MAF1, KIAA1875, C8orf30A, AC145291.2, AC145291.2, AC145291.2, C8orf30B, AC110280.8, BOP1, SCXA, HSF1, N/A, DGAT1, SCRT1, AF205589.5, FBXL6, GPR172A, ADCK5, CPSF1, hsa-mir-939, SLC39A4, VPS28, NFKBIL2, GPT, AC084125.9, CYHR1, KIFC2, FOXH1, PPP1R16A, MFSD3, RECQL4, LRRC14, LRRC24, AC084125.9, AC084125.9

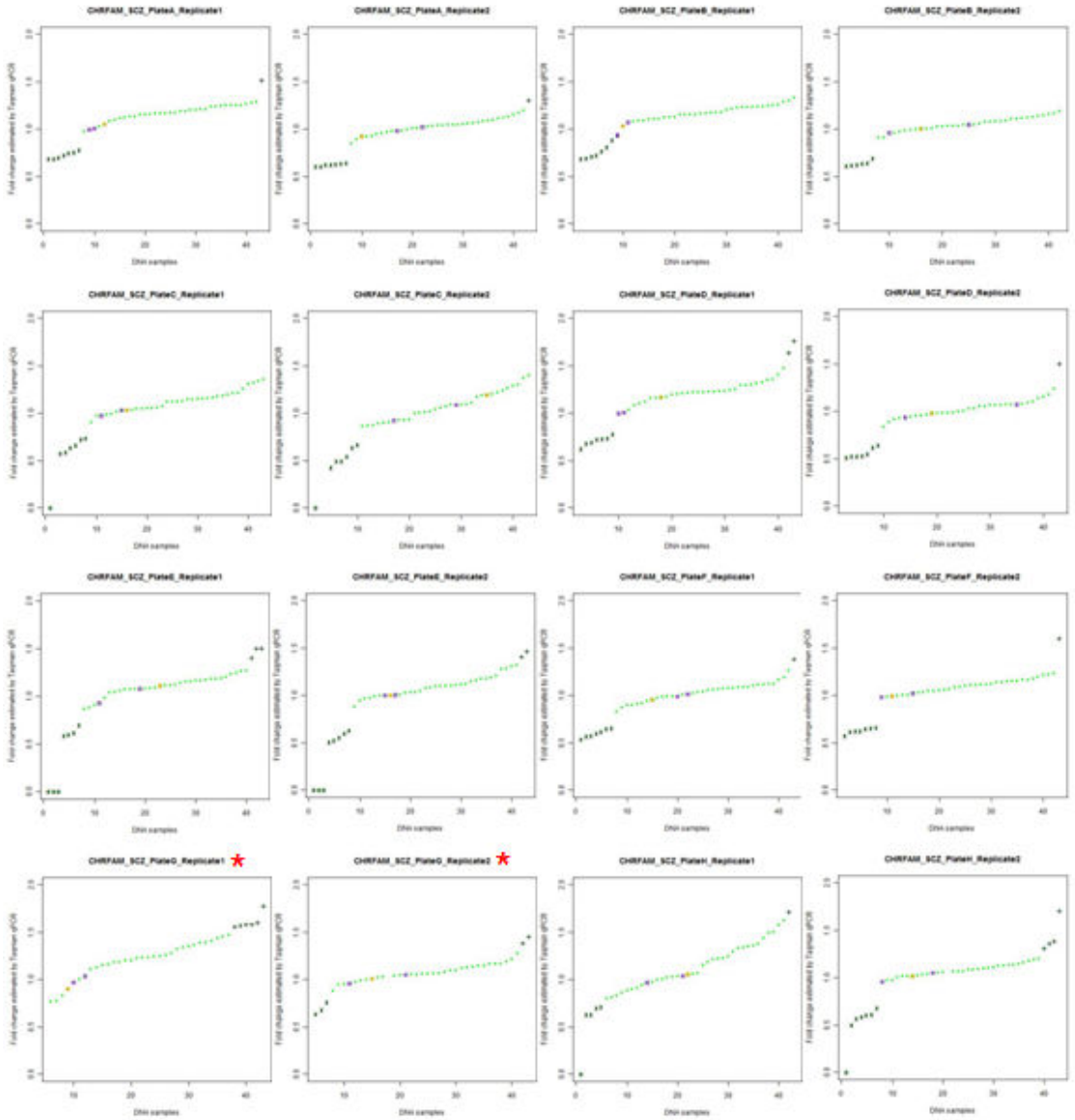
	clones	SCZ class	LBC class	p-val	chrom coordinates	Genes
6	Chr9tp-3C2,**Chr9tp-3E10**,Chr9tp-2H8	54_14	76_2	0.00132	chr9:137679970-138276710	LCN9, SOHLH1, KCNT1, CAMSAP1, QSOX2, UBAC1, BTBD14A, N/A, LHX3, N/A
7	**Chr22tp-7A3**,Chr22tp-7A10	72_9	91_0	0.00346	chr22:46903974-46908346	N/A
8	**Chr4-32k-3B10**,Chr4-32k-4N21	0_64_13	1_84_2	0.0037	chr4:135111708-135433169	N/A
9	**Chr7tp-5H9**,Chr7tp-11F4	35_34_0	18_50_2	0.00526	chr7:69823647-70123482	AUTS2
10	**Chr19tp-6D4**,Chr19-32k-3I21,Chr19-32k-2N13,Chr19-32k-3O11	52_16	75_5	0.00568	chr19:1331124-2345413	NDUFS7, AC005329.1, GAMT, DAZAP1, RPS15, N/A, APC2, AC027307.5, PCSK4, REEP6, ADAMTSL5, AC027307.5, RKHD1, AC005943.1, MBD3, UQCR, TCF3, ONECUT3, N/A, ATP8B3, REXO1, KLF16, FAM108A4, AC012615.4, ADAT3, AC012615.4, CSNK1G2, BTBD2, MKNK2, MOBKL2A, C19orf36, AP3D1, DOT1L, PLEKHJ1, SF3A2, AMH, JSRP1, OAZ1, C19orf35, N/A, LSM7., AC004410.1, TMPRSS9
11	Chr19tp-6D4,**Chr19-32k-3I21**,Chr19-32k-2N13,Chr19-32k-3O11	48_14	79_5	0.00687	chr19:1331124-2345413	NDUFS7, AC005329.1, GAMT, DAZAP1, RPS15, N/A, APC2, AC027307.5, PCSK4, REEP6, ADAMTSL5, AC027307.5, RKHD1, AC005943.1, MBD3, UQCR, TCF3, ONECUT3, N/A, ATP8B3, REXO1, KLF16, FAM108A4, AC012615.4, ADAT3, AC012615.4, CSNK1G2, BTBD2, MKNK2, MOBKL2A, C19orf36, AP3D1, DOT1L, PLEKHJ1, SF3A2, AMH, JSRP1, OAZ1, C19orf35, N/A, LSM7., AC004410.1, TMPRSS9
