

**Identifying and characterising novel genomic variants and  
disease mechanisms in patients with rare and novel  
diseases**

A thesis submitted to The University of Manchester for the degree of  
Doctor of Philosophy  
in the Faculty of Biology, Medicine, and Health

**2022**

**J. Robert Harkness**

Faculty of Biology, Medicine and Health, Division of Evolution,  
Infection and Genomics

# Contents

List of Figures.....	6
List of Tables .....	7
Abbreviations.....	8
Abstract .....	12
Lay Abstract .....	13
Declaration .....	14
Copyright Statement .....	15
Acknowledgements .....	16
Rationale for Journal Format.....	17
Author contributions .....	18
Chapter 1: Introduction .....	21
1.1 Rare disease.....	21
1.1.1 Rare disease and the diagnostic odyssey.....	21
1.1.2 Aetiology of rare disease .....	22
1.1.3 <i>De novo</i> variants .....	24
1.1.4 Consanguinity and Endogamy: drivers of rare disease in South Asia .....	25
1.2 Genomic approaches in rare disease diagnosis .....	27
1.2.1 Advances in genomics.....	27
1.2.2 Strategies for NGS analysis .....	29
1.2.3 Processing and analysis of NGS data .....	30
1.2.4 Population databases .....	32
1.2.5 <i>In silico</i> modelling of genomic variants .....	34
1.2.5.1 Splicing models .....	38
1.2.5.2 Aggregating <i>in silico</i> predictions .....	39
1.2.6 Functional and disease annotation.....	40
1.2.7 Classifying variants.....	42
1.2.8 Analysis of structural variants.....	45
1.2.9 Current clinical approaches .....	46
1.3 Challenges in rare disease genomics .....	49
1.3.1 Missing heritability .....	49
1.3.2 Challenges in detecting poorly characterised genetic mechanisms in disease .....	50
1.3.2.1 Structural variants .....	50
1.3.2.2 Non-coding variants.....	51
1.3.2.3 DNA methylation .....	52
1.3.2.4 Variants in non-monogenic disease.....	53

1.3.3 Challenges in interpreting variants of unknown significance .....	54
1.3.3.1 Variants of Unknown Significance .....	54
1.3.3.2 Functional evaluation .....	55
1.3.3.3 Reassessment of unresolved data .....	57
1.4 Research aims and objectives.....	59
Chapter 2: Methods and Materials.....	61
2.1 Ethics approval .....	61
2.2 Patient samples .....	61
2.2.1 DNA extraction from patient blood samples .....	61
2.2.2 RNA extraction from patient cells.....	61
2.2.3 DNA and RNA quantification.....	62
2.3 Next generation sequencing.....	62
2.3.1 WGS .....	62
2.3.2 WES BGI .....	62
2.3.3 WES Illumina .....	62
2.3.4 NGS data processing.....	62
2.3.5 Analysis of single nucleotide variants.....	63
2.3.6 Analysis of copy number variants .....	64
2.4 Sanger sequencing.....	65
2.4.1 Designing primers .....	65
2.4.2 Polymerase chain reaction.....	65
2.4.3 Agarose gel electrophoresis.....	65
2.3.4 PCR product purification.....	66
2.3.5 Sanger sequencing reactions .....	66
2.3.6 Sequencing reaction product purification and capillary electrophoresis .....	66
2.3.7 Sanger sequencing analysis .....	67
Chapter 3: Early B-cell Factor 3-Related Genetic Disease Can Mimic Urofacial Syndrome .....	68
Chapter 4: Clinically diverse and perinatally lethal syndromes with urorectal septum malformation sequence.....	69
Chapter 5: Deep Intronic Variant Causes Aberrant Splicing Of <i>ATP7A</i> In A Family With A Variable Occipital Horn Syndrome Phenotype .....	70
5.1 Abstract .....	71
5.2 Introduction.....	71
5.3 Results: .....	74
5.3.1 Clinical features .....	74
5.3.2 Genetic analysis .....	79

5.3.3 Intronic <i>ATP7A</i> variant results in altered splicing .....	79
5.3.4 Proband <i>ATP7A</i> RNA analysis.....	81
5.4 Discussion: .....	84
5.5 Methods .....	87
5.5.1 SNP Array Genotyping .....	87
5.5.2 Exome Sequencing.....	87
5.5.3 Whole Genome Sequencing .....	87
5.5.4 Sanger Sequencing.....	88
5.5.5 Minigene Assay .....	88
5.5.6 RNA extraction from patient cells.....	89
5.5.7 Quantitative RT-PCR .....	89
5.6 Acknowledgements .....	90
5.7 References .....	91
Chapter 6: Biallelic Variants in <i>RCC1</i> result in fever-associated axonal neuropathy with encephalopathy .....	94
6.1 Abstract .....	95
6.2 Introduction .....	96
6.3 Results: .....	99
6.3.1 Clinical findings .....	99
6.3.2 Genetic analysis of Family A .....	103
6.3.3 Identification of further affected individuals in the 100,000 genomes project.....	103
6.3.4 <i>Rcc1</i> <sup>G43S</sup> retains Ran GEF activity.....	105
6.3.5 <i>Rcc1</i> localisation changes in fibroblasts of individuals with ANE.....	108
6.4 Discussion .....	110
6.4.1 Implications of <i>Rcc1</i> chromatin delocalisation .....	110
6.4.2 Induction of <i>Rcc1</i> dysfunction .....	111
6.4.3 Ran-related dysfunction in hereditary neuropathies.....	113
6.5 Materials and Methods .....	115
6.5.1 Autozygosity mapping .....	115
6.5.2 Exome Sequencing.....	115
6.5.3 Generation of constructs for expression of recombinant <i>Rcc1</i> and Ran proteins.....	115
6.5.4 Production of single-stranded DNA for site-directed mutagenesis .....	116
6.5.5 Site-directed mutagenesis of <i>RCC1</i> .....	116
6.5.6 Expression and purification of recombinant <i>Rcc1</i> and Ran .....	116
6.5.7 GEF assay .....	117
6.5.8 Thermal Shift Assay .....	117



6.5.9 Culture of patient fibroblasts.....	118
6.5.10 Immunofluorescence Imaging .....	118
6.6 References .....	120
Chapter 7: Discussion .....	125
7.1 Eligibility and participation in genetic screening programmes.....	126
7.2 Identifying rare disease aetiology in patients with complex phenotypes .....	128
7.3 Detection of novel variants associated with rare disease .....	130
7.3.1 Splicing .....	131
7.3.2 Oligogenic disease .....	132
7.3.3 Structural variation .....	133
7.3.4 Non-coding variants.....	134
7.4 Characterisation of variants of uncertain significance .....	136
7.5 Limitations of approach and study .....	139
7.6 Further work .....	140
7.7 Final Summary .....	141
References .....	142
Appendices .....	168
Appendix I: Supplementary data for Chapter 3 .....	169
Appendix II: Supplementary data for chapter 4 .....	173
Appendix III: Supplementary data for Chapter 5 .....	180
Appendix IV: Supplementary data and updates for Chapter 6.....	182
Appendix V: Data from an additional investigation.....	187

Word count: 36,815

## List of Figures

Figure 1.1	Overview of a standard WGS/WES bioinformatics workflow	31
Figure 2.1	Default filtering strategy used to analyse WGS and WES data	64
Figure 5.1	Family with apparent X-linked skeletal dysplasia	75
Figure 5.2	Evaluation of splicing changes from the <i>ATP7A</i> c.1544-872C>G variant	80
Figure 5.3	Analysis of <i>ATP7A</i> transcripts from P3 (V:2)	82
Figure 5.4	Quantitative PCR of <i>ATP7A</i> transcripts from P3 (V:2) lymphocyte RNA	83
Figure 6.1	Schematic of Ran GTPase action in nucleocytoplasmic transport	98
Figure 6.2	Pedigrees and genetic investigations of individuals with ANE	100
Figure 6.3	<i>In silico</i> and <i>in vitro</i> analyses of Rcc1 <sup>G43S</sup>	107
Figure 6.4	Immunofluorescence of heat-shocked fibroblasts	109

## Supplementary figures

Figure I-1	Screenshot of microarray data for <i>EBF3</i> in the developing murine lower urinary tract	170
Figure I-2	Filtering strategy used for whole exome sequencing data analysis	171
Figure II-1	Clinical images of fetus 1	173
Figure II-2	Clinical images of fetus 5	174
Figure II-3	Clinical images of fetus 6	175
Figure II-4	Clinical images of fetus 7	176
Figure II-5	Clinical images of fetus 8	177
Figure II-6	Clinical images of fetus 9	178
Figure II-7	Clinical images of fetus 10	179
Figure IV-1	Assessment of purification of recombinant Ran and Rcc1 proteins	183
Figure IV-2	Diagram of <i>RCC1</i> cDNA variants identified in patients with fever-associated axonal neuropathy	184
Figure V-1	Wars1/TrpRS catalytic domain amino acid sequence (154-362)	190
Figure V-2	Aminoacylation activity of recombinant Wars1 with variants identified in patients with ID/DD or neuropathy	191
Figure V-3	Temperature stability of recombinant Wars1 with variants identified in patients with ID/DD or neuropathy	192

## List of Tables

Table 1.1	<i>In silico</i> tools for functional predictions of variant effect	35
Table 1.2	Interpretation of CADD-PHRED scores	40
Table 1.3	Overview of criteria for variant pathogenicity classification	43
Table 1.4	Examples of models available for functional evaluation of genetic variants	56
Table 2.1	Thermal cycler program for PCR amplification	65
Table 2.2	Sequencing reaction cycles	66
Table 5.1	Phenotypes observed in five related individuals	76
Table 6.1	Summary of clinical findings in family A	101
Table 6.2	Identification of unrelated probands with biallelic <i>RCC1</i> variants from 100,000 Genomes Project Exomiser data	104

## Supplementary tables

Table I-1	Patient developmental abnormalities	169
Table I-2	<i>In silico</i> predictions for <i>EBF3</i> variant c.626G>A	169
Table I-3	Variants detected in <i>EBF3</i> targeted exome sequencing of VUR cohort	169
Table I-4	PCR reaction components	172
Table I-5	Sequencing reaction components	172
Table III-1	Rare variants identified in exome sequencing of IV:1	180

## Abbreviations

100kGP – 100,000 Genomes Project

aa – Amino Acid

ACMG – American College of Medical Genetics and Genomics

ADEM – Acute disseminated encephalomyelitis

AF – Allele Frequency

AI – Artificial Intelligence

ALS – Amyotrophic Lateral Sclerosis

ANE – Acute Necrotising Encephalopathy

ATP – Adenosine Triphosphate

BAM – Binary Alignment Map

bp – Base Pair

BSA – Bovine Serum Albumin

CADD – Combined Annotation-Dependent Depletion

cDNA – coding DNA

CFS – Chronic Fatigue Syndrome

CGH – Comparative Genomic Hybridisation

CNS – Central Nervous System

CNV – Copy Number Variant

COVID-19 – Coronavirus Disease 2019

CPK – Creatine Phosphokinase

CRISPR – Clustered Regulatory Interspaced Short Palindromic Repeats

CRM1 – Chromosomal Maintenance 1

DAPI – Diamidino Phenylindole

DDD – Deciphering Developmental Disorders

DGV – Database of Genomic Variants

DMEM – Duolbecco's Modified Eagle Medium

DNA – Deoxyribonucleic acid

DTT – Dithiothreitol

EBV – Epstein-Barr Virus

eIF2A – eukaryotic translation Initiation Factory 2A

EMG - Electromyography

ESE – Exonic Splicing Enhancer

ESS – Exonic Splicing Suppressor

ExAC – Exome Aggregation Consortium

FATHMM – Functional Analysis Through Hidden Markov Models

FISH – Fluorescence In Situ Hybridisation

FORGE – Finding of Rare Disease Genes

GATK – Genome Analysis ToolKit

GBS – Guillain-Barré Syndrome

GE – Genomics England

GEF – Guanine Exchange Factor

GO – Gene Ontology

GTE<sub>x</sub> – Genotype-Tissue Expression

GTP – Guanosine Triphosphate

GUaRDIAN – Genomics for Understanding Rare Diseases: India Alliance Network

GVGD – Grantham Variation and Grantham Deviation

GWAS – Genome Wide Association Study

HADDs – Hypotonia, Ataxia and Delayed Development Syndrome

HGMD – Human Gene Mutation Database

HHV6 – Human Herpes Virus 6

HLA – Human Leukocyte Antigen

hnRNPH3 – heterogeneous nuclear Ribonuclear Protein H3

HPO – Human Phenotype Ontology

HRA – Health Research Authority

HSF – Human Splicing Finder

IL-17 – Interleukin

IL-6 – Interleukin 6

KEGG – Koyoto Encyclopaedia of Gene and Genomes

LFT – Liver Function Test

LOEUF – Loss of Function Observed/Expected Upper Bound Fraction

LoF – Loss of Function

LOX – Lysyl Oxidase  
LOXL – Lysyl Oxidase-like  
LVNC – Left Ventricular Non-Compaction  
MAF – Minor Allele Frequency  
MGI – Mouse Genome Informatics  
MIM – Mendelian Inheritance in Man  
MKL – Multiple Kernel Learning  
MMP – Matrix Metalloprotein  
MRI – Magnetic Resonance Imaging  
MRKH – Mayer-Rokitansky-Kuster-Hauser Syndrome  
mRNA – messenger RNA  
NAD – Nicotinamide Adenine Dinucleotide  
NARP – Neuropathy, Ataxia, and Retinitis Pigmentosa  
NCV – Nerve Conduction Velocity  
NGS – Next Generation Sequencing  
NHS – National Health Service  
NPC – Nuclear Pore Complex  
OFC – Occipital Frontal Circumference  
OHS – Occipital horn syndrome  
OMIM – Online Mendelian Inheritance in Man  
PBS – Phosphate Buffered Saline  
PCOS – Polycystic Ovary Syndrome  
PCR – Polymerase Chain Reaction  
pLI – Probability of LoF Intolerance  
qPCR – quantitative PCR  
RCC1 – Regulator of Chromatin Condensation  
RDMM – Rare Disease Models and Mechanisms network  
REVEL – Rare Exome Variant Ensembl Learner  
RNA – Ribonucleic Acid  
RNAi – RNA interference  
RT-PCR – Reverse-Transcription PCR

SARS-CoC-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SIFT – Sorting Intolerance from Tolerant

SMAX3 – Spinomuscular Ataxia, X-linked type 3

SNP – Single Nucleotide Polymorphism

SNV – Single Nucleotide Variant

SV – Structural Variant

TAD – Topologically Associated Domain

TGN – Trans-Golgi Network

TLR3 – Toll-Like Receptor 3

UFS – Urofacial Syndrome

URSMS – Urorectal Septum Malformation Sequence

VCF – Variant Call File

VEP – Variant Effect Predictor

VLCFA – Very Long Chain Fatty Acids

VUR – Vesicoureteric reflux

VUS – Variant of Unknown Significance

WES – Whole Exome Sequencing

WGS – Whole Genome Sequencing

WT – Wild Type

## Abstract

Rare diseases are individually uncommon, but cumulatively affect 1 in 17 people. Development of next generation sequencing technologies including whole genome sequencing (WGS) and whole exome sequencing (WES), and associated data analysis algorithms allow the identification of rare and novel genetic variants that underly rare inherited diseases. However, for many patients diagnosis is challenging due to the complexity involved in predicting effect of novel variants in non-coding regions and genes that are poorly described. The aim of this work is to identify novel genomic variants which contribute to monogenic disease in patients with rare diseases, and to characterise any novel mechanisms of disease. We carried out analysis of WGS or WES data from several individuals affected by different rare diseases.

