Identifying and modelling genes that are associated with rare developmental disorders

Keren Jacqueline Carss
Wellcome Trust Sanger Institute
Queens' College
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy
August 2014

Declaration

I declare that this dissertation describes my own, original work. It only includes work done in collaboration where clearly and specifically indicated in the text. No part of this dissertation has been submitted for any other degree, diploma, or qualification at any university or institution. This dissertation does not exceed 60,000 words.

Keren Jacqueline Carss

August 2014

Abstract

While individually rare, collectively developmental disorders are common, affecting around 3% of live births in the UK. The aetiology of these disorders often includes a genetic component. Next generation sequencing provides a powerful tool with which to identify variants that cause rare developmental disorders. This dissertation describes three distinct projects in which next generation sequencing was used for this purpose, along with statistical or functional follow-up approaches.

A cohort of 30 fetuses with a diverse range of structural abnormalities, along with their parents, was exome sequenced. I analysed these data to identify rare, high quality, coding variants consistent with a *de novo* or recessive inheritance model. I investigated several methods of variant interpretation, including manual and computational methods, and found causative variants for 10% of the cohort. These results suggest that next generation sequencing is a promising method for prenatal genetic diagnostics.

As part of the UK10K project, 996 patients with moderate to severe intellectual disability (ID) underwent sequencing of 565 known or candidate ID-associated genes. I developed and implemented a pipeline to identify likely causative loss of function (LOF) variants through extensive quality filtering. From these data, causative variants were identified for ~14% of the cohort, and the novel ID-associated gene *SETD5* was identified. Next, I performed a series of case-control enrichment analyses to evaluate the contribution of different classes of possibly pathogenic variants. Patients with ID had a significant enrichment of both LOF and missense variants in known ID-associated genes, compared to controls with non-syndromic congenital heart defects.

One strategy to investigate the consequences of a potentially pathogenic variant is to inhibit expression of the gene in an appropriate animal model, and assess the extent to which aspects of the human phenotype are recapitulated. I applied this technique to two genes identified from the UK10K project as likely to be associated with dystroglycanopathy, a subtype of muscular dystrophy. I inhibited the expression of both genes, *B3GALNT2* and *GMPPB*, in zebrafish embryos using morpholino oligonucleotides. The phenotype of both models mimicked several aspects of the human phenotype including morphological defects such as micropthalmia and hydrocephalus, structural defects of the tissue such as disordered muscle fibres, and the precise molecular defect, which is hypoglycosylation of α-dystroglycan.

Acknowledgements

Throughout my PhD, I have been involved in several large collaborative projects, so there are a great many people who I would like to thank. My first and biggest thank you is to my primary supervisor, Dr Matthew Hurles. He provided me with the opportunity to work on several exciting and fruitful projects. He also provided the perfect balance of support and freedom, along with a constant stream of outstanding ideas, and encouragement when I needed it. Thank you also to my secondary supervisor, Dr Derek Stemple. He introduced me to the zebrafish, and gave me valuable support throughout my time working in the zebrafish laboratory. Thank you to the additional members of my thesis committee: Dr Helen Firth and Dr David Adams, for guidance and advice throughout the last four years.

Thank you to the following clinical collaborators at the University of Birmingham: Prof Eamonn Maher, Prof Mark Kilby, Dr Dominic McMullan, and Dr Sarah Hillman. I very much enjoyed working on the abnormal fetal development project, and I learned a lot from it. Thank you to Dr Vijaya Parthiban, Dr Alejandro Sifrim and Dr Damian Smedley for running the CoNVex, eXtasy, and PhenoDigm programs respectively during this project.

I am grateful to the UK10K consortium, particularly the rare disease group, for the opportunity to be involved in two exciting UK10K projects. Thank you to Dr Lucy Raymond for the opportunity to work on the UK10K intellectual disability cohort, in an exciting and fruitful collaboration. I also thank the many other people involved in this project, most importantly Dr Detelina Grozeva, Dr Olivera Spasic-Boskovic, and Dr James Floyd.

Thank you to Prof Francesco Muntoni for giving me the opportunity to work on the UK10K dystroglycanopathy project. Also to other members of this project including Dr Elizabeth Stevens, Dr Silvia Torelli, Dr Sebahattin Cirak, and Dr Reghan Foley. I owe a particularly large thank you to Dr Yung-Yao Lin, who provided a huge amount of support, supervision, and training to me during this project. Thank you also to Dr Sebastian Gerety for help and support with designing zebrafish experiments.

Thank you to all patients who participate in research studies, without whose generosity none of this work would have been possible.

I am very grateful to the Wellcome Trust for generously funding my PhD. I have been fortunate to do my PhD at the Wellcome Trust Sanger Institute, which provides not only

a very stimulating environment, but also unparalleled facilities. I therefore thank the many people who work in the various pipelines and services teams at the Wellcome Trust Sanger Institute, including the sample reception and sequencing teams, the GAPI team, Human Genome Informatics, and the many other people who keep the laboratory and computational facilities running. I am grateful to Dr Alex Bateman, Dr Julian Rayner, Dr Annabel Smith, Christina Hedberg-Delouka, and Carol Dunbar for practical and administrative support.

I thank all past and present members of teams 29 and 31 for help with experiments or coding, and for many fun and enlightening conversations over the years. I would like to especially thank Dr Saeed Al Turki, Dan King, and Dr Vijaya Parthiban for their great generosity in sharing their scripts, skills, and time with me. I am also particularly grateful to Dr Ana Cvejic and Dr Jovana Serbanovic-Canic for teaching me basic experimental techniques in zebrafish when I first started.

