Genetic and Phenotypic Investigations into Developmental Disorders

Dr Wendy Dawn Jones
Wellcome Trust Sanger Institute
Newnham College,
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy
July, 2017

Abstract

Genetic and Phenotypic Investigations into Developmental Disorders, Dr Wendy Dawn Jones

Genetic developmental disorders cause distress to families and substantial mortality, morbidity and costs to the health service. However not all genetic diseases have been discovered or had their genetic cause elucidated, and the phenotypic spectrum of many molecularly solved disorders is not fully understood.

Wiedemann-Steiner syndrome (WSS), resulting from mutations in *KMT2A*, is a multiple congenital-anomaly syndrome associated with hypertrichosis, intellectual disability and a distinctive facial appearance. In order to understand the broader spectrum of WSS I identified 84 individuals with WSS and mutations in *KMT2A* and performed a detailed phenotypic evaluation. My cohort is 15 times larger than the biggest cohort reported so far. I identified new phenotypic features, and defined the mutational spectrum and growth profile associated with WSS and mutations in *KMT2A*. In addition, I ran a clinician facial recognition experiment that confirmed WSS is distinguishable from other developmental disorders. To investigate the genetic architecture of hypertrichosis more generally, I assembled a cohort of 228 individuals with hypertrichosis. I showed by analysing their exome variant profiles that there is a burden of mutations in genes that play a role in maintaining the structure and function of chromatin in this group compared to other individuals with developmental disorders. I showed, in principle, grouping by hypertrichosis is a successful method for gene discovery.

Finally, I investigated autosomal recessive disease in 1080 individuals with developmental disorders in the DDD study, for which I generated a population matched control dataset using the parental untransmitted alleles. My work gives the first insight into the contribution of autosomal recessive disease to developmental disorders, by studying untransmitted haplotypes. The themes of this thesis include those important in current Clinical Genetics practice in the whole exome sequencing era: loss of function versus missense variants, the use of next generation sequencing to unravel the underlying causes of developmental disorders and the challenges of assigning pathogenicity to variants.

Declarations

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed 60,000 words.

Dr Wendy Dawn Jones, July 2017

Acknowledgements

First and foremost, I would like to thank my supervisors Jeff Barrett and Matt Hurles. In particular, for our punchy and exciting conversations about science, their highly constructive feedback and for inspiring me by their own achievements and ideas. Special thanks also to David Fitzpatrick for his advice with my study of individuals with Wiedemann-Steiner syndrome.

I also wish to thank all those colleagues and collaborators without whom this work would not have been possible, especially, Michael Simpson, Meriel McEntagart, Charu Despande and Hans Bjornsson. Special thanks also, to all the clinicians who recruited individuals to the DDD study and to the WiSH study and welcomed me into their clinics to meet individuals with Wiedemann-Steiner syndrome. I also thank the Hurles and Barrett Groups and the DDD study team for all their help and stimulating conversations, both about science and life. In particular, to Margriet, Art and TJ. I remain forever indebted to those who helped me in my journey to become a programmer, in particular, Tomas, Dan and TJ.

On a personal note, thanks to all my friends, old and new who helped, often without realising, with their love and support and sense of fun. To name but a few, thanks to Sharmeen, Kate, Angela, Elora, Alex, Laura and Lucy. To Cor and Jeannette for their support and understanding. To my grandfathers who taught me to aim high in life and my grandmothers who taught me about hard work and perseverance. To my wonderful parents for all their support and encouragement throughout my life. To David for making me laugh. To Rogier for everything, especially his endless support and understanding.

Finally, and most importantly, I would like to thank the individuals affected with developmental disorders and their families for kindly taking part in this research, without whom, this work would have not been possible.

Attributions

Below is a summary of the contributions of other scientists and clinicians to the work described in this dissertation. I carried out all of this work under the supervision of Dr Jeff Barrett and Dr Matt Hurles.

Chapter 2:

Patient recruitment to the Genotype-phenotype study was by Clinical Geneticists in the UK, Europe and worldwide. Local Clinicians filled out questionnaires and provided phenotype information. Dr Roman Laskowski performed the protein modelling experiments and created the figures associated with these and the figure legends. I was responsible for the execution of this investigation from start to finish. I managed the flow of patient data for all 84 patients and liaised with clinicians to obtain phenotype data. I was responsible for the clinician facial recognition survey. I analysed all the phenotypic and molecular data and performed the statistical analysis and was responsible for the clinical interpretation of findings.

Chapter 3:

Whole exome sequencing was carried out by the Wellcome Trust Sanger Institute (WTSI) Core facility. SNV and INDEL detection was carried out by Mr Martin Pollard. I received help with variant annotation from Stephen Clayton (VEP), Mr Martin Pollard and Mr TJ Singh. The clinical filtering programme was written by Dr Jeremy McRae based on code originally written by Dr Saeed Al Turki and Dr Jeff Barrett. Mr TJ Singh provided help with performing statistical analysis on missense variants and python programming. The programme used to calculate the burden of variants in chromatin genes was written by Dr Jeremy McRae. I was responsible for the execution of this investigation from start to finish. I designed the recruitment criteria and liaised with UK clinical genetics clinicians to identify individuals with relevant phenotypes within the DDD and recruit 20 trios from overseas. I analysed the exome sequencing data from the variant call format stage, writing my own QC scripts and scripts to analyse the exome data to identify rare coding variants. I was responsible for reviewing all of the variants in confirmed disease genes and candidate genes, performing statistical analysis to analyse the significance of missense variants and performing the burden analysis for variants in chromatin genes.

