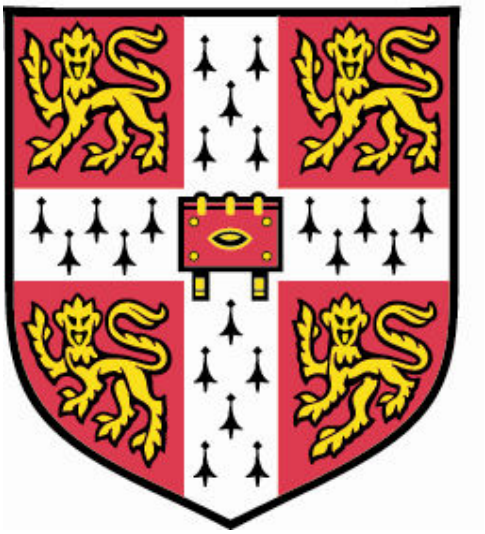

Copy Number Variation and Schizophrenia

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

List of Tables

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