12	Chr2tp-25B10,**Chr2tp-10C11**,Chr2tp-5E12,Chr2tp-4E12,Chr2tp-28C3,Chr2-32k-4A6,Chr2tp-17D7,Chr2tp-16E3,Chr2-32k-4M13	2_8_39_9_1_1	1_1_55_3_6_0	0.00737	chr2:87099273-87904292	AC111200.7, AC092651.3, AC083899.6, AC068279.6, AC133644.1, AC133644.3, N/A, N/A, PLGLB2, RGPD2
13	**Chr7tp-2C3**,Chr7tp-19G6	1_87_0	5_78_6	0.01029	chr7:148684699-148876192	AC004941.2, AC004941.2, ZNF777, ZNF746, ZNF767
14	**Chr13tp-4E6**,Chr13tp-10G6,Chr13tp-1D12,Chr13tp-3B4	22_49_1	10_70_0	0.01222	chr13:112568275-112939870	ATP11A, MCF2L, AL137002.19, F7, F10, PROZ, PCID2, CUL4A
15	Chr4tp-24A9,**Chr4tp-17E1**,Chr4tp-5C8,Chr7tp-16B7	1_7_83	0_0_92	0.01457	chr4:189169360-189702507	TRIML2, AC138781.4, TRIML1
16	**Chr5-32k-6H23**,Chr5tp-5F1	74_9_0	88_1_1	0.01545	chr5:3262387-3435889	N/A
17	Chr1tp-23D1,Chr1tp-4A8,Chr1tp-36B4,**Chr1tp-20H2**,Chr1tp-28G3,Chr1tp-28G4,Chr1tp-4B12	81_6_1	88_0_0	0.02612	chr1:68325266-71516648	N/A, GPR177, RPE65, DEPDC1, LRRC7, AL391728.19, N/A, LRRC40, AL353771.8, SFRS11, ANKRD13C, HHLA3, CTH, AL354872.9, PTGER3, ZRANB2, hsa-mir-186
18	**Chr15tp-12A11**,Chr15tp-4A10	85_0	84_7	0.02618	chr15:63048238-63255120	SPG21, MTFMT, AC013553.14, RASL12, PDCD7, CLPX
19	**Chr12-32k-3H21**,Chr12tp-18A3,Chr12tp-11G1,Chr12tp-10D9	1_5_52_10	2_17_38_9	0.02823	chr12:81598000-81881752	TMTC2
20	**Chr1tp-22B12**,Chr1tp-8H4,Chr1-32k-1C17	6_6_47_9_5_2	3_4_57_110_0	0.02937	chr1:159672939-159891049	FCGR2A, HSPA6, FCGR3A, AL451067.12, FCGR3B
21	**Chr4tp-14G7**,Chr4tp-1H7	9_45_27	8_32_46	0.02937	chr4:34313166-34632938	N/A
22	Chr17tp-11A6,Chr17tp-6D1,Chr17tp-2G12,**Chr17tp-14C12**,Chr17tp-3D7	26_45_12	34_31_23	0.03077	chr17:41349282-41752876	MAPT, KIAA1267, AC005829.1, AC005829.1, AC005829.1, AC005829.1, ARL17P1, LRRC37A
23	**Chr22tp-2A1**,Chr22tp-8C6	65_20	61_6	0.03138	chr22:41097245-41332085	AL022316.2, NFAM1, SERHL, CTA-126B4.5, SERHL2, Z93241.11, Z93241.11, POLDIP3