In an individual with atypical urofacial syndrome WES analysis identified a variant in *EBF3*, which causes hypotonia, ataxia, and developmental delay syndrome, within which urological phenotypes are common. WGS analysis was carried out for a cohort of foetuses with urorectal septum malformation sequence following detailed foetal autopsy. WGS was not diagnostic in this cohort, but did identify candidate variants in *HOXD9*, *NALCN*, *SLIT6*, *TMEM132A*, implying an oligogenic model of disease.

WGS data were reanalysed from a family with a variable occipital horn syndrome phenotype, leading to identification of a novel deep intronic *ATP7A* variant. Analysis of splicing in a minigene assay and patient RNA revealed a leaky splicing effect predictive of protein truncation.

Autozygosity mapping and WES analysis of a consanguineous family with several children affected by acute necrotising encephalopathy secondary to febrile infection identified a novel homozygous variant in *RCC1*. Further patients were identified though reverse phenotyping. Enzymatic and stability assays using recombinantly expressed protein refuted a hypothesis of thermolability. Immunofluorescent imaging suggests delocalisation of Rcc1 from chromatin, prompting further investigation into the involvement of nucleocytoplasmic transport dysfunction.

This work displays the value offered by undertaking genomic analysis of patients in a research setting and highlights the challenges that remain in detecting genomic variants in patients with complex phenotypes.



## Lay Abstract

Diseases that occur in fewer than 1 in 2000 people are defined as rare diseases. Added together, different rare diseases affect 1 in 17 people. The genetic cause for rare diseases can be identified by sequencing either a patient's whole genome, or part of the genome which contains only genes. However, many tools used to analyse genomic data have limitations. This prevents many patients affected by rare diseases from getting a diagnosis. This work aims to identify genetic changes which have not been seen before in patients with rare disorders, and to discover how these genetic changes result in disease.

Analysis of a patient with a bladder disorder called urofacial syndrome identified a genetic change in the *EBF3* gene. This gene is already known to cause a disorder that mostly affects nerve function and movement, but sometime causes problems with bladder nerves that are similar to patients with urofacial syndrome.

Genome analysis of foetuses with severe impaired bladder development as well as multiple other developmental complications did not identify genetic changes in genes previously identified. The cause of this complex disorder is likely to result from multiple genetic changes having a combined effect that are difficult to detect.

In a family where males are affected by a copper uptake disorder resulting in abnormal bone and joint growth, digestive problems, and wiry hair, genome analysis revealed a change outside of the regions which contains genes. This genetic change causes the *ATP7A* gene to be read incorrectly resulting in incomplete expression of the gene.

In a large family, multiple children were affected by a brain degenerative disorder triggered by a fever. Genetic analysis identified a change in the *RCC1* gene. By looking at a national genome database, we identified other patients with changes in *RCC1* with similar symptoms. By expressing the altered Rcc1 protein from the *RCC1* gene, we found that the genetic change did not affect the stability of the protein when heated to simulate a fever. By looking at patient cells under a microscope we found that the Rcc1 protein changes its location after fever simulation. This suggests that transport of other proteins in the cell may be impaired, and we plan to investigate this further in the future. Overall, these investigations have shown the value in taking a deeper look into the genomes of patients with rare disease, and the challenges that remain in detecting certain kinds of genetic changes.

## **Declaration**

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

## Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and they have given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, the University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in the University’s policy on Presentation of Theses.

## Acknowledgements

Firstly, I would like to thank my supervisory team Prof. William Newman, Prof. Siddharth Banka, and Prof. Simon Hubbard for their guidance and support. I would like to thank Prof. William Newman for sharing his vast experience and advice which were instrumental in the planning and execution of analysis, experiments, and preparation of manuscripts. Special thanks also go to Prof. Raymond O’Keefe for teaching me many laboratory techniques used in the preparation of this thesis.

Thanks also go to the other members of the Newman and O’Keefe research groups, and the wider Genomic Medicine department at St Mary’s Hospital. Particular thanks to Dr. Glenda Beaman, Dr. Huw Thomas, and Dr. Cristina Perez-Becerril for all their support and the scientific expertise they have shared with me. I am also grateful to the Wellcome Trust programme directors including Dr. Sarah Woolner my advisor, and Prof. Hilary Ashe for recognising my potential, and supporting me during my time at Manchester.

I would also like to acknowledge my collaborators who have worked closely with me on different aspects of this work. I am particularly grateful to Prof. Katta Girisha and Dr. Shalini Nayak for their special collaboration and providing me with their wonderful hospitality at the University of Manipal. Thanks also to Dr. Leigh Demain and Dr. Jill Urquhart for their early work in the identification of the *RCC1* variant, and to Dr. John McDermott for his clinical insight.

I must also thank all my family and friends for the support they have shown despite many not understanding about the world of genetics or academia. I owe much to the steadfast encouragement and care of my parents Julia and John during my progression through higher education. Specifically, thanks go to Rosie, Annabel, Neil, and Aris for their advice, support, and encouragement during the research and writing of this thesis.

## Rationale for Journal Format

This thesis contains four separate investigations united under one common theme. In each experimental chapter, genomic variants were identified and/or characterised in patients, families, or cohorts with different rare diseases. Different experimental approaches have been used to identify and characterise variants in individuals with different rare diseases, so distinct chapters formed around each case study. The case study approach exemplifies the breadth of the field relating to conducting genetic research into the aetiology of disease in individuals with rare disease. The timeline of each case study also varied largely as a result of delays relating to sample and data acquisition during the COVID-19 pandemic from March 2020.

Chapters 3 and 4 have been published in *Kidney International Reports* (Kidney Int Rep (2020) 5, 1823–1827. DOI: 10.1016/j.ekir.2020.07.001) and *American Journal of Medical Genetics* (Part A, 1–12. DOI: 10.1002/ajmg.a.63067) respectively. Chapter 5 is prepared for publication but has not yet been submitted. Chapter 6 is not yet ready for publication as further work is being planned to complete the characterisation of homozygous *RCC1* variants in disease.

## Author contributions

Identifying novel genetic variants and characterising their role in rare disease can take time, and so several collaborators have made contributions to much of the work presented in this thesis. Contributions for each experimental chapter are described below.

### **Chapter 3: Early B-cell Factor 3-Related Genetic Disease Can Mimic Urofacial Syndrome**

Kidney International Reports, 5, 1823–1827; DOI: 10.1016/j.ekir.2020.07.001

J. Robert Harkness, Glenda M. Beaman, Keng W. Teik, John A. Sayer, Heather J. Cordell, Huw B. Thomas, Katherine Wood, Helen M. Stuart, Adrian S. Woolf, William G. Newman.

I assisted with whole exome data analysis with guidance from GBM. I conducted targeted sequencing of *EBF3* exons in the vesicoureteric reflux (VUR) study group cohort. I reviewed literature and prepared the manuscript under guidance from WGN and ASW.

GBM extracted DNA for whole exome sequencing. JAS, HJC and HMS oversaw accumulation of the VUR study group cohort. KWT was the clinician involved in care of the proband. HBT and KW were consulted regarding splicing effects of synonymous variants identified in the VUR study group. All authors commented on the manuscript. The planning and execution of this work was supervised by WGN.

### **Chapter 4: Clinically diverse and perinatally lethal syndromes with urorectal septum malformation sequence**

American Journal of Medical Genetics Part A, 1–12; DOI: 10.1002/ajmg.a.63067

Shalini S Nayak, J. Robert Harkness, Anju Shukla, Siddharth Banka, William G Newman, Katta M Girisha.

I conducted whole genome sequencing analysis of eight fetuses with urorectal septum malformation sequence (URSMS). I contributed to the writing of the methods, results, and discussion of the manuscript.

The URSMS cohort autopsies, sample collection and DNA extraction were performed by SSN, following clinical consultation and recruitment by AS and KMG. The manuscript was prepared by SSN and me, under supervision of KMG, WGN. All authors commented on the manuscript. The planning and execution of this work was supervised by KMG, WGN, and SB.

## **Chapter 5: Deep Intronic Variant Causes Aberrant Splicing Of ATP7A In A Family With A Variable Occipital Horn Syndrome Phenotype**

J. Robert Harkness, Huw B. Thomas, Helen M. Kingston, Jill E. Urquhart, Genomics England Research Consortium, Simon Hubbard, Raymond T. O’Keefe, Siddharth Banka, William G. Newman, Charu Deshpande

I analysed WGS data from the 100,000 genomes project research environment and identified the deep intronic variant in *ATP7A*. I prepared and conducted the minigene assay with guidance from HBT and RTO. I prepared and conducted RNA extraction and transcript analysis. I prepared the manuscript under guidance of WGN and CD.

Clinical characterisation of the family was conducted by CD, HMK, and WGN. Initial autozygosity mapping and whole exome sequencing was conducted by JEU. Whole genome sequencing and data processing was conducted by GERC. The planning and execution of this work was supervised by WGN, SB and SH.

## **Chapter 6: Biallelic Variants In *RCC1* result in fever-associated axonal neuropathy with encephalopathy**

J. Robert Harkness, John McDermott, Kay Metcalfe, William L. Macken, Ataf Sabir, Saikat Santra, Jill E. Urquhart, Leigh A. M. Demain, Christopher J. Record, Wyatt W. Yue, Helen Byers, Robert D.S. Pitceathly, Robert W. Taylor, Simon Hubbard, Siddharth Banka, Genomics England Research Consortium, Mary M. Reilly, Raymond T. O’Keefe, William G. Newman

I planned and undertook recombinant protein expression under guidance from RTO. I conducted GTPase assays, thermal shift assays, and associated data analysis. I cultured, stimulated, and processed proband fibroblasts for immunofluorescence imaging. I conducted immunofluorescence imaging of associated data processing. I identified families

B and C using data from the 100,000 genomes project research environment. I prepared the manuscript under guidance from WGN and JM.

JEU, LAMD and HB conducted autozygosity mapping, whole exome sequencing, identified the *RCC1* variant, and conducted sanger sequencing to confirm segregation. WWY conducted structural modelling of the Rcc1 variant. Clinical characterisation of family A was conducted by MD and KM. Clinical characterisation of family B was conducted by AS and SS. Clinical characterisation of family C was conducted by WLM, RDSP, and MMR. Whole genome sequencing and data processing of samples for families B and C was conducted by GERC. Screening for respiratory chain defects was arranged through RWT. The planning and execution of this work was supervised by WGN, RTO, SB, and SH.



# Chapter 1: Introduction

## 1.1 Rare disease

### 1.1.1 Rare disease and the diagnostic odyssey

A rare disease is defined as a condition which has very low prevalence in the general population. Numerically, this is defined by the European Commission as one which affects fewer than one in 2,000 people (Eurordis, 2020), and in the United States as one which affects fewer than 1 in 5,000 people (Álvaro-Sánchez et al., 2021; Nguengang Wakap et al., 2020). There are more than 8,000 different rare diseases, involving all organs and tissues of the body, ranging from rare cancers to autoimmune disorders, and neuromuscular syndromes. Rare diseases may be congenital or develop at any stage of life. Although rare diseases are uncommon individually, at some point in their lives, 1 in 17 people will be affected by a rare disease (Eurordis, 2020). The cumulative burden of rare diseases on society is therefore substantial. Management of rare diseases varies, with different care pathways, support, and therapies to consider.

The biggest barrier to care for individuals with rare diseases is diagnosis, which typically takes 5-10 years or more, and often requires referral to numerous different specialists (Austin et al., 2018). Termed the ‘diagnostic odyssey’, the time from symptom onset to diagnosis is difficult for patients and their wider support environment, and incurs considerable healthcare costs. In a report by Mendelian (Angelis et al., 2015), an organisation which advocates for rare disease patients, £ 3.4 billion was spent by the UK National Health Service on care of rare disease patients in the ten years prior to diagnosis (Angelis et al., 2015). One important finding which contributes to this is that rare disease patients are more likely to visit Accident and Emergency departments, outpatient clinics and be admitted to hospital wards than the general population (Angelis et al., 2015). Streamlining diagnostic pathways for individuals with rare disease is therefore important for patient welfare.

Providing an accurate diagnosis in a shorter amount of time may also lead to a reduction in unnecessary medical procedures and treatments, by better targeting therapies or

interventions to patients. However, with 94 % of rare diseases lacking approved treatments (Austin et al., 2018), more research is necessary to enable faster diagnosis and greater targeting of therapies for rare disease patients. Indeed, a socioeconomic analysis suggests that greater investment should be made in the management and treatment of rare diseases, as the direct healthcare cost of treatment is outweighed by the reduction of indirect costs associated with patient quality of life and productivity (Fernandez-Marmiesse et al., 2018).

For rare diseases to be managed most effectively by patients, carers, and healthcare teams, a molecular diagnosis is necessary. A molecular diagnosis describes the specific genetic variant (genotype) which causes a disease's traits or characteristics (phenotype). A genetic cause is expected in most cases of rare diseases, and it is estimated at least 80 % of rare diseases display a genetic origin (Eurordis, 2020; Rajasimha et al., 2014). Genetics therefore plays an important part in the diagnostic journey for most patients.

#### 1.1.2 Aetiology of rare disease

Transmission of common familial traits - such as eye colour or facial appearance - through generations has been remarked since well before the rise of genetic science. The thesis of Gregor Mendel in the 19th century fostered a culture for the observation of heredity, which precipitated the understanding of dominant and recessive traits (Bateson and Mendel, 2013; Singh, 2021). This concept of trait inheritance applies just as much to the study of rare diseases. Indeed in monogenic diseases, where a disease phenotype is associated with a genetic variant in a single gene that can be transmitted either in a dominant or recessive manner is described as mendelian as a result of these seminal observations.

Inherited autosomal rare diseases, which affect genes on chromosomes 1-22, may be transmitted either in a dominant or recessive manner. In autosomal dominant conditions only one disease allele is required for disease penetrance, such as transmission of a genetic variant from an affected parent to an affected child. Arrhythmogenic right ventricular cardiomyopathy (Mendelian Inheritance in Man (MIM): 107970) is an example of an

autosomal dominant disease, caused by heterozygous variants in various genes involved in cell-cell binding, such as *DSP* (MIM: 125647) (McNally et al., 1993; Norgett et al., 2000).

Alternatively, diseases which are transmitted in a recessive manner, where two disease alleles (biallelic) are necessary for disease penetrance usually requires inheritance of one disease allele from each unaffected carrier parent by an affected child. Autosomal recessive transmission can be exemplified with cystic fibrosis where typically disease occurs in individuals who carry biallelic variants in the *CFTR* gene (MIM: 602421) (Savant et al., 1993). In recessive disease, an affected individual can be homozygous for a pathogenic variant where the same variant is inherited from each heterozygous parent. Alternatively, an affected individual can be compound heterozygous, where different pathogenic variants are inherited from each parent.

