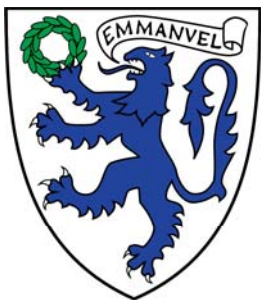

# The contribution of rare variants to risk of schizophrenia and neurodevelopmental disorders




**Tarjinder Singh**

Wellcome Trust Sanger Institute
University of Cambridge

This dissertation is submitted for the degree of
*Doctor of Philosophy*

January 2017

# Declaration

I hereby declare that I carried out the work described in this Thesis between September 2012 and August 2016 under the supervision of Dr. Jeffrey C. Barrett at the Wellcome Trust Sanger Institute. The contents of this Thesis has not been submitted in whole or in part for any other degree or qualification at the University of Cambridge, or any other University. This Thesis does not exceed the specified length limit, and is formatted according to the requirements set by the Biology Degree Committee and the Board of Graduate Studies.

Tarjinder Singh
January 2017

# Acknowledgements

I would first like to thank my supervisor Jeff for giving me the opportunity to write my thesis in his group and for his support and encouragement in the past four years. I arrived in Cambridge in 2012 with plans to complete a one-year MPhil degree in the genetics of blood traits before attending medical school the following year. Little did I know that I would be fortunate enough to stay at the Sanger and explore an area of genetics and medicine that I had never previously encountered. Now, four years later, I am more motivated than ever to delve deeper into the world of statistical genetics and use its advances to understand the fundamental causes of mental illnesses. This formative experience would not be possible if not for the patience, mentorship, and guidance that Jeff has shown me. I sincerely hope we keep in touch, and perhaps find an opportunity to work together again in the future.

The research in this Thesis requires the coordinated efforts of numerous collaborators in the UK and around the world. I thank the many clinicians and scientists in the UK10K consortium who designed the initial study that laid the foundation for this Thesis. In particular, I would like to thank Mike Owen, Mick O'Donovan, Dave Curtis, and Matthew Hurles for productive discussions, advice, and contributions to this work. I want to express my gratitude to the Wellcome Trust and Williams College for their generous financial support. Most importantly, I want to thank the tens of thousands of patients and participants who enrolled in the studies described in this Thesis, without whom none of this work is possible and for which I am indescribably grateful.

On a more personal note, I want to extend my gratitude to the members of the Barrett team, past and present, for engaging team meetings, memorable retreats, and entertaining lunch discussions. I also want to thank my friends at Emmanuel College for all the enjoyable Formal dinners, pub crawls, and European travels that have made life in Cambridge so entertaining. In particular, I want to thank Albert and Uttara for their unwavering friendship, support, and honesty. I am grateful to the PhD students at the Sanger Institute, especially the Class of 2012, for all the enjoyable times we've had, from the scenic drives through the Peak District to the adventures in Warsaw. Lastly, I want to thank my parents and sister for their patience, encouragement, love and support throughout the years.

# Abstract

In recent years, whole-exome sequencing has successfully identified genes in which rare variants confer substantial risk for neurodevelopmental disorders, such as autism spectrum disorders and intellectual disability. In many of these studies, the same gene is implicated in a wide variety of diagnoses and presentations. Despite a number of rare variant studies in schizophrenia, no gene has been significantly implicated using rare coding variants. In this Thesis, I compiled the largest rare variant data set in schizophrenia to date, and meta-analysed the whole-exome sequences of 1,077 trios, 4,268 cases, and 9,343 matched controls. With these data, I identified a genome-wide significant association between rare loss-of-function (LoF) variants in *SETD1A* and risk for schizophrenia. I additionally found that *SETD1A* is substantially depleted of LoF variants in the general population, and that LoF variants in this gene increased risk for a range of neurodevelopmental disorders. Combined, our results implicate epigenetic regulation, specifically histone modification, as a mechanism in the pathogenesis of schizophrenia, and suggest that rare risk alleles may potentially be shared between schizophrenia and other neurodevelopmental disorders.

To better understand if *SETD1A* finding can be generalized to a larger number of rare schizophrenia risk variants, I jointly analysed the trio and case-control exome data with array-based copy number variant calls from 6,882 cases and 11,255 controls. I found that individuals with schizophrenia carried a significantly higher burden of rare damaging variants in 3,488 "highly constrained" genes with a near-complete depletion of truncating variants. Rare variant enrichment analyses demonstrated that the rare schizophrenia risk variants were most strongly enriched in autism risk genes, and genes diagnostic of severe developmental disorders. I further showed that schizophrenia patients with intellectual disability had a greater enrichment of rare damaging variants in highly constrained genes, but that a weaker but significant enrichment existed throughout the larger schizophrenia population. Combined, these results demonstrate that schizophrenia risk loci of large effect across a range of variant types implicate a common set of genes shared with broader neurodevelopmental disorders, suggesting a path forward in identifying additional risk genes in psychiatric disorders and further supporting a neurodevelopmental etiology to the pathogenesis of schizophrenia.

# Table of contents

# List of figures

# List of tables