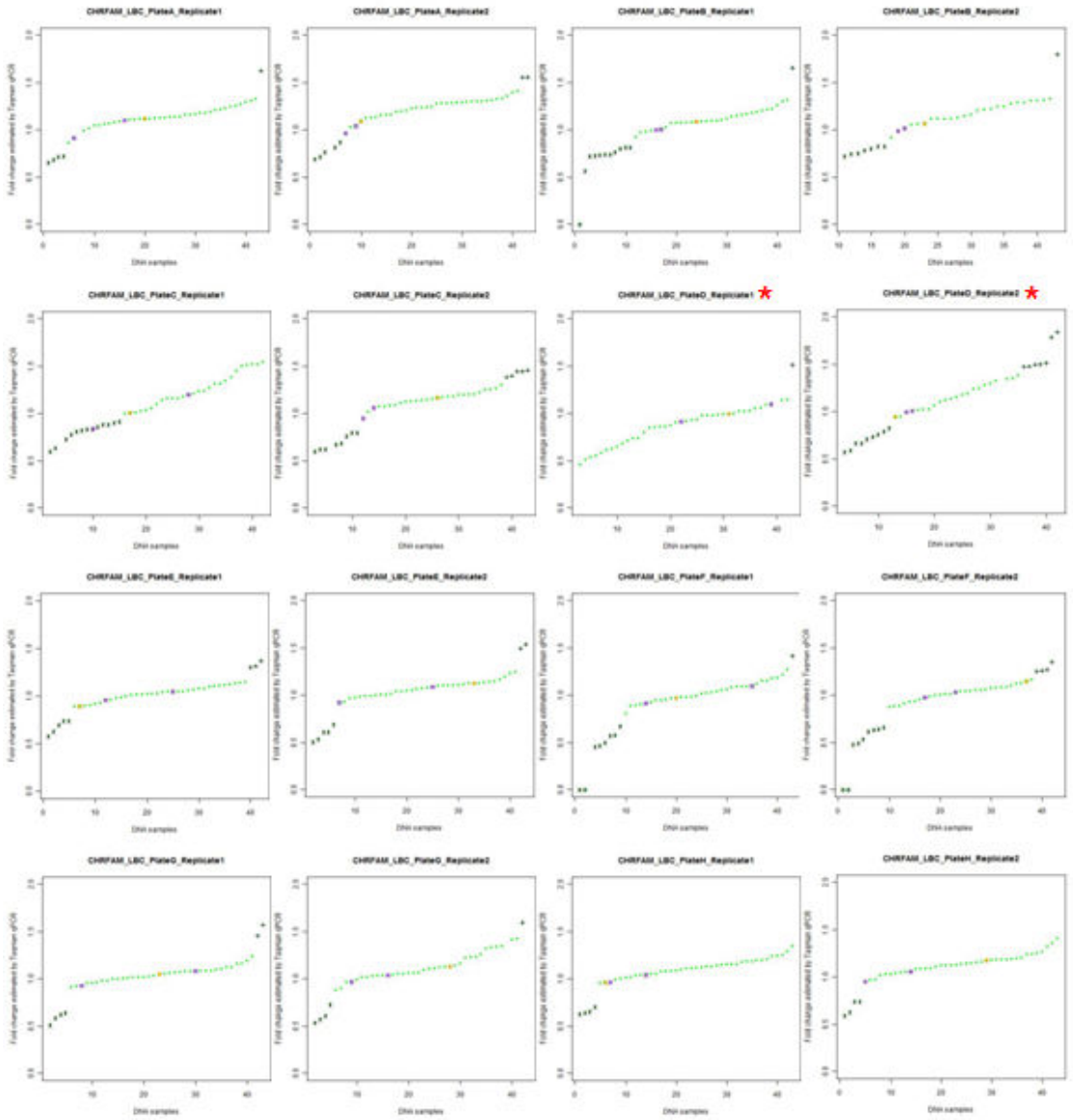
	clones	SCZ class	LBC class	p-val	chrom coordinates	Genes
24	Chr13tp-3D4,**Chr13tp-11B6**,Chr13tp-11A8	1_61_9	0_77_2	0.0317	chr13:113166605-113658050	DCUN1D2, AL442125.13, TMCO3, TFDP1, ATP4B, GRK1, GAS6, N/A, FAM70B
25	Chr15tp-2F10,Chr15tp-11F12,Chr15tp-1A1,**Chr15tp-4H6**,Chr15tp-12B1	4_1_14_4_35_19_1	1_2_10_2_23_34_6	0.03236	chr15:32330313-32775539	SLC12A6, AC021822.21, NOLA3, C15orf55, AGPAT7, GOLGA8A, AC025678.7, GOLGA8B, AC027139.5, AC100834.3
26	Chr12tp-2H2,**Chr12tp-2D2**,Chr12tp-6B6	51_12_0	28_18_1	0.03489	chr12:43845685-44127313	PLEKHA9, TMEM16F, AC063924.17
27	**Chr9tp-3C2**,Chr9tp-3E10,Chr9tp-2H8	54_16	71_7	0.03571	chr9:137679970-138276710	LCN9, SOHLH1, KCNT1, CAMSAP1, QSOX2, UBAC1, BTBD14A, N/A, LHX3, N/A
28	Chr20tp-6F1,Chr20tp-7A1,**Chr20tp-6F8**,Chr20tp-1E3,Chr20tp-2E8,Chr20tp-2F1	9_64_0_0	2_80_2_1	0.03897	chr20:55604184-59362789	ZBP1, TMEPAI, AL354984.17, C20orf85, ANKRD60, AL354776.15, PPP4R1L., RAB22A, VAPB, APCDD1L, N/A, STX16, NPEPL1, AL139349.36, hsa-mir-296, hsa-mir-298, GN/AS, GN/AS, TH1L, CTSZ, TUBB1, ATP5E, C20orf45, ZNF831, EDN3, C20orf83, PHACTR3, SYCP2, C20orf177, PPP1R3D, CDH26, C20orf197, N/A, hsa-mir-646, AL121918.26, AL117372.35, N/A, CDH4, N/A
29	Chr16tp-13A11,**Chr16-32k-3E20**,Chr16tp-11B1,Chr16tp-1C4	3_80	12_76	0.04086	chr16:28140185-28751472	SBK1, AC138904.3, EIF3CL, AC138894.2, AC138894.2, CLN3, AC138894.2, IL27, AC130449.1, CCDC101, SULT1A2, SULT1A1, AC145285.2, AC145285.2, EIF3S8, AC145285.2, ATXN2L
30	Chr19tp-5F6,**Chr19-32k-1L21**,Chr19tp-7A7,Chr19tp-2B5	2_3_61_7_1_1	0_6_68_2_2_8	0.04689	chr19:47734587-48590893	CEACAM8, N/A, PSG8, N/A, PSG6, N/A, PSG2, PSG4, PSG9, AC005392.2, N/A, CD177., AC005392.2, TEX101
31	**Chr16tp-13A11**,Chr16-32k-3E20,Chr16tp-11B1,Chr16tp-1C4	4_84	13_78	0.04916	chr16:28140185-28751472	SBK1, AC138904.3, EIF3CL, AC138894.2, AC138894.2, CLN3, AC138894.2, IL27, AC130449.1, CCDC101, SULT1A2, SULT1A1, AC145285.2, AC145285.2, EIF3S8, AC145285.2, ATXN2L