Sex-linked disorders - in which genes on the X or Y chromosomes are affected - have different patterns of transmission, resulting from female possession of XX chromosomes and male possession of XY chromosomes (Lau, 2020). The most common mechanism of sex-linked disorder is X-linked recessive disorders. These are more commonly identified due to their penetrance in hemizygous males who only inherit one X-chromosome from their mothers. A male that is hemizygous for an X-linked recessive disorder will be affected, and there is a 50 % chance of a carrier female having an affected son. One example of X-linked disease is haemophilia type A, where inheritance of a pathogenic variant in X-chromosome coagulation factor VIII gene *F8* (MIM: 300841) causes disease predominantly in males (Antonarakis et al., 1995; Konkle and Nakaya Fletcher, 1993). For females to be affected by an X-linked recessive disorder, X-linked variants in the same gene must be inherited from each parent. Additionally, skewed X-chromosome inactivation in females can lead to variable penetrance of X-linked recessive phenotypes (Ross et al., 2005). Y-linked disorders only affect males and are less commonly identified likely due to the reduced gene content of the Y-chromosome compared to the X-chromosome (Lau, 2020). Y-chromosome instability, and microdeletions affecting numerous genes on the Y-chromosome are associated with spermatogenic failure (MIM: 415000) and male infertility (Zhu et al., 2020). So, inheritance of Y-chromosome disorders is unusual, and usually only possible following

participation of clinically affected fathers in *in vitro* fertilisation using intracytoplasmic sperm injection (Fan and Silber, 1993).

Finally, maternally inherited disorders can occur as a result in variants in genes located in the mitochondrial genome. Mitochondria are organelles which are the major source of energy production by facilitating the biochemical process of oxidative phosphorylation. As intracellular organelles, mitochondria are inherited by offspring via the maternal ovum, and so mitochondrial disorders display a characteristic maternal inheritance. Genetic variants in mitochondrial gene *MT-ATP6* are associated with maternally inherited Neuropathy, ataxia, and retinitis pigmentosa (NARP) and Leigh Syndrome (MIM: 516060) (D'Aurelio et al., 2010; Heuer and Seibert, 2022).

Understanding inheritance patterns of a rare disease is important for two reasons. Firstly, it is useful to be able to predict risk of future offspring being affected by a familial disease during genetic counselling (Abacan et al., 2019). Secondly, knowledge of the expected mode transmission is useful to incorporate into diagnostic workflows when trying to identify a causal variant.

### 1.1.3 *De novo* variants

Rare diseases by definition occur very infrequently in a population, however a rare disease can occur for the first time in a previously unaffected family due to spontaneous genetic mutation. These are termed *de novo* genetic variants and are often identified in patients where there is no family history of disease. Mutation of germline DNA that results in the formation of *de novo* variants results from the numerous rounds of cell division during the process of gametogenesis (Conrad et al., 2011). The average mutation rate for single nucleotide variants (SNVs) per generation is  $1.18 \times 10^{-8}$  per nucleotide, which produces ~74 *de novo* variants in an individual's genome (Veltman and Brunner, 2012). *De novo* variants are less stringently exposed to evolutionary selection compared to inherited variants, which contributes to the tendency of *de novo* variants to be more deleterious on average than inherited variants (Eyre-Walker and Keightley, 2007). Identification of *de novo* variants in an individual with a negative family history of disease is therefore a priority when conducting

genetic analysis. This approach is possible by sequencing the genome or exome of both biological parents in addition to the affected child in order to establish which variants are inherited, and which are *de novo*. This is a powerful approach available as a result of the introduction of massively parallel next-generation sequencing (NGS) methods, which allow detection of novel variants and to formulate novel mendelian gene-disease associations (Gilissen et al., 2011).

#### 1.1.4 Consanguinity and Endogamy: drivers of rare disease in South Asia

Individuals born from consanguineous relationships are predisposed to transmission of recessive conditions following inheritance of a rare damaging variant from an ancestor common to both parents. This effect is illustrated in ‘founder’ population groups, where factors such as the geographical isolation of the non-European Finnish population (Lim et al., 2014), or the cultural and religious practice in the Ashkenazi Jewish population (Cox et al., 2018) contribute to the formation of communities with little gene admixture that are predisposed to propagation of specific *de novo* variants. Both these example populations suffer disproportionately from disease-causing loss-of-function variants, making them useful cases for understanding the generation of novel genetic diseases.

Similarly, some populations in south-Asia are highly endogamous. Endogamy is the isolation of a population as a result of cultural and ethnic differences. In India, the adoption of the caste structure over many generations has given rise to a multitude of communities which are characterised by a high degree of reproductive isolation (Majumder and Basu, 2014). Four main castes among the Hindu population are subdivided into more than 4000 sub-castes, plus the co-existence of Muslim, Christian and Sikh minorities have led to the development of a heterogeneous population characterised by high rates of endogamy (Mastana, 2014). This genetic phenomenon is neglected in many Caucasian-centric studies, and publicly available population-level genetic datasets such as gnomAD (<https://gnomad.broadinstitute.org/>) (Karczewski et al., 2019) regard the south-Asians as a homogeneous population group.

The unique heredity architecture among south-Asians has led to a higher predisposition of founder events compared to the Ashkenazi Jewish population (Nakatsuka et al., 2017). Indeed, an estimation by Organisation for Rare Diseases India (ORDI) predicted that 56 million individuals (4.67 % of the population) may suffer from monogenic diseases in India (Rajasimha et al., 2014). The establishment for the Genomics for Understanding Rare Diseases: India Alliance Network (GUaRDIAN) consortium, which oversees genetic research to diagnose patients in India with rare and unique disorders (Bajaj et al., 2019), has led to molecular diagnoses unique to Indian populations. Novel *KRT5* (MIM 148040) (Vellarikkal et al., 2014) and *KRT14* (MIM: 148066) (Yenamandra et al., 2017) variants associated with Epidermolysis Bullosa in Indian pedigrees, and rare variants in *MLC1* (MIM: 159552) were identified in nine patients from a single north-Indian community with leukodystrophy (Vellarikkal et al., 2018). Given the number of individuals predicted to suffer from rare disease in India, there remain a lot of unsolved and unreached cases, which represents a great endeavour for genomics in India. Analysis of these cases will likely yield novel insights into the aetiology and mechanism of rare and unknown diseases.

## 1.2 Genomic approaches in rare disease diagnosis

### 1.2.1 Advances in genomics

Initial sequencing of the 3.2 billion nucleotides that make up the human genome (Lander et al., 2001) represented a huge step forward in being able to understand human biology at a fundamental level. By harnessing the original Sanger sequencing method, a plethora of alternative methods now offer massively parallel sequencing of more individuals, in less time, than ever before. From 2010, the cost of sequencing started to decline, and next-generation sequencing (NGS) methods started to become a tool available to many research and diagnostic laboratories. Genetic diseases are relevant to all medical specialities, and rare diseases represent a continuous global healthcare challenge (Brittain et al., 2017). Advances in NGS technologies provide scope for assessing diagnosis of rare conditions on an individual, patient-by-patient, or family-by-family basis. This section describes the advances in genomics, associated NGS analytical pipelines, and implementation in a diagnostic setting.

NGS platforms are particularly valuable when considering monogenic diseases with low incidence rates, such as Juvenile Arthritis (1 in 20,000, MIM: 618795) (Bernatsky et al., 2007), or Epidermolysis bullosa (1 in 40,000, MIM: 226650) (Angelis et al., 2015). In certain cases, NGS technologies, which include targeted gene sequencing panels, whole exome sequencing (WES) and whole genome sequencing (WGS) are able to provide the data necessary to reach a molecular diagnosis and end a long ‘diagnostic odyssey’ that patients often face (Fernandez-Marmiesse et al., 2018). Prior to the introduction of NGS technologies, it would not be uncommon for rare disease patients to undergo multiple diagnostic investigations, re-diagnoses, being passed between clinical services. The paradigm shift resulting from the application of NGS, particularly introduction of exome-wide sequencing of the protein-coding parts of the genome to the diagnostic pathway has had a major impact on patient diagnosis. Before the advent of NGS, targeted sequencing of exons in known disease-causing genes did not easily allow novel identification of gene-disease associations in patients with rare disease (Samocha et al., 2014). This change in diagnostic ability is exemplified in numerous early studies which implemented WES, demonstrating utility as a diagnostic tool (Deciphering Developmental Disorders Study,

2015; Gilissen et al., 2014; Yang et al., 2014). One example this in practice is the Canadian Finding Of Rare Disease GENes (FORGE) study, which was able to implement WES in order to diagnose 105 of 362 families with rare disease (diagnostic rate 29 %) (Sawyer et al., 2016). Of the patients who were diagnosed 72 % were diagnosed by WES after their first clinic visit, which reinforces WES as a powerful diagnostic tool in terms of patient time-to-diagnosis and healthcare costs (Sawyer et al., 2016). Current clinical approaches in the UK including the Deciphering Developmental Disorders (DDD) study and 100,000 genomes project are outlined further in Section 1.2.9.

Furthermore, the proliferation of microarray-based comparative genomic hybridisation (aCGH) methods for detecting single nucleotide polymorphisms (SNPs) also represented a cost and time-effective mode of determining specific genetic features in patient cohorts with a view to identifying chromosomal changes such as copy number variants (CNVs). Additionally, SNPs were identified that could be used as markers to aid in diagnosis of common diseases, particularly in cancer (Pant et al., 2014). The use of different technologies to assess genetic susceptibility or make a genetic diagnosis is being adapted as WES/WGS becomes cheaper. Polygenic risk scores for common diseases such as breast cancer or coronary artery disease can be determined from WGS data with similar imputation accuracy, while also overcoming the intrinsic ascertainment bias of selecting variants to include in a certain genotyping array (Homburger et al., 2019). Similarly, mapping of genome-wide data in WGS permits the more accurate detection of CNVs breakpoints compared to microarray methods, as well as the possibility to detect other structural variants including inversions or translocations (Rapti et al., 2022). This represents a shift from retrieving targeted genotyping data for specific analyses, to sequencing whole genomes and conducting genotyping analyses post-hoc.

Moreover, WES and WGS offer an unbiased approach to gene discovery. WES has been preferred historically due to the ease of interpretation of variants within coding regions of the genome, and feasibility studies such as one conducted by Columbia University Medical Center established that WES has clinical use in precipitating medical interventions and improving patient and family education through genetic counselling (Iglesias et al., 2014). Over time, aided by reduction in cost, WGS has become increasingly applied to rare disease



studies because of the comprehensiveness of the data provided (Brittain et al., 2017). WGS is now more affordable, and when investing in sequencing it makes more sense to have genomic data available rather than be limited to exome data, as WGS also has the advantage that exome data can still be analysed as per similar analysis conducted with WES (Fernandez-Marmiesse et al., 2018).

### 1.2.2 Strategies for NGS analysis

When considering genetic analysis of patients with rare diseases, there are different options for structuring the process of identifying the aetiological variant. Initially, a hypothesis will be determined, and if the phenotypes have previously been associated with a disease, a gene-specific hypothesis can direct genetic analysis. When a patient has phenotypes which are not associated with a specific gene, or testing for variants in a known gene has returned a negative test, a gene agnostic approach is justified (MacArthur et al., 2014). In a gene agnostic approach different frameworks can be applied depending on the number of affected family members, the hypothesised mode of transmission, presence of non-related individuals with the same phenotypes, and participation of unaffected family members.

If there is a positive family history of disease, a family linkage approach can be used to identify variants inherited by affected individuals, and not inherited by unaffected family members. Autozygosity mapping is particularly useful in identifying the cause of recessive diseases in consanguineous families (Carr et al., 2013).

Where there is no family history of disease, parent-child trios can be analysed to identify *de novo* variants. In this scenario, the cause of disease is hypothesised to be a dominant *de novo* variant. Inherited variants are assumed to be non-pathogenic, reducing the pool of variants for consideration (Fernandez-Marmiesse et al., 2018). If no causal *de novo* variant is identified, trio sequencing also permits the identification of genes with biallelic variants inherited from each parent. This makes the trio approach effective in identifying causative variants compared to patient-only analyses (Farwell et al., 2015). In situations where trio analysis has either not successfully identified a variant, or where trio analysis is not possible as a result of parental non-participation, an intersection approach can be devised if a cohort

of non-related individuals have been identified. In this approach, intersection of pathogenic variants in the affecting the same gene in multiple individuals can be particularly advantageous in the identification of novel gene-disease relationships (Fernandez-Marmiesse et al., 2018).

With numerous approaches to identify variants in cases where there is no previous disease-gene association to form a hypothesis, diagnosis of rare disease patients can be challenging. Many cases remain unsolved, often decades after initial investigations begin. Reanalysis of patient data, and re-sequencing of patient DNA using WGS and implementation of updated bioinformatic pipelines can be useful for investigating unsolved cases (Liu et al., 2019). It is important when considering probands with rare - or potentially unique - phenotypes that the diagnostic approach does not rely on analysis of coding regions. Methods of investigating non-coding variants are often underused, partly due to the inability to confidently interpret the consequences of a genetic variant which may not directly impact protein structure or function.

### 1.2.3 Processing and analysis of NGS data

Since the adoption of NGS into genetic research, the workflow bottleneck has shifted from the cost of sequencing to the processing of large sets of genomic data (H. Y. Chen et al., 2019). The workflow for processing genomic sequencing data, depicted in Figure 1, outlines the main steps in converting raw sequencing data of usually small reads (<200 base pairs) to a meaningful and appropriate list of genetic variants. For standardisation of bioinformatic pre-processing of NGS data and calling of SNVs and structural variants (SVs), best practice is outlined by Genome Analysis Toolkit (GATK; <https://gatk.broadinstitute.org/hc/en-us>). Following variant calling and exporting of files as Variant Call Files (VCFs), computational annotation and filtering of genetic variants is necessary to provide the context for interpretation (Butkiewicz and Bush, 2016).

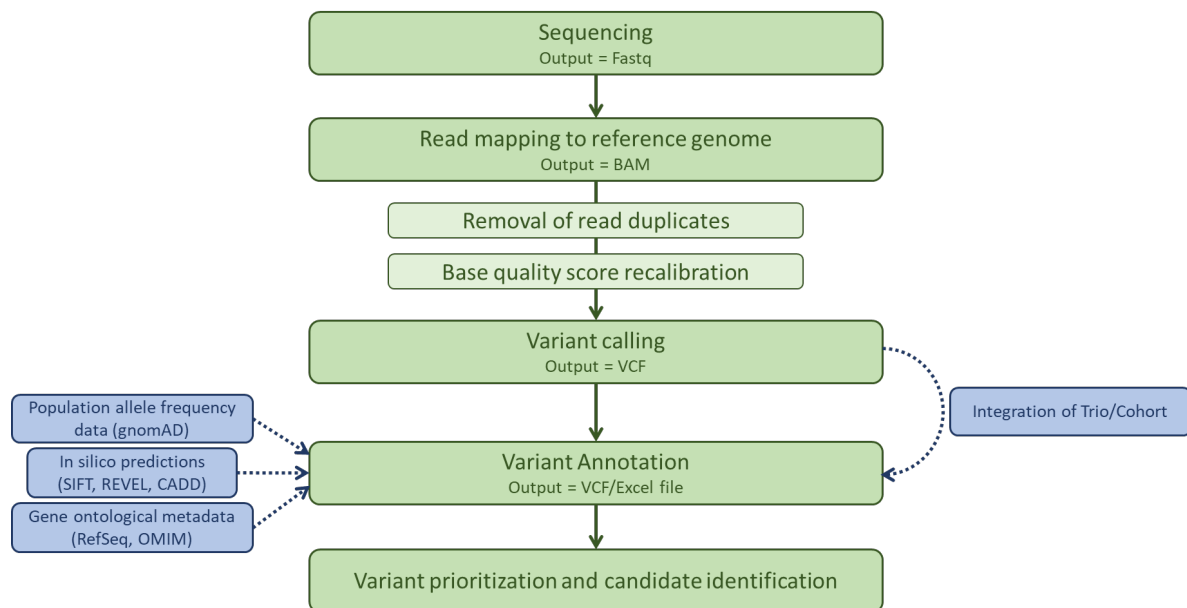


Figure 1.1: Overview of a standard WGS/WES bioinformatics workflow. Raw reads are assessed for quality and aligned to a reference genome to produce a Binary Alignment Map (BAM) file. Aligned reads in the BAM file can then be used to call variants and are exported to Variant Call Files (VCF), which allow annotation and filtering of variant data to identify variants of interest. Based on figure from GATK (<https://gatk.broadinstitute.org/hc/en-us>).