On a more personal note, I thank the many additional scientists who have encouraged, mentored, and inspired me throughout my career, both directly and by example. These include Prof Graham Anderson, Dr Kevin O'Shaughnessy, Dr Annie Mercier Zuber, Dr Darren Logan, Dr Inês Barrosso, Dr Eleanor Wheeler, Dr Eleftheria Zeggini, Prof Jane Worthington, and Prof Eamonn Maher. Last but not least I thank my friends and family, especially the fabulous PhD10, the wonderful Norman and Carss families, and my endlessly supportive husband.

Publications

Publications arising from work associated with this thesis:

- Mackie FL, **Carss KJ**, Hillman SC, Hurles ME, & Kilby MD. Exome sequencing in fetuses with structural malformations. [Review article] ***Journal of Clinical Medicine.*** 2014 July, 3(3), 747-762.

- Grozeva D, **Carss KJ**, Spasic-Broskovic O, Parker M, Archer H, Firth HV, *et al. De novo* mutations in *SETD5* cause intellectual disability and associated features of 3p25 microdeletion syndrome. ***American Journal of Human Genetics****.* 2014 April, 3;94(4):618-24.

- **Carss KJ**, Hillman SC, Parthiban V, McMullan DJ, Maher ER, Kilby MD & Hurles ME. Exome sequencing improves genetic diagnosis of structural fetal abnormalities revealed by ultrasound. ***Human Molecular Genetics.*** 2014 June, 15,23(12):3269-77.

- **Carss KJ***, Stevens E*, Foley AR, Cirak S, Riemersma M, Torelli S, *et al*. Mutations in *GDP-Mannose pyrophosphorylase B* cause congenital and limb-girdle muscular dystrophies associated with hypoglycosylation of α-dystroglycan. ***American Journal of Human Genetics.*** 2013 July, 11;93(1):29-41.

- Stevens E*, **Carss KJ***, Cirak S, Foley AR, Torelli S, Willer T, *et al.* Mutations in *B3GALNT2* cause congenital muscular dystrophy and hypoglycosylation of α-dystroglycan. ***American Journal of Human Genetics.*** 2013 March, 7;92(3):354-65.

* Jointly contributing authors.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

aCGH	Array comparative genomic hybridisation
α-DG	α-dystroglycan
ASD	Autism spectrum disorder
β-DG	β-dystroglycan
BioGPS	Biology Gene Portal System
bp	Base pair
CAST	Cohort allelic sums test
CDG	Congenital disorder of glycosylation
cfDNA	Cell-free DNA
CHD	Congenital heart disease
CK	Creatine kinase
CMD	Congenital muscular dystrophy
CNS	Central nervous system
CNV	Copy number variant
CRISPR	Clustered regularly interspaced short palindromic repeat
DDD	Deciphering developmental disorders
DDG2P	Developmental Disorder Gene2Phenotype
DGC	Dystrophin-glycoprotein complex
DNA	Deoxyribonucleic Acid
Dol-P-Man	Dolichol phosphate mannose
EBD	Evans blue dye
ECM	Extracellular matrix
ENU	*N*-ethyl-*N*-nitrosourea
ER	Endoplasmic reticulum
ESP	Exome Sequencing Project
FCMD	Fukuyama-type CMD
FISH	Fluorescence in situ hybridisation
FORGE	Finding of rare disease genes
GalNAc	N-acetylgalactosamine
Gb	Gigabases
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GlcNAc	N-acetylglucosamine
GPI	Glycosylphosphatidylinositol
GWAS	Genome-wide association study
HDL-C	High-density lipoprotein cholesterol
HMQ	High mapping quality
Hpf	Hours post fertilisation
HPO	Human phenotype ontology
IC	Information Content
ID	Intellectual disability

IEM	Inborn error of metabolism
IGV	Integrative Genomics Viewer
IKMC	International knockout mouse consortium
Indel	Insertion deletion
IQ	Intelligence quotient
Kb	Kilobase
LD	Linkage disequilibrium
LGMD	Limb girdle muscular dystrophy
LOF	Loss of function
Mb	Megabase
MEB	Muscle-eye-brain disease
MO	Morpholino oligonucleotide
MPO	Mammalian phenotype ontology
MRI	Magnetic resonance imaging
NGS	Next generation sequencing
NHGRI	National Human Genome Research Institute
NHLBI EVS	National Institute of Health: Heart, Lung, and Blood Institute exome variant server
NIPT	Non-invasive prenatal testing
OMIM	Online Mendelian Inheritance in Man
PAGE	Prenatal Assessment of Genomes and Exomes
PCA	Principal component analysis
PCR	Polymerase chain reaction
PKU	Phenylketonuria
PSD	Postsynaptic density
PTR	Primary target region
QF-PCR	Quantitative fluorescent polymerase chain reaction
RD	Retinal dystrophy
RNA	Ribonucleic Acid
RT-PCR	Reverse transcription polymerase chain reaction
SB	Splice blocking
SimJ	Jaccard Index
SKAT	Sequence kernel association test
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
TALEN	Transcription activator-like effector nuclease
TB	Translation blocking
UTR	Untranslated region
VCF	Variant call format
VEP	Variant effect predictor
VOUS	Variant of unknown significance
WTSI	Wellcome Trust Sanger Institute
WWS	Walker-Warburg Syndrome
ZFIN	Zebrafish information network
ZFN	Zinc finger nuclease