Table D-4 Results from qPCR Genotyping of a Candidate near *CHRFAM7A*

Analysis of all qPCR experiments to determine *CHRFAM7A* genotypes. In each 96-well plate 40 samples (+2 controls) were genotyped. Fold change of all samples in the same experiment (sorted in ascending order) were plotted on the same graph (green dots) and CNV genotypes were estimated by the change of the slope of the graph.

(o: Homozygous deletion samples; x: heterozygous deletion samples; +: duplication samples (: control sample NA10851; : control sample NA12776)




CHRFAM_LBC_PlateA_Replicate1
CHRFAM_LBC_PlateA_Replicate2
CHRFAM_LBC_PlateB_Replicate1
CHRFAM_LBC_PlateB_Replicate2
CHRFAM_LBC_PlateC_Replicate1
CHRFAM_LBC_PlateC_Replicate2
CHRFAM_LBC_PlateD_Replicate1
CHRFAM_LBC_PlateD_Replicate2
CHRFAM_LBC_PlateE_Replicate1
CHRFAM_LBC_PlateE_Replicate2
CHRFAM_LBC_PlateF_Replicate1
CHRFAM_LBC_PlateF_Replicate2
CHRFAM_LBC_PlateG_Replicate1
CHRFAM_LBC_PlateG_Replicate2
CHRFAM_LBC_PlateH_Replicate1
CHRFAM_LBC_PlateH_Replicate2

Appendix E

Validation of SCZ Specific CNVRs

E1 Validation by quantitative PCR

Ten CNV regions of various sizes and types (gains/losses, single/recurrent) were selected from the 113 schizophrenia cohort-specific calls for qPCR validation. Up to three sets of primers were designed tiling across each CNV region for qPCR. SYBR quantitative PCR was performed to quantify the fold change of the patient DNA against the reference DNA NA10851. Eight regions were validated (Table E-1).

Table E-1 10 regions selected for CNV validation by qPCR

CNV Identifier	Chromosome Coordinates	CNV Type	Genes	Single or Recurrent	qPCR Validated
3409-loss-23	chr12:17761730-18154145	Loss	RERGL	Single	√
850-loss-22	chr5:110212743-110577198	Loss	WDR36, TSLP	Single	√
5324-loss-21	chr6:22212251-22607763	Loss	PRL	Single	√
1085-loss-13	chr6:97469365-98475130	Loss	KLHL32, C6orf167	Single	√
ED1176-loss-17	chr7:17795286-18873121	Loss	SNX13, HDAC9	Single	√
4203-loss-28	chr7:93658283-94064936	Loss	SGCE, CASD1, COL1A2	Single	√
3789-loss-42	chr8:9670799-9988807	Loss	TNKS, MSRA	Single	
3766-gain-3	chr1:71346768-71893785	Gain	NEGR1	Single	√
3584-gain-11	chr3:139094479-139416525	Gain	ARMC8, DZIP1L, CLDN18, A4GNT, DBR1	Single	
323-gain-32, 4179-gain-19, 5758-gain-47	chr9:137112070-137464520	Gain	C9orf62, OLFM1	Recurrent	√

E2 validation by custom oligonucleotide CGH microarrays

We designed a high resolution oligo array with Roche Nimblegen Inc., with a median probe spacing of 40 bp, to validate the presence and delineate the breakpoints of 7 CNV regions in the schizophrenia cohort.

Among them were two CNVRs that failed to validate by qPCR. With oligo array data we confirmed the presence of the deletion on chromosome 8 at 9 Mb, and narrowed down the CNV to ~chr8 9.83-9.87 Mb (Figure E-2a). No obvious sign of CNV was detected near the duplication call on chromosome 13 at 139Mb, suggesting a false positive call ascertained by WGTP analysis (Figure E-2b).

Furthermore, we confirmed the presence and resolved the boundary of another 3 CNVRs at 86-87Mb on chromosome3 (Figure E-3a), at 61 Mb on chromosome5: (Figure E-3b) and at 137 Mb on chromosome9 (Figure E-3c). WGTP array data suggested that these regions overlap the candidate genes chromatin modifying protein 2B (*CHMP2B*), kinesin heavy chain member 2 (*KIF2A*) and olfactomedin 1 (*OLFM1*), respectively. In each of these 3 cases, oligo array CGH validated the CNV regions, but restricted CNV locations to intervals either upstream or downstream of the mentioned genes.

The oligo array also indicates that two recurrent CNVRs detected on WGTP were likely false positive events. The former is a CNV in sample 5390 at chr14:103Mb overlapping the gene kinesin family member 26A (*KIF26A*), and the latter is in sample 4179 at chr20:44Mb spanning cadherin 22 precursor (*CDH22*) (data not shown). We cannot confirm whether the other affected DNA samples with CNV calls on WGTP carried any change at these loci.