Multiple tools are available for annotation, including open-source tools such as Variant Effect Predictor (VEP) (McLaren et al., 2016) and Annotate Variation (ANNOVAR) (Yang and Wang, 2015), and licensed software such as VarSeq ([www.goldenhelix.com/products/VarSeq/](http://www.goldenhelix.com/products/VarSeq/)) and Alamut Visual (<https://www.sophiagenetics.com/platform/alamut-visual-plus/>). Annotation allows for genetic variants to be described within a biological and disease context. In order to narrow down the amount of putative diagnostic variants, filtering strategies are implemented based on annotation of genetic variants. Initially this includes annotation of basic gene ontological features including gene names and cDNA positions through catalogues such as ENSEMBL ([www.ensembl.org/index.html](http://www.ensembl.org/index.html)) or RefSeq ([www.ncbi.nlm.nih.gov/refseq/](http://www.ncbi.nlm.nih.gov/refseq/)) (Pabinger et al., 2014; Pruitt et al., 2007). Genetic variants can be annotated with further relevant metadata from additional databases such as the minor allele frequency (MAF) of the variant in the general population (section 1.2.4), *in silico* modelling of protein function using bioinformatic tools (section 1.2.5), and whether variants have previously been associated with disease or

specific phenotypes (section 1.2.6) (Butkiewicz and Bush, 2016). The overall annotation and filtering approach can be refined depending on factors such as the expected mode of transmission or incidence rate of disease, however standardisation across different rare disease patients allows for more consistent reporting of findings.

#### 1.2.4 Population databases

A key principle in identifying causative genetic variants in rare disease patients is that rare diseases result from genetic variants that are rare, or unique, in the wider healthy population. In order to determine if a variant is rare, it is useful to annotate genomic variants with allele frequency data from the general population. Absence or scarcity of a variant in a large control dataset supports its putative involvement in rare disease. Sequencing projects such as the 1000 genomes sequencing project allowed for discrimination of polymorphisms, or variants that were observed in more than 1% of the population (1000 Genomes Project Consortium et al., 2010). This approach was used in large clinical sequencing programmes such as the Deciphering Developmental Disorders (DDD) study to filter out polymorphisms, using a minor allele frequency cut off of 1 % (Wright et al., 2015). The sequencing of more and more exomes and genomes improves the use of allele frequency data as a metric in rare variant identification. This encouraged the use of low coverage sequencing approaches which sacrifice individual genotyping accuracy to allow for sequencing of more individuals, thus maximising determination of population allele frequency (1000 Genomes Project Consortium et al., 2012). Additionally, completion of the 1000 genome project saw mapping of genomes of 2,504 individuals from 26 populations, allowing for determination of polymorphisms unique to diverse population groups (1000 Genomes Project Consortium et al., 2015). Analysis of the finalised 1000 genomes project dataset identified that 1-4 % of variants in the genome of an average individual have a frequency in the population of < 0.5 %, representing between 40,000-200,000 rare variants per genome (1000 Genomes Project Consortium et al., 2015).

As more sources of data from control from control individuals became available, aggregation of data from different institutions was necessary to provide better determination of minor allele frequency reporting, particularly for minority ethnic

populations. The establishment of the exome aggregation consortium (ExAC) was an important development which allowed anonymised exome sequencing-derived genetic variants to be deposited in a central database freely accessible to researchers working to identify variants in individuals with rare mendelian disease (Lek et al., 2016). Exome sequencing data from 60,706 individuals enabled some key observations in the field of rare disease genetics. Objective metrics of pathogenicity could be assessed, and 3,230 genes were identified which could be characterised by depletion of predicted loss-of-function variants (Lek et al., 2016). Crucially, 72 % of these genes had no established involvement in human disease, opening up many avenues for research into mechanisms of rare disease (Lek et al., 2016). This supports previous findings that variants are more likely to be deleterious if a gene is depleted of functional variation in a control population, which is to say that genes that are more constrained for nonsense or missense variants are more likely to be mutated in disease (Samocha et al., 2014).

The most recent iteration of ExAC, which combines exome and whole genome data from dozens of sources is the genome aggregation database (gnomAD), which contains exome and genome data for more than 150,000 individuals unaffected by inherited disease (Karczewski et al., 2019). The scale of the data available through gnomAD allow highly accurate estimates of allele frequency, including for rare variants  $< 0.01\%$  (Gudmundsson et al., 2022). Using gnomAD-derived data is therefore useful when filtering genomic variants to identify rare SNVs. One caveat of the use of gnomAD data is possible inclusion of individuals with late-onset genetic diseases, or individuals with low penetrance of rare disease phenotypes (Rexach et al., 2019). Additionally, it is important to consider that rare variants may be benign as well as pathogenic, and so the use of allele frequency should only comprise part of a more sophisticated bioinformatic filtering pipeline.

Large-scale aggregation of genetic data has permitted a more accurate quantitation of selective pressure, allowing systematic identification of genes which are predicted to be intolerant to heterozygous loss of function (LoF) variants (Karczewski et al., 2020; Lek et al., 2016). This assumes that LoF variants will be depleted in disease-causing genes among a cohort of disease-free individuals. The pLI (probability of LoF intolerance) metric is an estimation derived from ExAC data of the probability that a gene is LoF-intolerant (Lek et al.,

2016). A pLI score > 0.9 indicates that LoF variants are observed in fewer than 10 % of cases than are expected based on the size of the gene (Lek et al., 2016). Utility of intolerance metrics are demonstrated in systematic gene-driven approaches to identify variants of interest in individuals with rare disease, such as identification of pathogenic variants in 22 histone lysine transferase genes with pLI >0.9 in DDD study participants with developmental disorders (Faundes et al., 2018). The pLI metric was refined using data from >140,000 exomes and genomes available in gnomAD to allow more reliable predictions for small genes, and to consider the variability in selective pressure that results from different types of variation on different genes (Karczewski et al., 2020). The loss-of-function observed/expected upper bound fraction (LOEUF) metric produces continuous estimates of LoF effect based on a confidence interval around the observed/expected ratio (Karczewski et al., 2020). The use of constraint metrics is an important tool that has evolved from aggregation of genomic data that can contribute to the refinement of putative pathogenic variants in individuals with rare disease.

#### 1.2.5 *In silico* modelling of genomic variants

The goal of bioinformatic filtering of variants is to identify variants which cause a deleterious change in protein expression or function. In order to assess the impact of genetic variation on protein function, bioinformatic tools have been developed to enable *in silico* prediction of deleterious effects (Peterson et al., 2013). Approaches have also been designed to indicate changes in splicing effects (see section 1.2.5.1). Various tools are available which use different approaches to predict conservation and pathogenicity of SNVs (Table 1.1) (Adzhubei et al., 2013; Desmet et al., 2009; Ioannidis et al., 2016; Jaganathan et al., 2019; Kircher et al., 2014; Reva et al., 2011; Schwarz et al., 2010; Shihab et al., 2015, 2014; Tavtigian et al., 2006; Vaser et al., 2016). These include FATHMM (Functional Analysis Through Hidden Markov Models; [fathmm.biocompute.org.uk/index.html](http://fathmm.biocompute.org.uk/index.html)) which discriminates between disease-causing and neutral polymorphisms based on sequence conservation using hidden Markov models (Shihab et al., 2014), and MutationTaster ([www.mutationtaster.org/](http://www.mutationtaster.org/)) which provides a probability of an SNV being pathogenic based on known SNVs derived from the human gene mutation database (HGMD) and 1000 Genomes Project (Schwarz et al., 2010). These are useful in providing an immediate view of

the potential impact of variants being considered during analysis of WES/WGS data. However, the disparity between functional prediction tools' ability to correctly predict effect has been remarked in various studies. One key criticism is the inadequate reflection of biological networks, including protein-protein interactions or relative functional importance of different proteins domains in predictions (Ganakammal and Alexov, 2019). As prediction tools are imperfect, and each assesses the impact of genetic variation differently, tools such as CADD and REVEL (see 1.2.5.2) have been developed in order to aggregate predictions from different predictors in order to offer a more reliable prediction.

Table 1.1: *In silico* tools for functional predictions of variant effect.

<b><i>In silico</i> prediction tool</b>	<b>Functionality</b>	<b>Output</b>	<b>Reference</b>
Align-GVGD (Align Grantham Variation and Grantham Deviation)	Predicts functional changes in amino acid substitutions based on Grantham scoring for biophysical properties of amino acids and multiple protein alignments.	Risk estimate score 0.00-4.00, divided into seven categories based on likeliness of mutation to interfere with function.	<a href="http://agvgd.hci.uta.h.edu/about.php">agvgd.hci.uta.h.edu/about.php</a> (Tavtigian et al., 2006)
SIFT 4G (Sorting Intolerant From Tolerant For Genomes)	Predicts consequence of amino acid substitution for protein function. Assesses protein conservation and severity of amino acid substitution.	Classification system (Tolerated or Deleterious)	<a href="http://sift.bii.a-star.edu.sg/">sift.bii.a-star.edu.sg/</a> ; (Vaser et al., 2016)
PolyPhen2	Assesses likely pathogenicity of missense variants via analysis of homology alignment, CpG context, protein structural changes in	Classification system (Unknown, benign, possibly damaging, probably damaging)	<a href="http://genetics.bwh.harvard.edu/pph2/">genetics.bwh.harvard.edu/pph2/</a> ; (Adzhubei et al., 2010)

	hydrophobicity and surface area		
MutationTaster	Assesses pathogenicity based on evolutionary conservation, changes in splice sites, mRNA and protein structure and function	Classification system (Disease causing, Disease causing automatic, polymorphism, polymorphism automatic).	<a href="http://www.mutationaltaster.org/">www.mutationaltaster.org/</a> ; (Schwarz et al., 2010)
MutationAssessor	Assesses evolutionary conservation of amino acids	Functional impact score with scores >1.935 classed as pathogenic.	<a href="http://mutationassessor.org/r3/">mutationassessor.org/r3/</a> ; (Reva et al., 2011)
FATHMM (Functional Analysis Through Hidden Markov Models)	Uses Hidden Markov Models to evaluate a variant using disease-specific weighting. Can choose from inherited disease, cancer or disease-specific analyses.	Classification system (Pathogenic, or Benign).	<a href="http://fathmm.biocompute.org.uk">fathmm.biocompute.org.uk</a> ; (Shihab et al., 2014)
FATHMM-MKL (Multiple Kernel Learning)	As for FATHMM but integrates functional annotations from kernel matrices including databases such as ENCODE to make predictions for coding and non-coding variants	Classification system (Pathogenic, or Benign) alongside a confidence measure.	<a href="http://fathmm.biocompute.org.uk/fathmmMKL.htm">fathmm.biocompute.org.uk/fathmmMKL.htm</a> ; (Shihab et al., 2015)



Human Splicing Finder (HSF)	Predicts changes to splicing signals using previously generated stand-alone matrices to analyse 5' and 3' splice sites, branchpoints, and cis-acting exon enhancers and silencers (ESE and ESS).	Probability of splicing being altered, ESE and ESS predictions integrated into single pathogenicity prediction score, and graphical output.	<a href="http://www.umd.be/HSF/">www.umd.be/HSF/</a> ; (Desmet et al., 2018)
SpliceAI	Assesses likelihood of a variant leading to formation of splice donor or acceptor site, or disruption of an existing splice site. Applicable to non-coding variants as well as coding variants and canonical splice sites.	Delta-score of likelihood of effect, with reference effect on local splice sites.	<a href="http://spliceailookup.broadinstitute.org/">spliceailookup.broadinstitute.org/</a> (Jaganathan et al., 2019)
REVEL (Rare Exome Variant Ensembl Learner)	Integration of 18 existing tools, trained specifically on rare pathogenic and rare benign coding variants.	Scores range on a scale of 0-1 with 1.	<a href="https://sites.google.com/site/revelgenomics/">sites.google.com/site/revelgenomics/</a> (Ioannidis et al., 2016)
CADD (Combined Annotation-Dependent Depletion)	Assesses pathogenicity based on scores from a combination of <i>in silico</i> tools which include evolutionary conservation, allelic diversity, variant annotation, splice score perdition, functional data.	Phred-scale (Log10-based score) pathogenicity ranking based on all possible SNVs in the whole genome.	<a href="http://cadd.gs.washington.edu/">cadd.gs.washington.edu/</a> ; (Kircher et al., 2014)

#### 1.2.5.1 Splicing models

There is similar availability of bioinformatic tools to assess the impact of variants on messenger RNA (mRNA) splicing. Changes in splicing that result in disease include disruption of canonical donor and acceptor splice sites, branch sites, cis-acting exonic splicing enhancers (ESE) and silencers (ESS), and introduction of novel splice sites within intron or exon regions (Cormier et al., 2022; Leman et al., 2020; Ohno et al., 2018). Numerous diseases have been associated with splicing disruption (Anna and Monika, 2018; Cunha et al., 2016; Li et al., 2018; Parthasarathy et al., 2022), and it is estimated that 9-11 % of pathogenic variant in patients with rare disease result from splicing changes (Jaganathan et al., 2019). Currently, splicing variants only make up 8.5 % of the disease-causing genes catalogued by HMGD (<https://www.hgmd.cf.ac.uk/ac/index.php> as of 24<sup>th</sup> November 2022), which suggests that although variants which affect splicing are a significant driver of genetic disease, more effort is required to ensure adequate assessment and detection of pathogenic splicing variants.

Largely consequential following its development, Human Splicing Finder (HSF) integrated analysis of motifs evaluated by stand-alone tools including branch points, and cis-acting elements, and generating easily interpretable predictions (Desmet et al., 2009). More recently, deep learning approaches have greatly influenced development of latest tools (Rowlands et al., 2019). One additional widely used example is SpliceAI, which utilises a deep neural network to predict splice site junctions and allow accurate prediction of splicing probability arising from non-coding variants (Jaganathan et al., 2019). Validation of this model using RNA-sequencing led to the finding that many of the rare non-coding splicing variants identified are highly deleterious, often leading to nonsense protein products. Additionally, more recent approaches that consider the constraint of non-coding regions to splicing variants can also be helpful when prioritising variants predicted to affect splicing (Cormier et al., 2022). These examples illustrate how ability to sensitively predict splicing changes in non-coding regions is improving at great speed, and therefore potentially allows a major uplift in diagnostic yields.