Chapter 4:

Some parts of this investigation have been published (1, 2). Patient recruitment into the DDD study was carried out by Clinical Geneticists and Genetic Counsellors in the UK and Ireland. The DDG2P was devised by Professor David Fitzpatrick and is updated by Professor David Fitzpatrick and Dr Helen Firth. The Decipher team supported and ran the portal in Decipher into which clinical information was uploaded. Whole exome sequencing was carried out by the Wellcome Trust Sanger Institute (WTSI) Core facility. SNV and INDEL detection was carried out by the GAPI pipeline team at the WTSI. *De novo* mutation detection was carried out by the DDD Study Bioinformatics team. The DDD management team comprises Dr Matt Hurles, Dr Jeffrey Barrett, Dr Caroline Wright and Dr Helen Firth and Professor David Fitzpatrick, Consultants in Clinical Genetics.

The untransmitted diplotypes was an idea conceived by Dr Jeff Barrett. I had help with Perl programming from Dr Dan King, Dr Tomas Fitzgerald and Dr Ray Millar. The analysis looking at variant numbers to adjust QUAL score was carried out working with Dr Tomas Fitzgerald. Dr Tomas Fitzgerald carried out the Mann-Whitney test of the numbers of rare SNPs per sample in probands and untransmitted diplotypes and carried out the PCA. I was responsible for the untransmitted diplotypes dataset generation and analysis from start to finish. I was responsible for the QC analysis and wrote the scripts for untransmitted diplotype generation.

Publications

Resulting from this work:

Peer reviewed articles:

Fitzgerald TW*, Gerety SS*, **Jones WD***, van Kogelenberg M* et al. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015 ;519(7542):223-8.

Wright CF, Fitzgerald TW, **Jones WD**, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. The Lancet. 2015;385(9975):1305-14.

King DA, **Jones WD**, Crow YJ, Dominiczak AF, Foster NA, Gaunt TR, et al. Mosaic structural variation in children with developmental disorders. Human molecular genetics. 2015.

Akawi N, McRae J, Ansari M, Balasubramanian M, Blyth M, Brady AF, *et al*. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nature genetics. 2015;47(11):1363-9.

McRae, JF, Clayton, S, Fitzgerald, Kaplanis J *et al*. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542(7642):433-8.

Book chapters

Jones WD, Hypertrichosis Chapter. Oxford Desk Reference: Clinical Genetics (2nd edition) by Firth HV, Hurst JA. Oxford University Press. In press.

* Denotes joint first author.

Contents

List of Abbreviations

2D	Two dimensional
3D	Three dimensional
aCGH	Array comparative genomic hybridisation
bcf	Binary variant call format
BWA	Burrows-Wheeler Aligner
ADHD	Attention deficit hyperactivity disorder
ALT	Alternate allele
BAM	Binary alignment map
CdLS	Cornelia de Lange Syndrome
CM	Centimetres
CNV	Copy number variant
CpG	5’-C-phosphate-G-3’
DAM	Damaging
DD	Developmental disorder
DNA	Deoxyribonucleic acid
DDD	Deciphering developmental disorders
DDG2P	Developmental Disorder Gene2Phenotype
DNA	Deoxyribonucleic Acid
EMG	Electromyogram
ESP	Exome Sequencing Project
EURODIS	Rare Diseases Europe
ExAC	The Exome Aggregation Consortium
FDNA	Facial Dysmorphology novel analysis
FISH	Fluorescence *in situ* hybridisation
FORGE	Finding of rare disease genes
FUNC	Functional
GAPI	Genome analysis production informatics
GATK	Genome analysis toolkit
GO	Gene ontology
GRCh37	Genome Reference Consortium human genome (build 37)

HbF	Fetal haemoglobin / haemoglobin F
HET	Heterozygous
HOM	Homozygous
HPO	Human phenotype ontology
H3K4	Lysine 4 of histone H3
INDEL	Insertion or deletion
IGV	Integrative Genomics Viewer
Kg	Kilogram
LoF	Loss of function
MAF	Minor allele frequency
Mb	Megabase
MM	Millimetres
MRI	Magnetic resonance imaging
NA	Not available
NHS	National health service
OFC	Occipital frontal circumference
OMIM	Online Mendelian Inheritance in Man
PCA	Principle component analysis
PCR	Polymerase chain reaction
PEG	Percutaneous endoscopic gastrostomy
PEJ	Percutaneous endoscopic jejunostomy
PHD	Plant homeo-domain
PolyPhen	Polymorphism Phenotyping
PROB DAM	Probably damaging
QC	Quality control
QUAL	Variant quality score from GATK
REF	Reference allele
RNA	Ribonucleic acid
RT-PCR	Real time polymerase chain reaction
SD	Standard deviations
SNP	Single-nucleotide-polymorphism
SNV	Single nucleotide variant

SWI/SNF	Switch-Sucrose Non-Fermentable
TDT	Transmission disequilibrium test
UTR	Untranslated region
UK	United Kingdom
VCF	Variant call format
VEP	Variant Effect Predictor
VQSLOD	Variant quality score log-odds
VQSR	Variant Quality Score Recalibration
WISH	Wiedemann-Steiner syndrome or related phenotypes or hypertrichosis
WiSH-WES	Wiedemann-Steiner syndrome and hypertrichosis whole exome sequencing
WSS	Wiedemann-Steiner syndrome
WSSP	Wiedemann-Steiner syndrome phenotype
WTSI	Wellcome Trust Sanger Institute