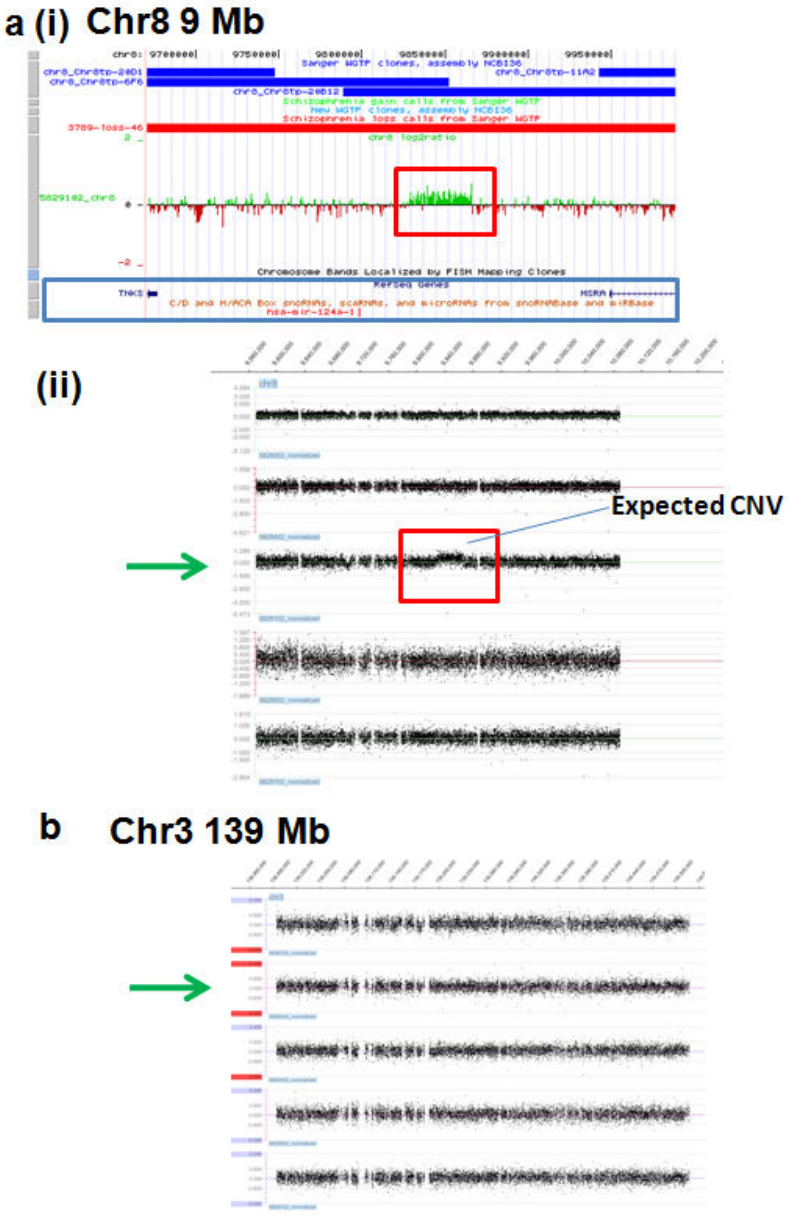
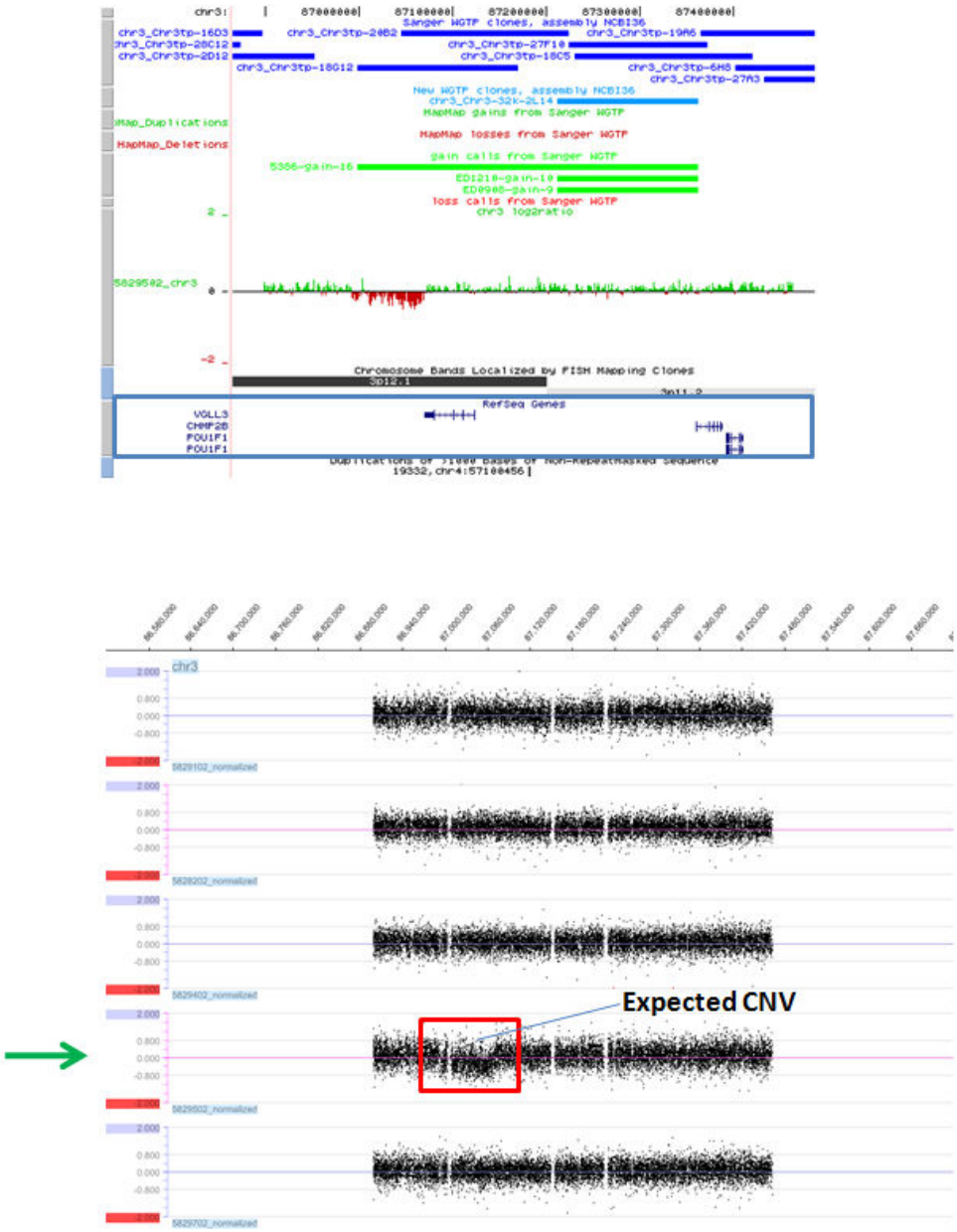