#### 1.2.5.2 Aggregating *in silico* predictions

While *in silico* tools offer immediate information as to the likelihood of a variant being deleterious, few combine data which reflect evidence from biological assays. As a result, many benign variants are often erroneously classed a disease-causing (false positive). One approach taken to ensure that fewer false positive variants are identified is by using tools which aggregate predictions from separate prediction tools. One such example is the development of CADD (Combined Annotation-Dependent Depletion) scoring (Kircher et al., 2014), which allowed ranking pathogenicity of all possible SNVs genome-wide by integrating data from >60 annotation tools including Ensembl Variant Effect Predictor (McLaren et al., 2016), epigenetic information from the ENCODE Consortium (ENCODE Project Consortium, 2012), Grantham Distance scoring (Grantham, 1974), SIFT/Polyphen2 scoring (Vaser et al., 2016), and splicing predictions (Rentzsch et al., 2021). CADD derives a single score using a machine learning model that is transformed on to a PHRED-like scale, producing a single meaningful output (Table 1.2). Although there is no set or recommended threshold for a variant to be classed as pathogenic, it is broadly accepted among users that a CADD-Phred score < 20 is unlikely to be pathogenic, and so can be regarded as benign (Rentzsch et al., 2019). While this arbitrary cut off is useful when analysing multiple variants from WGS/WES data, variants falling below this threshold should not be entirely ignored. For example, sensitivity is using a CADD score of 20 to predict pathogenicity declined to 75 % when non-coding variants more than four nucleotides from canonical splice sites are evaluated in patients with congenital hypothyroidism (Albader et al., 2022).

Table 1.2: Interpretation of CADD-PHRED scores. Scores rank all possible genomic variants by predicted pathogenicity, from (Kircher et al., 2014).

CADD score	Variant pathogenicity ranking
>10	Bottom 90%
10-20	Top 10%
20-30	Top 1%
30-40	Top 0.1%
40-50	Top 0.01%
>50	Top 0.001%

Development of REVEL (Rare Exome Variant Ensembl Learner) also represents a key development in the prediction of rare-disease specific SNVs (Ioannidis et al., 2016). This approach integrates predictions from distinct tools as achieved by CADD, but random forest algorithms are trained using newly identified rare pathogenic and rare benign variants. This addresses the overlooking of rare variants, and the predisposition of stand-alone tools to inaccurately mark ultra-rare or novel variants as pathogenic due to their absence from databases such as the 1000 genomes project (Li et al., 2013). Unlike CADD however, REVEL does not incorporate splicing predictions, and does not assess non-coding variants. The advantages and drawbacks of the *in silico* tools available therefore requires the use of multiple different methods when attempting to predict the effect of different types of rare variants identified in NGS data (Butkiewicz and Bush, 2016; Richards et al., 2015).

#### 1.2.6 Functional and disease annotation

In order to improve annotation and interpretation of rare, putatively damaging genetic variants identified from NGS approaches, it can be useful to consider databases of variant-, gene-, or phenotype-specific functional data. As most ontological and functional terms tend to be verbose, they permit the incorporation of experimental data into the consideration stage of variants post-filtering (Ejigu and Jung, 2020). Catalogues of human variants such as

HGMD (Stenson et al., 2012) or the Single Nucleotide Polymorphism database (dbSNP) (Kitts and Sherry, 2011) provide variant-level data on phenotype associations (Pabinger et al., 2014). Meanwhile molecular and cellular functional data can be integrated using the Kyoto Encyclopaedia of Gene and Genomes (KEGG) pathway database and gene ontology (GO) terms (Kanehisa et al., 2016). Compiled by the GO Consortium, GO terms provide a description of the molecular function, cellular component, and biological processes associated with a particular gene product, based on experimental data (Gene Ontology Consortium, 2015). Data regarding gene expression across different human tissues and cell types, provided by Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/>) can also be useful when linking a variant in a poorly characterised gene to expression within an affected tissue (Aguet et al., 2017).

To provide better context of human phenotypes, databases such as Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2009), ClinVar (Landrum et al., 2014), and the Genome Wide Association (GWAS) catalog (Hindorff et al., 2009) provide a catalogue of disease-gene relationships. ClinVar annotation also annotates information regarding the pathogenicity or non-pathogenicity of variants that are present in the database, allowing for previously characterised rare pathogenic or benign variants to be flagged. OMIM is a curated database which describes over 16,000 genes, 6,514 mendelian phenotypes with known molecular basis, and an additional 1,510 mendelian phenotypes which are yet to be associated with a molecular basis (as of 23<sup>rd</sup> November 2022, <https://www.omim.org/>). Phenotype and genomic data available through OMIM can be useful in establishing if a gene has previously been associated with a phenotype.

Finally, phenotypes identified in model organisms can add a further layer of insight, such as The Mouse Genome Informatics (MGI) database (<http://www.informatics.jax.org/>) which integrates genomic, gene expression, and phenotype data relating to wild-type and genetically engineered mice. Animal model databases can be particularly useful when identifying novel gene-phenotype associations not previously observed in humans.

### 1.2.7 Classifying variants

In order for variant pathogenicity to be consistently reported in a clinical setting, a standard variant classification system and terminology is necessary (MacArthur et al., 2014). This allows for genetic testing and molecular diagnoses to support clinical decision making. The first use of a classification system in the interpretation of genetic tests was in predicting cancer susceptibility (Plon et al., 2008). In 2015, the American College of Medical Genetics and Genomics (ACMG) produced guidelines for classification of genetic variants involved in Mendelian disease into five categories: Pathogenic, Likely Pathogenic, Uncertain Significance, Likely Benign, and Benign (Richards et al., 2015). Classification was based on multiple criteria which provide support for either a pathogenic or benign classification. This system considers the strength of variant-specific evidence including variant population frequency, presence in disease databases, computational *in silico* modelling, functional data, and family segregation data (Richards et al., 2015). Criteria and evidence strength are summarised in Table 1.3.

In cases where criteria for pathogenic classification are not met, or where criteria supporting benign or pathogenic classification is contradictory, a variant is classed as a variant of unknown significance (VUS). In such cases, additional evidence is required to make a classification. For example, if one affected family is identified, identification of further unrelated individuals can support a pathogenic classification. Furthermore, it is often necessary to undertake new functional investigations to provide greater molecular evidence of variant pathogenicity. Further investigations might include more detailed phenotyping, imaging studies, *in vitro* molecular assays, or modelling of disease using an animal model.

Table 1.3: Overview of criteria for variant pathogenicity classification. Guidelines as recommended by ACMG and the Association of Molecular Pathology, Richards et al., (2015).

Evidence strength	Strength	Criterion	Description
Pathogenic	Very strong	PVS1	Null variants (nonsense, frameshift, splice site) where LOF is known cause of disease.
	Strong	PS1	Amino acid change observed previously in disease.
		PS2	<i>De novo</i> variant in patient with no family history of disease.
		PS3	<i>In vitro</i> or <i>in vivo</i> functional data support damaging effect on gene or gene product.
		PS4	Prevalence of variant in affected individuals is significantly increase compared to unaffected family members.
	Moderate	PM1	Located in mutational hotspot and/or critical and well-established functional domain without benign variation.
		PM2	Absent from controls.
		PM3	For recessive disorders, detected in trans with pathogenic variant.
		PM4	Protein length changes due to in-frame deletions/insertions in non-repeat region or stop-loss variants.
		PM5	Novel missense change at AA residue where different change has been determined to be pathogenic.
		PM6	Assumed <i>de novo</i> , but without confirmation of paternity or maternity.

	Supporting	PP1	Co-segregation with disease in multiple affected members of the family, where the gene has previously been associated with the disease.
		PP2	Missense variant in constrained gene, where missense variants are a known mechanism of disease.
		PP3	Prediction of deleterious effect among multiple <i>in silico</i> computational predictive models (conservation, evolutionary, splicing impact).
		PP4	Phenotype is highly specific for a disease with a monogenic aetiology.
		PP5	Reported as pathogenic, but functional data is not available to perform independent evaluation.
Benign	Supporting	BP1	Missense variant in a gene where truncating variants are known mechanism of disease.
		BP2	Observed in trans with pathogenic variant in a fully penetrant dominant disorder; observed in cis with pathogenic variant in any inheritance pattern.
		BP3	Inframe indel in repetitive region without a known function.
		BP4	Prediction of benign effect among multiple <i>in silico</i> computational predictive models (conservation, evolutionary, splicing impact).
		BP5	Variant identified in case with alternative molecular basis of disease.
		BP6	Reported as benign, but evidence is not available to perform independent evaluation.



		BP7	Synonymous variant which is not predicted to have an effect on canonical splicing.
	Strong	BS1	Allele frequency is greater than expected for disease.
		BS2	Observed in health adult individual.
		BS3	<i>In vitro</i> or <i>in vivo</i> functional data support benign effect on gene or gene product.
		BS4	No segregation in affected members in a family.
	Very strong	BA1	Polymorphism, allele frequency >5 %.

#### 1.2.8 Analysis of structural variants

Analysis of structural variants (SVs) is another form of variation in the human genome, and a crucial aspect when assessing patients with rare genetic diseases (Sanchis-Juan et al., 2018). Classically, structural variants (SVs) could be detected using targeted assays including fluorescent *in situ* hybridisation (FISH), and array comparative genomic hybridisation (CGH) (Kannan and Zilfalil, 2009). The advent of NGS technologies and particularly WGS offer a new approach to identify SVs, which include large (> 50 bp) insertions and deletions, copy number variant (CNV) duplications and deletions, and genomic rearrangements (inversions, translocations) (Alkan et al., 2011). Different options are available for calling SVs, such as generating *de novo* genome assemblies (Iqbal et al., 2012), assessing differences in read depth (Plagnol et al., 2012), and analysing alignment of read pairs (Jiang et al., 2012). These techniques are optimised for using WGS data due to the comprehensive unbiased coverage. Determination of some structural variants such as CNVs can be accomplished using WES by identifying exonic regions that have higher or lower depth of coverage relative to other sequenced regions (Ellingford et al., 2017; Krumm et al., 2012).

Just as for SNV data, annotation of SV data is essential when considering pathogenicity. Many of the population databases discussed in section 1.2.4 also contain data relating to SVs identified in the population, allowing filtering of rare SVs. The Database of Genomic

Variants (DGV) is one SV-specific catalogue of SVs from > 75 studies (Ejigu and Jung, 2020; MacDonald et al., 2014). Databases such as ClinVar and DECIPHER (Firth et al., 2009) contain SVs identified in rare disease patients. Functional annotation of SVs is more complicated compared to SNVs. Specific stand-alone bioinformatic tools can be used for SV-specific data annotation, including defining genes affected by the SV. Implementation of SV analysis remains an area of active research, with new pipelines and analysis tools continuing to emerge. The challenges in SV detection are further described in section 1.3.2.1.

#### 1.2.9 Current clinical approaches

In the last decade, numerous large scale clinical NGS initiatives have been initiated in order to harness the ability of sequencing for benefit of patients with rare disease. Some key advances in the integration of NGS into clinical diagnostic services within the United Kingdom are described. Common developments from the DDD study and the 100,000 genomes project have been the utility in standardisation of approaches across multiple diagnostic centres. These include the use of Human Phenotype Ontology (HPO) terms (Köhler et al., 2014) to accurately and reliably describe rare diseases identified in different referral centres.

#### DDD

The Deciphering Developmental Disorders (DDD) study began recruiting patients with rare diseases in 2011, focusing on patients with neurodevelopmental disorders, congenital anomalies, and dysmorphic features. In the DDD study, recruitment of patient-parent trios was a key approach to maximise the chances of pathogenic variant identification (Firth et al., 2011). Using this approach, streamlining and standardisation of bioinformatic approaches was necessary. In order to narrow down the number of putatively diagnostic variants, a filtering strategy is utilised: variants with a minor allele frequency > 1%, and variants not predicted to be protein-altering were filtered out (Wright et al., 2015).

During the study, the release of a large dataset of genetic data for individuals with developmental phenotypes allowed new approaches to identify novel genes and improve diagnostic rates. Using a genotype-driven approach to identify novel genes in 1,133

participants with complex phenotypes, novel variants from 35 participants were identified in 12 genes previously associated with developmental disorders, allowing a rise in the diagnostic rate of the whole DDD study from 28 % to 31 % (Deciphering Developmental Disorders Study, 2015). More recent publication of genetic findings indicated an increase in diagnostic rate to 41 % (Wright et al., 2023). Factors which contributed to increased likelihood of a diagnosis being reported included recruitment of family trios, which was associated with lower amounts of candidate variants compared to patients recruited alone (Wright et al., 2023). This exemplifies the value of recruiting family trios in the detection of *de novo* autosomal dominant disorders. Anonymised genetic and phenotype data from the DDD study is available through the DECIPHER portal, allowing researchers to support or discount novel diagnoses. This will ensure that the diagnostic yield of participants in the DDD study continues to rise as new analysis pipelines are established.

#### 100,000 Genomes Project

Established in 2013, the 100,000 Genomes Project (100kGP) aimed to improve clinical diagnostic detection rates in rare diseases and cancer using WGS (100,000 Genomes Project Pilot Investigators et al., 2021). This has been possible by the reduction in cost of genome sequencing from £2 billion for the initial genome sequence to less than £ 500 in the last 20 years (Turnbull et al., 2018). The project includes tracking and integration of phenotype data to improve variant interpretation. In the long-term, accelerated molecular diagnosis through 100kGP enables delivery of targeted treatments, enables reproductive planning, offers opportunities for carrier testing, and more appropriate genetic counselling of affected individuals and families (Griffin et al., 2017; Snape et al., 2019).

In order to streamline interpretation of clinically relevant genetic variants, a tiering approach is used (100,000 Genomes Project Pilot Investigators et al., 2021). Initially, variants are selected for the tiering system if that variant is: rare in the healthy population, is protein altering in at least one transcript, follows the relevant mode of inheritance, and segregates appropriately in the family. If the variant is not in a gene listed on an applies virtual disease gene panel, the variant is placed in Tier 3, and is not usually reviewed for clinical interpretation. Of the remaining variants in genes pre-specified by virtual panels, protein altering variants are placed in Tier 2, and *de novo* and truncating variants and

previously reported pathogenic variants are placed in Tier 1. The diagnostic rate is 22 % from variants placed in tier 1 and 2 (Turnbull et al., 2018). Although Tier 3 variants are not considered during clinical interpretation, availability of de-identified data through the Genomics England virtual research environment for research use (Turnbull et al., 2018). This allows independent assessment in research settings of Tier 3 variants, and variants which were not selected for the tiering system, such as non-coding variants. This approach is particularly valuable to identify novel mechanisms of disease. Review of the 100kGP pilot study which utilised this interpretation pipeline revealed that 14 % of molecular diagnoses made resulted from a combination of research and automated processes, and that research involvement was particularly critical in identification of non-coding, structural, and mitochondrial variants (100,000 Genomes Project Pilot Investigators et al., 2021).