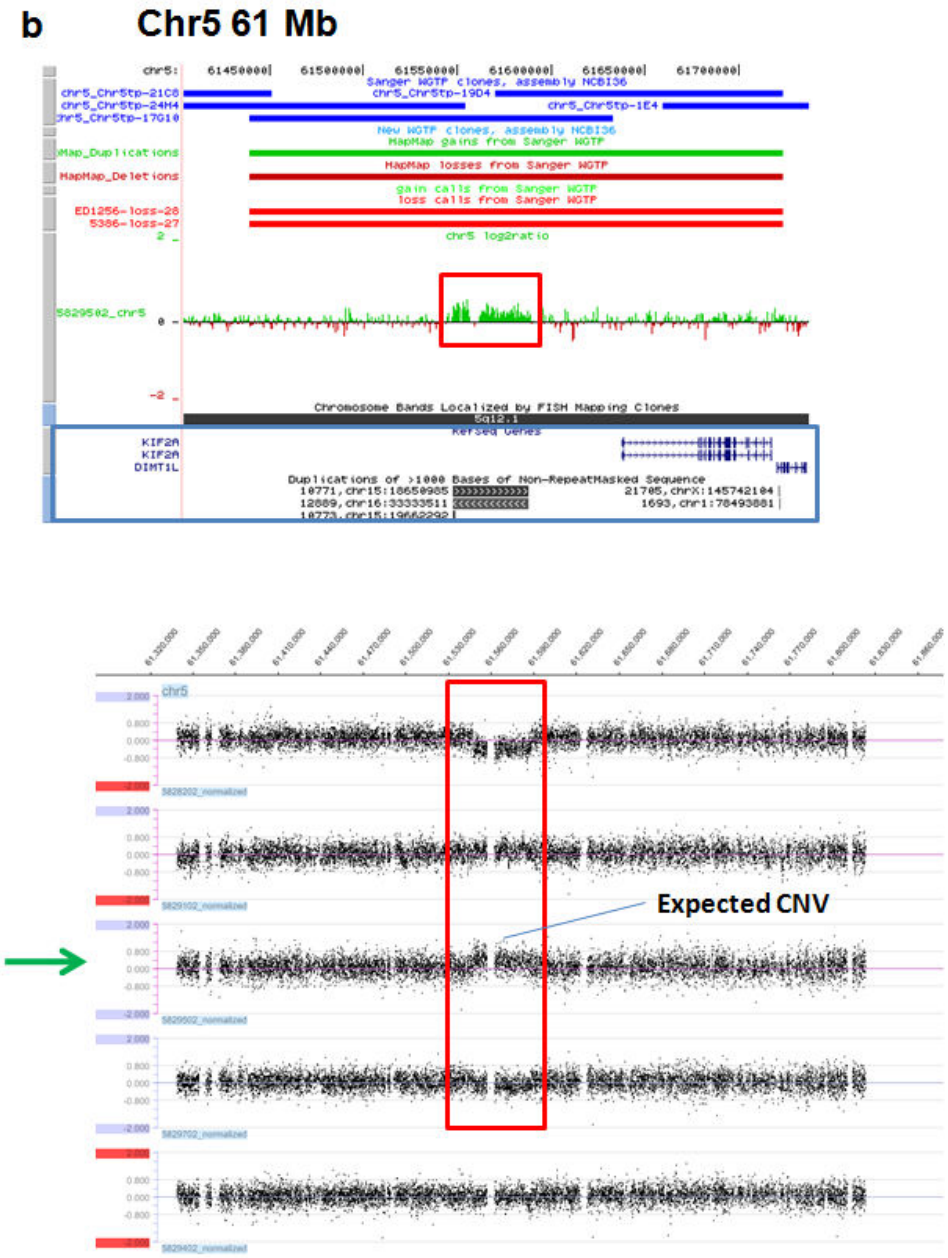


Figure E-2 Oligonucleotide array CGH validation of 2 CNV regions that previously failed to validate with qPCR. a) (i) The deletion at chr8:9Mb detected in sample 5398 was confirmed by oligo array CGH, with the CNV boundaries resolved. (**ii)** Results from 5 hybridization experiments on the oligo array revealed the expected deletion in sample 5398 (green arrow, hybridized as reference). **b)** The duplication call at chr3:139Mb showed no sign of CNV on the five pairs of DNA hybridized. *(red filled block: CNV region as determined by WGTP array; red unfilled box: the expected CNV determined by oligo array; green arrows: pair of DNA samples we expected to see the CNV)*

a **Chr3 86-87 Mb**



(Figure E-3 to be continued)



(Figure E-3 to be continued)

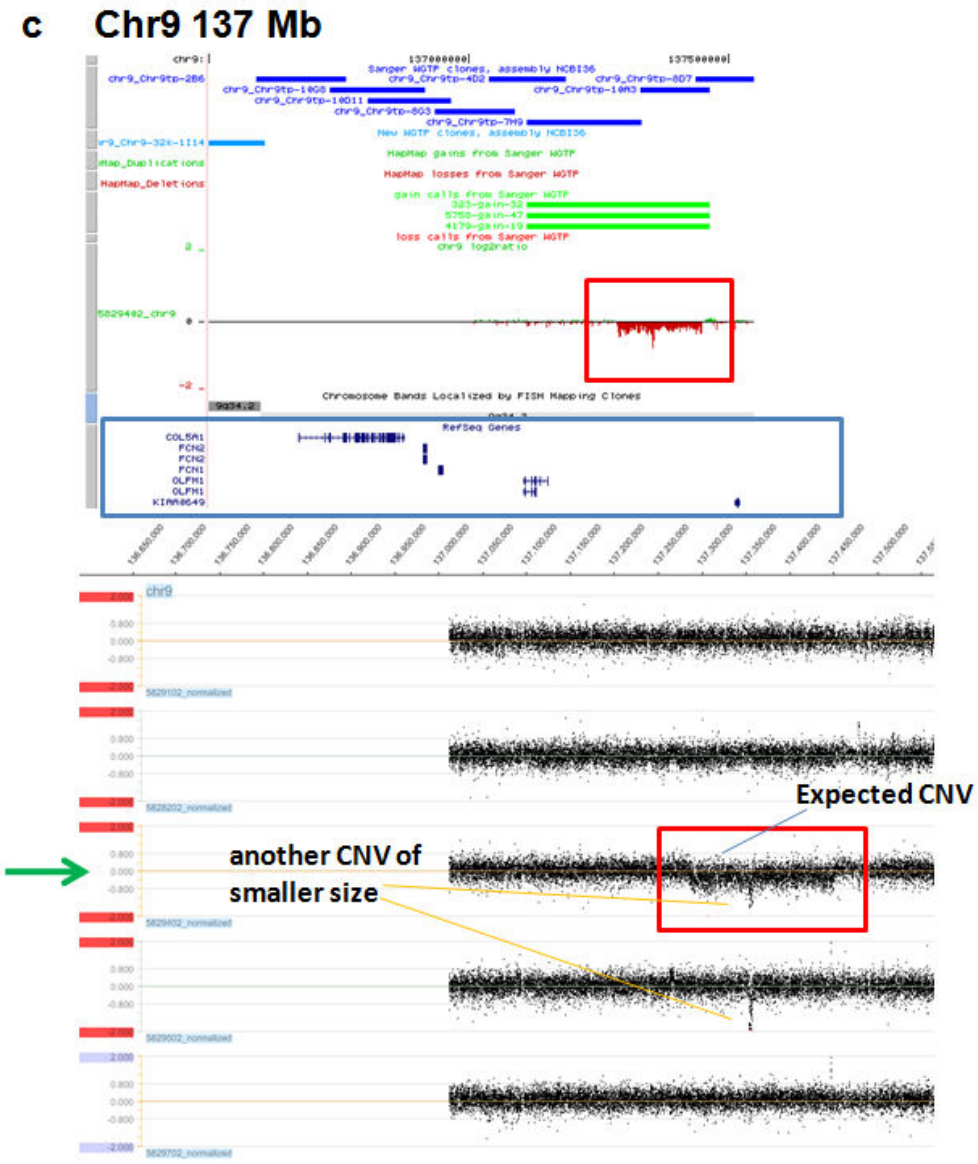


Figure E-3 Three CNVs confirmed by oligo array CGH.
a) A duplication (sample 5386, hybridized as reference) at chr3 86-87 Mb was validated. CNV boundaries were refined to ~chr3 87.00-87.06 Mb **b)** A deletion (sample 5386, hybridized as reference) at chr5 61 Mb was validated. CNV boundaries were refined to ~chr5 61.55-61.58 Mb **c)** A duplication (sample 4179, hybridized as reference) at chr9 137 Mb was validated.CNV boundaries were refined to ~chr9 137.28-137.45 Mb
(*red or green solid box: CNV region as determined by WGTP array; red unfilled box: expected CNV determined by oligo array; green arrows: pair of DNA samples expected to display the CNV)*

Appendix F

186 Genes in the NMDA Receptor Complex (NRC/MASC)

Gene	ID	chr	start	end
GNB1	ENSG00000078369	1	1706590	1812355
CAPZB	ENSG00000077549	1	19537857	19684594
PRDX1	ENSG00000117450	1	45749295	45760196
PRKACB	ENSG00000142875	1	84316329	84476765
ATP1A1	ENSG00000163399	1	116717359	116748917
ARNT	ENSG00000143437	1	149048810	149115815
PKLR	ENSG00000143627	1	153526254	153537843
PRDX6	ENSG00000117592	1	171713084	171724567
GLUL	ENSG00000135821	1	180618292	180627573
PLA2G4A	ENSG00000116711	1	185064655	185224726
ATP2B4	ENSG00000058668	1	201862312	201984201
ACTN2	ENSG00000077522	1	234916422	234993863
PRKCE	ENSG00000171132	2	45732547	46268632
CALM1	ENSG00000143933	2	47240809	47257159
SPTBN1	ENSG00000115306	2	54536958	54752086
YWHAE	ENSG00000181399	2	138762069	138762767
STK39	ENSG00000198648	2	168518777	168812365
SLC25A12	ENSG00000115840	2	172349127	172458991
MYO1B	ENSG00000128641	2	191818375	191998360
MAP2	ENSG00000078018	2	209997016	210307087
ARPC2	ENSG00000163466	2	218790119	218827324
TUBA1A	ENSG00000127824	2	219822677	219827514
IRS1	ENSG00000169047	2	227308182	227372719
ATG16L1	ENSG00000085978	2	233825058	233869059
ARPC4	none	3	9809227	9823789
RAF1	ENSG00000132155	3	12600108	12680678
CTNNB1	ENSG00000168036	3	41216000	41256938
BSN	ENSG00000164061	3	49566926	49683974
APPL1	ENSG00000157500	3	57236805	57282539
SLMAP	ENSG00000163681	3	57716997	57889935
ATP6V1A	ENSG00000114573	3	114948598	115013591