## 1.3 Challenges in rare disease genomics

### 1.3.1 Missing heritability

The concept of “missing heritability” was described over a decade ago by Manolio et al., (2009) and has come to represent the major ongoing challenge faced in rare disease diagnostics. In genetics, heritability is defined as the proportion of a disease that results from genetic variants. By definition, missing heritability therefore is the proportion of disease phenotype which is not attributable to detectable genetic variants or environmental factors. The missing heritability can be addressed by improvements in the ability to effectively (i) detect and (ii) interpret genetic variants.

Variants that have limited detectability include structural variants, non-coding variants and changes in epigenetic signatures such as DNA methylation in regions of the genome which are poorly characterised. Additionally, the cumulative influence of different genetic variants with variable effect sizes in several genes can contribute to penetrance of rare disease phenotypes in a non-monogenic disease model (Manolio et al., 2009). Diagnosis of rare monogenic diseases where there is a direct gene-disease relationship assumes that variants with large effect sizes result in high phenotypic penetrance (Manolio et al., 2009). However, the aetiology of more common diseases such as coronary heart disease or schizophrenia derives from a combination of common variants with low effect sizes and environmental factors, each of which impart an increased risk of developing disease. Identifying uncommon variants (MAF > 1 %) which segregate with disease is possible, but requires large cohorts of affected and unaffected individuals to provide adequate statistical power (Neumeyer et al., 2019; Wellcome Trust Case Control Consortium, 2007).

Aside from challenges in variant detection, variants for which interpretation is limited are commonly termed as variants of unknown significance (VUS) and occur within both characterised and new genes. Variants which are difficult to interpret can also be considered as a consequence of limited detectability of poorly characterised mutational mechanisms, such as non-coding variants which result in regulatory changes. These can result in non-mendelian inheritance patterns, adding an additional layer of complexity. These may cause changes in epigenetic status such as DNA methylation or histone

modifications (Zhang and Pradhan, 2014), and effect expression of known or yet uncharacterised genes.

### 1.3.2 Challenges in detecting poorly characterised genetic mechanisms in disease

#### 1.3.2.1 Structural variants

While WGS offers the ability to acquire an individual's entire genome, different sequencing chemistries need to be considered when assessing SVs compared to SNVs. Short read sequencing is optimal for SNV detections due to high accuracy of base calling, and while the use of paired short reads also allows SVs to be detected, effective calling and mapping of large and complex structural variants from short reads (often around 200 base pairs) is limited (Mitsuhashi and Matsumoto, 2020). Long read sequencing enables complex SVs, such as translocations and inversions to be identified and for breakpoints to be established, however, base calling accuracy is reduced compared to short sequencing alternatives (Eichler, 2019; Rang et al., 2018; Weirather et al., 2017). Long read sequencing approaches are therefore used when there is a priori hypothesis of complex SV involvement, usually in individuals with complex or heterogeneous disorders (Jenko Bizjan et al., 2020; Pauper et al., 2021).

In addition to effective SV detection from a technological viewpoint, it is also important to consider challenges in detecting variants which are likely to be disease-causing. Tools are continually being developed and optimised to ascertain and classify variants for interpretation. One such tool, ClinSV, uses metrics based on combinations of discordancy mapping of read pairs, split-read mapping, and changes in read depth coverage to identify SVs to overcome some of the disadvantages inherent to short read data (Minoche et al., 2019). This approach has been used to identify rare SVs in various rare disease cohorts including familial dilated cardiomyopathy (Minoche et al., 2019), and dystonia (Kumar et al., 2019). ClassifyCNV (Gurbich and Ilinsky, 2020) is a similar CNV-specific tool with extended functionality to annotate and classify CNVs according to a recent extension to the ACMG framework (Riggs et al., 2020). In order to implement the ACMG classification criteria, ClassifyCNV uses a scoring system with variants receiving higher pathogenic scores for specific attributes such as the inclusion of > 35 protein-coding genes, the presence of

dosage-sensitive genes, or low population frequency (< 1 %) (Gurbich and Ilinsky, 2020). This tool has provided a useful means for ensuring that SVs that are detected are novel and rare, and therefore of interest when considering interpreting the effect of such SVs.

In terms of SV interpretation, assessment of Topologically Associated Domains (TADs) can provide important regulatory context during SV analysis. TADs are functional units of self-interacting chromatin, often containing groups of co-regulated genes (McArthur and Capra, 2021). Disruption of boundaries between TADs as a result of SVs can therefore disrupt gene expression and is observed in rare disease patients (de Bruijn et al., 2020). ClinTAD is a bioinformatic tool for assessing the impact of CNVs on TADs and was successfully implemented to identify deletion of a TAD boundary between two cardiac-related genes MCTP2 and NR2F2 (MIM: 107773) in a patient with complex cardiac anomalies (Spector and Wiita, 2019). Improvements in ability to detect and interpret SVs continue at great pace and provide great promise in improving the evaluation of SVs, particularly for rare disorders that are complex in nature.

#### 1.3.2.2 Non-coding variants

While SV detection is limited by methodological approach use for WGS, one crucial benefit of WGS compared to WES is access to discerning variants in non-coding regions of the genome. However, due to the extent of non-coding variant present in each individual's genome, determining which are relevant when considering a molecular diagnosis is enormously challenging. Firstly, non-coding variants in introns and intergenic regions have the potential to induce novel splicing acceptors and donors, potentially disrupting processing of transcribed genes (Bauwens et al., 2019). Tools which enable improved in silico prediction of non-coding variants continue to emerge (Sullivan et al., 2023).

Secondly, variants in non-coding regulatory elements of the genome such as enhancers (Green et al., 2021) and promoters (Wood et al., 2022) have also been associated with rare diseases. SNVs within regulatory regions may impair binding of regulatory machinery or alter the binding of transcription factors. For example, non-coding variants in non-coding regulatory elements upstream of *TBX3* (MIM 601621) have been associated with cardiac conduction (van Weerd et al., 2020). Deletion of the associated non-coding regulatory element in a mouse model resulted in increased expression of *Tbx3* and dysfunctional

cardiac conduction (van Weerd et al., 2020). In order to successfully considering non-coding variants, tissue specific databases of gene expression, DNA modification are necessary. Additionally, targeting regions around known genes can be a useful method of reducing overall variants for consideration. Further integration and adoption of these methods will identify further non-coding variants which are associated with rare disease.

#### 1.3.2.3 DNA methylation

In addition to differential gene expression resulting from non-coding variants, DNA methylation signatures are an additional heritable element which might explain missing heritability (Guimarães et al., 2022; Moore et al., 2013). However, methylation of CpG dinucleotides, associated with transcriptional repression, cannot be determined using standard WGS (Zhang and Pradhan, 2014). Genomic methylation profiling relies on bisulfite sequencing, where unmethylated cytosine bases are converted to uracil followed by DNA sequencing (Adusumalli et al., 2015). Bisulfite sequencing of patient DNA is therefore key in detecting changes in methylation patterns which may underly disease segregation and indicate the influence of environmental factors in missing heritability of disease (Adusumalli et al., 2015; Xu et al., 2022).

Combining genotype data with bisulfite sequencing, as well as other functional genomic tools is challenging in terms of data processing, but an area of interest when considering the challenge of variant interpretation. The challenge of including epigenetic changes can be useful when considering potential mechanisms of variable disease penetrance within families affected by often complex phenotypes (Herlin et al., 2019). One such example is the association between occurrence of congenital heart disease in children affected by Down's syndrome and maternal methylation status of the promoter of the Methylenetetrahydrofolate Reductase (*MTHFR*; MIM 607093) gene (Asim et al., 2017). This illustrates how determining complex phenotype heredity can be complex and require a combination of different functional genomic approaches. The added complexity in capturing modifications to DNA bases, rather than changes to the DNA sequence holds back the integration of genomic DNA methylation mapping into rare disease diagnostic pipelines.



#### 1.3.2.4 Variants in non-monogenic disease

The missing heritability hypothesis suggests that rare diseases may also derive from a combination of genetic variants with variable effect sizes and environmental factors (Manolio et al., 2009). Sometimes termed a polygenic disease model, this concept suggests that inheritance of a few uncommon, but not rare, variants with intermediate effect sizes may contribute to rare disease in an additive manner. This can lead to non-Mendelian ratios of affected individuals within families making identification and segregation of such variants in affected families by family association methods challenging (McCarthy et al., 2008). In addition, the uncommon nature of such variants (MAF < 2 %), and the low numbers of affected individuals means that cohorts cannot be constructed for GWAS approaches (Visscher et al., 2017). For example, polycystic ovary syndrome (PCOS) has a complex model of heritability (Dapas and Dunaif, 2020), and several GWAS have identified > 20 susceptibility loci. Analysis for gene-level associations of rare variants from aggregated WGS data from 261 affected individuals supported a model of rare variant inheritance in genes regulating relevant disease pathways such as *DENND1A*, *C9ORF3*, and *BMP6* (MIMs: 613633, 619600, 112266) (Dapas and Dunaif, 2020). Some of the rare variants from the *DENND1A* locus were predicted to disrupt transcription factor binding or RNA-binding protein motifs, suggesting a regulatory effect (Dapas and Dunaif, 2020). This highlights the importance of both conducting WGS, and of acquiring cohorts of rare disease patients to conduct statistical analyses where possible.

Another example is that of an oligogenic model of inheritance. This is an intermediate between polygenic diseases, where many variants with small effect sizes contribute to disease, and monogenic diseases. In a family affected by left ventricular non-compaction (LVNC), three affected offspring were found to have inherited three heterozygous variants, two from the father (in genes *MKL2* (MIM: 609463) and *MYH7* (MIM: 160760)) and one from the mother (in gene *NKX2-5* (MIM: 600584)) (Gifford et al., 2019). This pattern, confirmed to be disease causing via CRISPR-Cas9-generated murine models, suggests that the father has asymptomatic LVNC, and that the variant inherited from the mother acts as a modifier, resulting in a highly penetrant form of the disease (Gifford et al., 2019). These examples describe inheritance patterns which are more likely to be observed in ultra-rare disease settings, and although it cannot be assumed that complex forms of segregation may

be involved in every rare disease case, it is important to retain the possibility of such underlying mechanisms.

A hypothesis of oligogenic interaction is seldom suitable during initial analysis of genomic data. In order to determine pathogenicity relating to a combination of variants from two or genes, large cohorts of unresolved cases are required to provide sufficient statistical power to enable a robust diagnosis (Sackton and Hartl, 2016; Singh, 2021).

### 1.3.3 Challenges in interpreting variants of unknown significance

#### 1.3.3.1 Variants of Unknown Significance

One key ongoing challenge in medical genetics is the question of variants of uncertain significance regardless of whether a gene has previously been associated with disease. The Richards et al., (2015) ACMG framework argues that variants with weak or contradictory evidence should be classified as VUS. Labelling a variant as VUS means that it cannot be reported to patients or families as a causative variant. Often VUS-classified variants are ignored due to inability of interpretation. In order to make more diagnoses, it is important to consider VUS where there is no adequate molecular diagnosis. There are multiple factors which could improve the resolution of VUS such as improving classification of non-coding and structural variants, removing biases from current datasets, conducting functional evaluation of VUS, and intermittent reassessment of genomic variants from individuals without a molecular diagnosis.

Standardisation of reporting and classification systems is a crucial ongoing development (DiStefano et al., 2022). As outlined in section 1.2.7 and 1.2.8, guidelines have been published for interpretation coding SNVs and CNVs. Evaluation of CNVs is complex, and the authors of the classification framework recommend a nuanced approach to CNV evaluation, given the potential complexity of the gene regulatory nature of CNVs (Riggs et al., 2020). The ability to functionally characterise the effect of CNVs and SVs more generally, such as use of Hi-C mapping of chromatin interactions, is allowing for more SVs to be resolved, but often implications difficult to interpret (Shinkai et al., 2020).

The current classification approach is limited by inherent biases of the databases and metrics used to assess variant pathogenicity. One example of an annotation bias occurs with gnomAD, where there is potential inclusion of variants from individuals affected by rare disease but with reduced phenotypic penetrance, or with late-onset genetic disease (Gudmundsson et al., 2022). Additionally, there is a disparity in acquisition of genetic sequencing data from non-European populations. This may allow rare variants to be overlooked in patients from other ethnicities (Austin et al., 2018; Gudmundsson et al., 2022).

One further source of VUS misclassification is the contradictions in pathogenicity predictions from different *in silico* tools, and the over reliance on such algorithms (Ernst et al., 2018). By modelling the pathogenicity of *CFTR* variants in cases of cystic fibrosis, Pereira et al. showed the low specificity and positive predictive value of computational algorithms are sometimes inadequate, due to their inability to reflect the complexity of biological networks (Pereira et al., 2019). The use of neural networks which combine individual bioinformatic approaches do go some way to bridging this gap, however, the challenge remains to integrate gene-specific epidemiological data that can facilitate a more comprehensive assessment of idiosyncrasies in disease mechanisms.

#### 1.3.3.2 Functional evaluation

Where functional evidence of pathogenicity is unavailable, such as when investigating novel phenotypes or a characterising a ‘new’ gene previously not associated with disease, it is necessary to generate functional data in order to support a pathogenic variant classification (Gando et al., 2019). Functional evaluation can take different forms, and multiple techniques and approaches are available depending on the context of the genetic variant, the hypothesised molecular mechanism of disease, and access to patient tissues (Table 1.4).

Table 1.4: Examples of models available for evaluation of genetic variant functional effect.

Model	Use/Example	Advantages	Disadvantages
<i>In silico</i> protein modelling	Identifying potential changes in functional domains within a protein structure	Low cost. Can indicate if a variant is likely to affect structure of key functional domains, or interactions with other proteins/molecules.	Can be computationally intensive. Requires availability of published crystal structure.
<i>In vitro</i> assays	qPCR, enzymatic reporter assays, protein-protein binding.	Low cost. Quick to establish. Highly reproducible.	Low physiological context.
Cell lines	Viability, splicing assays, protein/transcription factor binding to DNA, localisation.	Low cost. Can assess dynamic changes in eg localisation, expression. No requirement for patient material. Use of patient-derived cells allows for assaying of variant impact in appropriate genomic context.	RNA/protein processing may be different to human/mammalian depending on model selected. Invasive if using patient-derived cells. Can be time intensive.
Model organism	Characterisation of tissue-level or systemic phenotype	Phenotypes are more relevant Allows assessment of phenotypes which cannot be confirmed in cell models e.g., neurological/ behaviour.	Cost. Time to establish. Ethical approval required.

Along with data processing and storage, functional evaluation represents a further bottleneck, largely due to time and resource constraints. In order to identify VUS which are good candidates for further study, linking patients acquired from different centres can be useful. Platforms such as GeneMatcher ([genematcher.org/](http://genematcher.org/)), MIMmatch (Amberger et al., 2009), or Matchmaker Exchange ([www.matchmakerexchange.org/](http://www.matchmakerexchange.org/)) can be invaluable in linking clinical teams which each have patients with diseases so rare that there are only a handful previously described (Philippakis et al., 2015). This system is also beneficial for describing novel rare diseases and accelerating the delineation of new disease-gene associations.