Gene	ID	chr	start	end
GAP43	ENSG00000172020	3	116824861	116923024
GSK3B	ENSG00000082701	3	121028238	121295203
KALRN	ENSG00000160145	3	125296275	125922726
PIK3CA	ENSG00000121879	3	180349005	180435189
GNB4	ENSG00000114450	3	180599696	180652065
DLG1	ENSG00000075711	3	198255819	198509844
MAPK10	ENSG00000109339	4	87156655	87593320
PPP3CA	ENSG00000138814	4	102163610	102487376
SLC25A4	ENSG00000151729	4	186301392	186305418
HOMER1	ENSG00000152413	5	78704215	78845796
LMNB1	ENSG00000113368	5	126140732	126200608
PPP2CA	ENSG00000113575	5	133560047	133589849
SPINK5	ENSG00000133710	5	147423728	147497045
CAMK2A	ENSG00000070808	5	149579247	149650047
DBN1	ENSG00000113758	5	176816220	176833271
GNB2L1	ENSG00000204628	5	180596531	180603508
MOG	ENSG00000206456	6	29732788	29748128
HSPA1B	ENSG00000204388	6	31903667	31906010
EHMT2	ENSG00000204371	6	31955515	31973443
SYNGAP1	ENSG00000197283	6	33495919	33533296
VEGFA	ENSG00000112715	6	43845926	43862202
LCA5	ENSG00000135338	6	80251427	80565440
GRIK2	ENSG00000164418	6	101953385	102624651
GRM1	ENSG00000152822	6	146390611	146800427
RAC1	ENSG00000136238	7	6380651	6410120
RALA	ENSG00000006451	7	39629687	39714240
CAMK2B	ENSG00000058404	7	44223280	44331749
STX1A	ENSG00000106089	7	72751472	72771925
YWHAG	ENSG00000170027	7	75794053	75826252
ADAM22	ENSG00000008277	7	87401394	87670140
AKAP9	ENSG00000127914	7	91408128	91577925

Gene	ID	chr	start	end
GNB2	ENSG00000172354	7	100109311	100114733
PRKAR2B	ENSG00000005249	7	106472375	106589491
CAPZA2	ENSG00000198898	7	116289799	116346548
FLNC	ENSG00000128591	7	128257719	128286572
NEFM	ENSG00000104722	8	24827188	24832508
NEFL	ENSG00000104725	8	24864385	24870541
DPYSL2	ENSG00000092964	8	26491327	26571607
PTK2B	ENSG00000120899	8	27224916	27372820
DUSP4	ENSG00000120875	8	29249539	29264104
RAB2A	ENSG00000104388	8	61592113	61696183
RPL7	ENSG00000147604	8	74365073	74375857
FABP5	ENSG00000164687	8	82355325	82359567
YWHAZ	ENSG00000164924	8	101999980	102034745
ARC	ENSG00000198576	8	143689412	143692835
GSN	ENSG00000148180	9	123009896	123134942
STXBP1	ENSG00000136854	9	129414389	129494816
DNM1	ENSG00000106976	9	130005484	130057348
GRIN1	ENSG00000176884	9	139152663	139183028
ATP5C1	ENSG00000165629	10	7870099	7889761
C10orf38	ENSG00000148468	10	15293661	15453103
VDAC2	ENSG00000165637	10	76640569	76661208
LGI1	ENSG00000108231	10	95507632	95547904
INA	ENSG00000148798	10	105026910	105040091
ABLIM1	ENSG00000099204	10	116180862	116434404
HRAS	ENSG00000174775	11	522242	525591
SLC25A22	ENSG00000177542	11	780478	786221
LDHA	ENSG00000134333	11	18372668	18386341
PTPN5	ENSG00000110786	11	18706053	18770844
SLC1A2	ENSG00000110436	11	35229329	35398186
MAPK8IP1	ENSG00000121653	11	45863778	45884592
BAD	ENSG00000002330	11	63793878	63808740

Gene	ID	chr	start	end
KLC2	ENSG00000174996	11	65781766	65791907
ACTN3	ENSG00000204633	11	66070967	66087373
CTTN	ENSG00000085733	11	69922292	69960337
SHANK2	ENSG00000162105	11	69991609	70536020
RAB6A	ENSG00000175582	11	73064342	73149849
DLG2	ENSG00000150672	11	82843701	85015962
GRM5	ENSG00000168959	11	87881006	88438761
DLAT	ENSG00000150768	11	111401374	111440324
GAPDH	none	12	6513918	6517797
TPI1	ENSG00000111669	12	6846544	6850373
PHB2	none	12	6944778	6950152
GRIN2B	ENSG00000150086	12	13605411	14024319
LDHB	ENSG00000111716	12	21679543	21702043
FGD4	ENSG00000139132	12	32546361	32684940
ARF3	ENSG00000134287	12	47616259	47637577
MYL6	ENSG00000092841	12	54838367	54841633
SYT1	ENSG00000067715	12	77781904	78369917
LIN7A	ENSG00000111052	12	79715306	79855825
ARPC3	ENSG00000111229	12	109357090	109376185
PPP1CC	ENSG00000186298	12	109642127	109665050
PTPN11	ENSG00000179295	12	111340919	111432099
NOS1	ENSG00000089250	12	116120497	116283965
CIT	ENSG00000122966	12	118607981	118799475
RAN	ENSG00000132341	12	129922521	129927316
RAP2A	ENSG00000125249	13	96884475	96918243
MYH6	ENSG00000092054	14	22946378	22974710
RTN1	ENSG00000139970	14	59132447	59407101
AKAP5	ENSG00000179841	14	64004256	64016821
TJP1	ENSG00000104067	15	27778863	27901998
TP53BP1	ENSG00000067369	15	41486699	41590028
MYO5A	ENSG00000197535	15	50386770	50608539