#### 1.3.3.3 Reassessment of unresolved data

A further consequence of undertaking functional evaluation of VUS is the ability to adjust variant classifications in further affected individuals. In this way, monitoring change in classification can also be useful when conducting reanalysis of genetic variants because as evidence evolves, application or disapplication of supporting criteria may change (Richards et al., 2015). For example, greater availability of sequencing data among controls in databases such as gnomAD allows variants at one time classified as VUS to be reclassified as 'benign' based on their greater frequency in a healthy population, compared to the occurrence rates of the disease (Yauy et al., 2022). Interestingly, Tavtigian et al., (2018) identified that the ACMG classification system was highly Bayesian in its approach to variant interpretation. This implies that the current primarily subjective classification framework is adequate in classifying variants, and that it can evolve to adopt a quantitative Bayesian basis, which has the benefits of reduced subjectivity and improved tracking of variant classification and reclassification over time (Tavtigian et al., 2018).

Tracking of changes in ClinVar classification over an 18-month period identified 1,247 changes in variant classification (Yauy et al., 2022). This underlines this importance of continuing to generate new functional data for gene-disease associations and the usefulness of intermittent reanalysis of unsolved cases. Identification of a molecular cause of disease in unsolved rare disease cases is the aim of the Solve-RD project (<http://solve-rd.eu/>). Solve-RD is a pan-European consortium of clinicians, geneticists, research and diagnostic

infrastructures, patient organisations, and leading experts in the fields of bioinformatics and data management (Matalonga et al., 2021). This approach exemplifies the need and benefit for patients of reanalysis and identification of novel disease genes.

## 1.4 Research aims and objectives

The field of genomics has progressed extensively since the advent and implementation of NGS in clinical pipelines, however many patients remain undiagnosed. The two aims of this thesis are to:

- (i) identify novel genetic variants in patients with rare monogenic diseases, and
- (ii) characterise novel disease mechanisms identified.

Novel genetic variants will be identified by reanalysis of clinical data where available, and by undertaking new WGS or WES analysis. Patients analysed will include individuals with complex or novel phenotypes who have received a negative genetic test if available.

Functional characterisation of novel genetic variants, or variant in genes without an existing known mechanism of disease will be conducted using methods and models appropriate to the variants and genes identified. Characterisation of novel gene involvement in disease is a useful conduit to improve understanding basic biology in human disease and permits the identification of affected individuals in the future.

These objectives will be achieved in a case-study approach. Analysis will focus on families and cohorts with undiagnosed Mendelian disorders. Given the timeline of the diagnostic journey for patients with rare diseases, the cases within this project fall at different points along the diagnostic timeline. An overview of the aim and objective of the cases studied in subsequent chapters is summarised below:

### Chapter 3:

- Aim: To identify a novel genetic variant in a case of suspected Urofacial syndrome.
- Objective: Undertake trio WES of the affected individual and screening in other unsolved cases with Urofacial syndrome.

### Chapter 4:

- Aim: To identify a monogenic cause of URSMS in a cohort of affected fetuses.
- Objective: Undertake WGS analysis of affected fetuses.

### Chapter 5:

- Aim: To identify the causative genetic variant in a family with a variable form of Occipital Horn syndrome.

- Objective: Undertake reanalysis of clinical WES and WGS data and undertake functional characterisation to explain phenotypic variability.

#### Chapter 6:

- Aim: To characterise a novel monogenic disorder of encephalopathy secondary to infection.
- Objective: Undertake reanalysis of clinical WES data and undertake functional characterisation of the disease mechanism.



## **Chapter 2: Methods and Materials**

This chapter details the methods and materials relating to sample handling and genetic analysis used in multiple experimental chapters of this thesis. Details relating to specific functional studies are available within each experimental chapter.

### **2.1 Ethics approval**

All patients provided written and informed consent in accordance with local regulations in the countries where ascertainment and recruitment took place. Ethical approval for this study was granted by the National Health Service (NHS) Ethics committee (IRAS:64321) and The University of Manchester.

### **2.2 Patient samples**

Patient samples were provided by the referring clinician as blood, DNA, or primary fibroblast cultures. Patient DNA was extracted and stored at -80 °C by the NHS Genomic Diagnostic Laboratory at The Manchester Centre for Genomic Medicine, an accredited medical laboratory. Similarly, patient primary fibroblast cultures were established and stored in liquid nitrogen by the NHS Genomic Diagnostic Laboratory.

#### **2.2.1 DNA extraction from patient blood samples**

M-PVA Magnetic Bead Technology was employed using the chemagic MSM I instrument (PerkinElmer, MA, USA) to extract DNA from peripheral blood lymphocytes from affected individuals by the NHS Genomic Diagnostic Laboratory at The Manchester Centre for Genomic Medicine following manufacturer's instructions.

#### **2.2.2 RNA extraction from patient cells**

Frozen patient cell pellets for RNA extraction were provided by Specialised Cell Culture Services at the The Manchester Centre for Genomic Medicine NHS Genomic Diagnostic Laboratory. To extract RNA, cell pellets were thawed on ice prior to extraction using the RNeasy RNA extraction kit (Qiagen), following manufacturer's instructions, including the optional DNase digestion step. 1 µg RNA was reverse transcribed to cDNA using High-

Capacity RNA-to-cDNA kit (Applied Biosystems), following manufacturer's instructions.

Patient cDNA was stored at -20 °C until further use.

### 2.2.3 DNA and RNA quantification

Extracted nucleic acids were quantified using the NanoDrop 8000 spectrometer (ThermoFisher).

## 2.3 Next generation sequencing

NGS data was sequenced using various platforms as per standard operation at the time.

### 2.3.1 WGS

Whole genome sequencing (WGS) was conducted using patient DNA samples sent to the Beijing Genomics Institute (BGI) facility, Shenzhen, China. Libraries of DNA nanoballs (DNBs) were prepared using rolling circle amplification. Sequencing data were generated to an average 30x depth with pair-end 100bp sequencing using the BGI-Seq500 platform.

### 2.3.2 WES BGI

Whole exome sequencing (WES) was also conducted using patient DNA samples sent to the Beijing Genomics Institute (BGI) facility, Shenzhen, China. For WES, library preparation was conducted using the BGI exome kit, version 4 (59M) 6G. Sequencing data were generated to an average 100x depth with pair-end 100bp sequencing using the BGI-Seq500 platform.

### 2.3.3 WES Illumina

WES was also conducted using the Illumina HiSeq 2500 platform (Illumina, UK) as previously described (Smith et al., 2014). The Agilent SureSelect Human All Exon Kit v4 (Agilent Technologies, UK) was used for library preparation, and sequencing was performed on the HiSeq 2500 system (Illumina, UK).

### 2.3.4 NGS data processing

Following sequencing on each platform, reads were aligned to the genomic reference sequence (hg19) via Burrows-Wheeler Aligner (BWA v.0.6.2; <http://bio-bwa.sourceforge.net>). Genome Analysis Tool Kit software (version 2.4.7; <https://www.broadinstitute.org/gatk>) was used for base quality score recalibration and the

unified genotype (<https://www.broadinstitute.org/gatk>) was used for variant calling. Data generated by BGI underwent mapping and variant calling at the sequencing facility. Illumina-generated WES data were processed locally.

### 2.3.5 Analysis of single nucleotide variants

Analysis and annotation of genome data was performed in-house using VarSeq™ v2.2 (Golden Helix, Inc., Bozeman, MT, USA). Variant filtering strategy (Figure 2) began with exclusion of variants with a genotype quality (GQ) score lower than 20. Variants in intergenic regions were filtered out. Rare variants (< 0.01 %) were identified from population frequencies observed in the Genome Aggregation Database (gnomAD, <https://gnomad.broadinstitute.org/>) and using population frequencies from in-house exome sequencing. Remaining variants were annotated with data from dbNSFP v4.0 (Liu et al., 2011), a compendium of data from RefSeq Genes 105 v1 (<https://www.ncbi.nlm.nih.gov/refseq/>), OMIM (<https://www.omim.org/>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), and data from *in silico* models, including Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Mutation taster (<http://www.mutationtaster.org/>), Mutation Assessor (<http://mutationassessor.org/r3/>), SIFT (Sorting Intolerant From Tolerant) (<http://sift.bii.a-star.edu.sg/>), CADD (GRCh37-v1.4) (Rentzsch et al., 2019), and REVEL (Ioannidis et al., 2016). Variants categorised as “pathogenic” or “likely pathogenic” by ClinVar were also retained. Variants were exported as an excel file format for further filtering according to expected mode of inheritance. Rare variants with potential splicing effects in candidate genes were assessed using SpliceAI (Jaganathan et al., 2019). Analysis of variants of interest on different transcripts was conducted using Alamut™ VisualPlus software (SOPHiA GENETICS). Visualisation of variants in associated BAM files was facilitated by Integrative Genomics Viewer software (<https://igv.org/>).

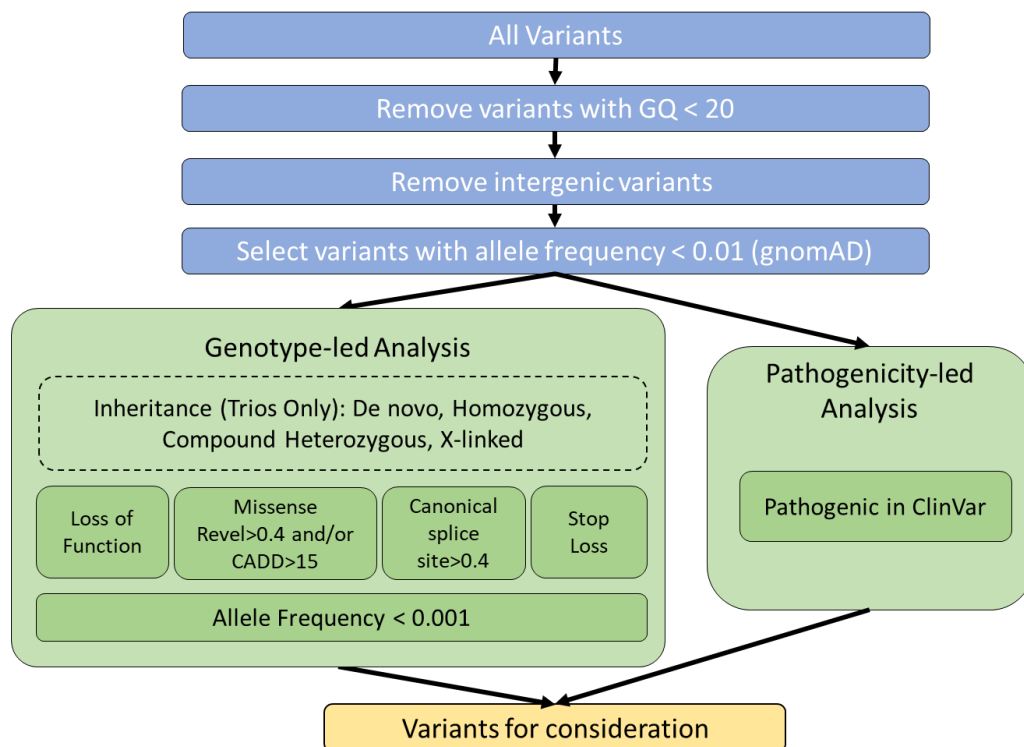


Figure 2.1: Default filtering strategy used to analyse WGS and WES data. Availability of patient-parent trio sequencing data permitted filtering of variants according to expected mode of inheritance. VCF files are filtered by genotype quality (GQ), before annotation with metadata for further filtering.

### 2.3.6 Analysis of copy number variants

Copy number variants (CNVs) were detected from aligned WGS data using CNVnator (Abyzov et al., 2011). CNVs were annotated using ClassifyCNV (Gurbich and Ilinsky, 2020). ClassifyCNV annotates CNVs with haploinsufficiency and triplosensitivity data from ClinGen and annotates scores following guidelines recommended by ACMG (Riggs et al., 2020).

## 2.4 Sanger sequencing

Sanger sequencing was performed in order to confirm variants identified from WES or WGS analysis and conduct segregation analysis as appropriate.

### 2.4.1 Designing primers

Primers were designed using Primer3 v4.1.0 software (<https://bioinfo.ut.ee/primer3-0.4.0/>) using default settings. Primers were checked to ensure no SNPs were present within genomic DNA by using the SNPCheck software (<https://genetools.org/SNPCheck/snpcheck.htm>). Primers were ordered from Sigma-Aldrich (St Louis, MO, USA) at a concentration of 100  $\mu$ M and stored at -20°C.

### 2.4.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was conducted in advance of Sanger sequencing to amplify the targeted region. PCR reactions contained GoTaq G2 Green 2X Master Mix (Promega), 10 ng DNA, and 10  $\mu$ M primers (forward and reverse). Reactions, made up to 25  $\mu$ l underwent PCR amplification using standard thermal cycles (Table 2.1) using the Veriti 96 well Thermal Cycler (Applied Biosystems).

Table 2.1. Thermal cycler program for PCR amplification.

Stage	1	2			3	4
	Activation of polymerase	Denaturation	Annealing	Elongation	Extension	Storage
Number cycles	1	35			1	1
Temperature	95°C	95°C	56-60°C	72°C	72°C	4°C
Time	3 min	30 sec	30 sec	30 sec	5 min	hold

### 2.4.3 Agarose gel electrophoresis

The product of the PCR reaction was confirmed via agarose gel electrophoresis. Agarose gels were prepared by dissolving 1 g agarose powder in 100 ml Tris-Borate-EDTA (TBE) and

adding 1:1000 SafeView nucleic acid stain (NBS Biologicals Ltd.). 7 µl PCR product was run on a 1 % (w/v) agarose gel in an electrophoresis tank containing 1x TBE buffer at 120 V for 20 min. Sample bands were visualised using an ultraviolet transilluminator.

#### 2.3.4 PCR product purification

The remaining 18 µl of the PCR reaction product was purified using Agencourt AMPure XP paramagnetic beads (Beckman Coulter). 32.4 µl AMPure beads were added to PCR products and processed using a Biomek NX robotics instrument (Beckman Coulter) following manufacturer protocol. Purified PCR product was eluted in 100 µl dH<sub>2</sub>O, and stored at -20 °C.

#### 2.3.5 Sanger sequencing reactions

Sanger sequencing reactions contained BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). Each 10 µl reaction contained 1 µl purified PCR product, 0.25 µl Big Dye Ready Reaction mix, 1.875 µl 5X Sequencing Reaction buffer, 1.6 µl of 2 µM forward or reverse primer, and 4.275 µl dH<sub>2</sub>O. Sequencing reactions were run in Veriti 96 well Thermal Cycler (Applied Biosystems) using the protocol described in table 2.2.

Table 2.2. Sequencing reaction cycles

Stage	1	2			3
	Incubation	Denaturation	Annealing	Extension	Storage
Number cycles	1	32			1
Temperature	96 °C	96 °C	55 °C	60 °C	4 °C
Time	1 min	10 sec	10 sec	4 min	Hold

#### 2.3.6 Sequencing reaction product purification and capillary electrophoresis

Sequencing products were purified using Agencourt CleanSeq paramagnetic beads (Beckman Coulter). 5 µl beads were added to each 10 µl reaction and processed using a Biomek NX robotics instrument (Beckman Coulter) following manufacturer protocol. Purified sequencing reactions were resuspended in 10 µl Hi-Di formamide, and capillary

electrophoresis was performed using a 3730 Genetic Analyser instrument (Thermo Fisher Scientific).

#### 2.3.7 Sanger sequencing analysis

Sanger sequencing traces were analysed using either GeneScreen software (Carr et al., 2011) or Staden Software package (<https://staden.sourceforge.net/>). Sequencing files were added to databases created using PreGap4 software and compared with control DNA processed as the reference sequence.

## Chapter 3: Early B-cell Factor 3-Related Genetic Disease Can Mimic Urofacial Syndrome

Kidney Int Rep (2020) 5, 1823–1827; DOI: 10.1016/j.ekir.2020.07.001

J. Robert Harkness<sup>1,2\*</sup>, Glenda M. Beaman<sup>1,2\*</sup>, Keng W. Teik<sup>3</sup>, John A. Sayer<sup>4,5,6</sup>, Heather J. Cordell<sup>7</sup>, Huw B. Thomas<sup>2</sup>, Katherine Wood<sup>1,2</sup>, Helen M. Stuart<sup>1,2</sup>, Adrian S. Woolf<sup>8,9</sup>, William G. Newman<sup>1,2</sup>

<sup>1</sup>Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK.

<sup>2</sup>Division of Evolution and Genomic Sciences, Faculty of Biology, Medicine and Health Sciences, University of Manchester, Manchester, UK.

<sup>3</sup>Genetic Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia.

<sup>4</sup>Clinical Medicine Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, UK

<sup>5</sup>Renal Services, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.

<sup>6</sup>NIHR Newcastle Biomedical Research Centre, Newcastle University, Newcastle upon Tyne, UK.

<sup>7</sup>Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK.

<sup>8</sup> Division of Cell Matrix Biology and Regenerative Medicine, School of Biological Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK.

<sup>9</sup> Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.



# Early B-cell Factor 3–Related Genetic Disease Can Mimic Urofacial Syndrome



J. Robert Harkness<sup>1,2,11</sup>, Glenda M. Beaman<sup>1,2,11</sup>, Keng W. Teik<sup>3</sup>, Sangeet Sidhu<sup>4</sup>, John A. Sayer<sup>5,6,7</sup>, Heather J. Cordell<sup>8</sup>, Huw B. Thomas<sup>2</sup>, Katherine Wood<sup>1,2</sup>, Helen M. Stuart<sup>1,2</sup>, Adrian S. Woolf<sup>9,10</sup> and William G. Newman<sup>1,2</sup>

<sup>1</sup>Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK; <sup>2</sup>Division of Evolution and Genomic Sciences, Faculty of Biology, Medicine and Health Sciences, University of Manchester, Manchester, UK; <sup>3</sup>Genetic Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; <sup>4</sup>Paediatric Department, Hospital Pulau Pinang, Pulau Pinang, Malaysia; <sup>5</sup>Clinical Medicine Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; <sup>6</sup>Renal Services, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; <sup>7</sup>NIHR Newcastle Biomedical Research Centre, Newcastle University, Newcastle upon Tyne, UK; <sup>8</sup>Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; <sup>9</sup>Division of Cell Matrix Biology and Regenerative Medicine, School of Biological Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK; and <sup>10</sup>Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

**Correspondence:** William G. Newman, Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Manchester, M13 9WL, UK. E-mail: [william.newman@manchester.ac.uk](mailto:william.newman@manchester.ac.uk)

<sup>11</sup>JRH and GMB contributed equally.

Received 27 April 2020; revised 24 June 2020; accepted 1 July 2020; published online 14 July 2020

*Kidney Int Rep* (2020) 5, 1823–1827; <https://doi.org/10.1016/j.ekir.2020.07.001>

© 2020 International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## INTRODUCTION

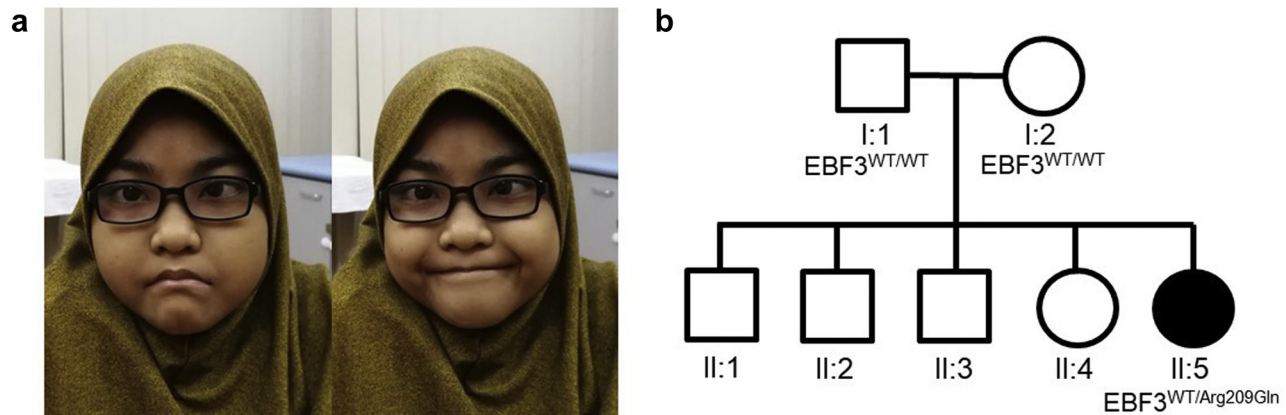
Several decades ago, Bernardo Ochoa<sup>1</sup> described a rare but potentially devastating inherited disease that is now called urofacial, or Ochoa, syndrome (UFS).<sup>2</sup> UFS is characterized by 2 features. First, a so-called “non-neurogenic neurogenic” dyssynergic bladder in which functional bladder outflow obstruction causes incomplete voiding, vesicoureteric reflux (VUR), ascending urosepsis, pyelonephritis, and renal failure. Second, a grimace where the corners of the mouth become downturned on smiling, so that the face appears to be crying. Severe constipation is reported in approximately two-thirds of cases.<sup>1</sup> Ochoa's<sup>1</sup> cohort was from Colombia, but UFS has subsequently been reported worldwide and cases totaled at least 150.<sup>2</sup> UFS is an autosomal recessive disorder and most cases are caused by biallelic pathogenic variants in either of 2 genes: *HPSE2* (Mendelian Inheritance in Man (MIM) 613469)<sup>3–5</sup> encoding heparanase-2, which inhibits the enzymatic activity of classical heparanase, or *LRIG2* (MIM 615112)<sup>6</sup> encoding leucine-rich repeats and Ig-like domains, a protein that may modulate growth factor signaling. Heparanase 2 and LRIG2 proteins are present in pelvic ganglia and in bladder autonomic nerves emanating from these ganglia,<sup>5–7</sup> and the patterns of bladder nerves are abnormal in mice carrying biallelic variants in either *Hpse2* or *Lrig2*.<sup>7</sup> Thus, UFS features a

peripheral neuropathy affecting the bladder, whereas the cause of the grimace requires further study.

A minority of people with an apparent clinical diagnosis of UFS do not carry variants in *HPSE2* or *LRIG2*.<sup>2</sup> The current report draws attention to a different genetic syndrome, which can mimic UFS and be associated with renal failure. We describe one such individual with a heterozygous missense predicted pathogenic variant in *early B-cell factor 3* (*EBF3*), a gene encoding a transcription factor and associated with hypotonia, ataxia, and developmental delay syndrome (HADDs; MIM 617330).<sup>8,9,51</sup> We proceeded to seek variants in *EBF3* in a UK cohort with familial primary nonsyndromic VUR<sup>52</sup> and reviewed the published literature of HADDs to determine whether other such individuals had renal tract disease.

## CASE PRESENTATION

A 3-year-old Malaysian girl was assigned a presumptive clinical diagnosis of UFS. Antenatal history of the proband was unremarkable but, after a normal delivery, she developed sepsis neonatally followed by recurrent urinary tract infections throughout her childhood. In the first year of life, investigations demonstrated a neurogenic bladder (in the absence of spinal cord damage) with a thickened wall, bilateral VUR, and bilateral hydroureter and hydronephrosis, the latter persisting on scans throughout childhood.



**Figure 1.** Face and family tree of the index case. (a) Proband face at rest (left frame) and the horizontal smile (right frame). (b) Pedigree showing affected proband with clinically unaffected mother, father, and elder siblings. Only the index case, II:5, carried the *EBF3* p.(Arg209Gln) variant.

When her face was at rest, the ends of her mouth were downturned and they failed to become upturned on smiling (Figure 1a), so producing an “abnormal horizontal smile” rather than the classical UFS grimace that features a downturned mouth.

The abnormal smile became less obvious as she grew up, perhaps due to improving myopathic face. She also suffered from constipation through childhood. Although the combination of renal tract disease, constipation, and abnormal smile were broadly compatible with UFS, other features were present that were atypical for UFS, including severe short stature and microcephaly; a deep philtrum, broad chin with midline cleft, and high arched palate; hypotonia with global developmental delay, being unable to walk independently and speak more than 2 words aged 3; tapering digits and fusion between the second and third toes; and a ventricular septal defect and patent ductus arteriosus that resolved spontaneously. Her brain magnetic resonance imaging scan was normal. These additional clinical features raised the possibility of a blended phenotype of UFS and another diagnosis co-occurring in the same individual<sup>53</sup> or of an alternative diagnosis with clinical overlap with UFS.

She died aged 17 years, after a brief illness featuring metabolic acidosis and acute kidney injury, with a plasma creatinine of 687  $\mu\text{M}$  (7.77 mg/100 ml), treated with hemodialysis; blood cultures were negative and her urine showed a mixed growth of gram-negative rods.

### Genetic Analyses of the Family

Parental consent was obtained for genetic analyses. Sanger sequencing of *HPSE2* and *LRIG2* exons failed to identify relevant pathogenic variants in these genes. Whole exome sequencing of the proband was then undertaken (see [Supplementary Methods](#) and [Supplementary Figure S1](#)), and no further variants

were identified in other genes (*BCN2*, *CHRM3*, the smooth muscle genes *ACTA2*, *ACTG2*, *MYH11*, *MYLK*, *MYL9*, *LMOD1*, or *MYOCD*) implicated in congenital bladder outflow obstruction.<sup>54–56</sup> Instead, we identified a heterozygous missense variant c.626G>A, p.(Arg209Gln) in *EBF3* predicted to cause a missense change in a DNA-binding domain of the encoded transcription factor protein. This variant had been reported previously in an individual with HADDs,<sup>8</sup> and it was absent from gnomAD control databases.<sup>57</sup> Sanger sequencing of parental DNA confirmed variant c.626G>A was *de novo*, being absent in the clinically unaffected nonconsanguineous parents (Figure 1b). The proband’s 4 siblings were also assessed as clinically unaffected, and so they were not tested for the variant. Of 8 *in silico* prediction tools used to consider the possible pathogenicity of p.(Arg209Gln), all but 1 predicted that the variant was deleterious ([Supplementary Table S1](#)).

### Sequencing *EBF3* Exons in a Cohort With Primary Nonsyndromic VUR

A previous linkage study using patient cohorts from the United Kingdom, the Republic of Ireland, and Slovenia indicated a susceptibility locus for familial primary nonsyndromic VUR on chromosome 10q26.<sup>58</sup> Given that *EBF3* is located at the edge of the 9 Mb region, we hypothesized that variants in *EBF3* may be enriched in such cases. Accordingly, we undertook Sanger sequencing of *EBF3* exons in 80 individuals randomly selected from the UK cohort<sup>52</sup> but did not identify any predicted deleterious variants ([Supplementary Table S2](#)).

## DISCUSSION

Heterozygous variants in *EBF3* result in the rare developmental disorder, HADDs.<sup>59</sup> The clinical

**Table 1.** Pathogenic *EBF3* variants associated with the renal tract and/or smiling, as reported in this study and in the published literature

<i>EBF3</i> coding variant	<i>EBF3</i> protein variant	Renal tract anomaly	Facial features relevant to smiling	Reference
c.280_283del	p.(Glu94Lysfs*37)	Neurogenic bladder, VUR, recurrent UTIs	Not recorded	Sleven <i>et al.</i> <sup>S1</sup>
c.469_477dup	p.(His157_Ile159dup)	Recurrent UTIs	Not recorded	Harms <i>et al.</i> <sup>9</sup>
c.471C>A	p.(His157Gln)	Neurogenic bladder, VUR	Not recorded	Tanaka <i>et al.</i> <sup>8</sup>
c.487C>T	p.(Arg163Trp)	Atonic bladder, urethral stricture, bilateral VUR and hydronephrosis, recurrent UTIs	Downturned corners of the mouth	Blackburn <i>et al.</i> <sup>S9</sup>
c.488G>A	p.(Arg163Gln)	Impaired bladder control	Triangular-shaped facies	Chao <i>et al.</i> <sup>S10</sup>
c.488G>T	p.(Arg163Leu)	Incomplete bladder emptying, VUR	Triangular-shaped facies	Chao <i>et al.</i> <sup>S10</sup>
c.512G>A	p.(Gly171Asp)	Neurogenic bladder, bilateral VUR	Facial asymmetry	Harms <i>et al.</i> <sup>9</sup>
c.554+1G>T	-	VUR, recurrent UTIs, renal dysplasia	Normal	Sleven <i>et al.</i> <sup>S1</sup>
c.579G>T	p.(Lys193Asn)	VUR	Downturned corners of the mouth, minimal facial expression	Sleven <i>et al.</i> <sup>S1</sup>
c.616C>T	p.(Arg206*)	Hydronephrosis, recurrent UTIs	Not recorded	Tanaka <i>et al.</i> <sup>8</sup>
c.626G>A	p.(Arg209Gln)	Normal	Lack of social smile	Tanaka <i>et al.</i> <sup>8</sup>
c.626G>A	p.(Arg209Gln)	Neurogenic bladder with bilateral VUR and hydronephrosis, recurrent UTIs	Abnormal horizontal smile	Proband in the current study
c.1402_1414del13	p.(Thr464Profs*10)	Recurrent UTIs	Not recorded	Tanaka <i>et al.</i> <sup>8</sup>

UTI, urinary tract infection; VUR, vesicoureteric reflux.

Each row represents a single nonrelated case. The renal tract features are as described in the original publications. In addition, facial features relevant to smiling are indicated.

features vary among individuals, and a subset of cases has urinary tract defects and facial features, including downturned corners of the mouth. Of the 30 cases with pathogenic *EBF3* variants recorded in the literature,<sup>8,9,S1,S9–S11</sup> 10 (33%) were reported to have structural or functional urinary tract defects, including neurogenic bladder and VUR (Table 1). Given that VUR can be asymptomatic, and that it was not evident that all individuals with variants in *EBF3* reported in the literature had undergone urinary tract investigations, this proportion may be an underestimate. Indeed, the proportion rises to 40% when cases with recurrent urinary tract infections are included (Table 1), and whether these additional cases have functional or structural defects of the urinary tract requires further investigation. Of these individuals with *EBF3* variants and urinary tract disease, a subset had facial features (Table 1) that we believe, as in the current case, could have led to confusion with UFS. On the other hand, UFS disease is confined to the urinary tract and the face, whereas *EBF3* variants resulting in HADDs can feature developmental delay and more widespread dysmorphology (Table 2), as evidenced by the current proband.

All but one of the known pathogenic variants in *EBF3* associated with urinary tract disease, fall within

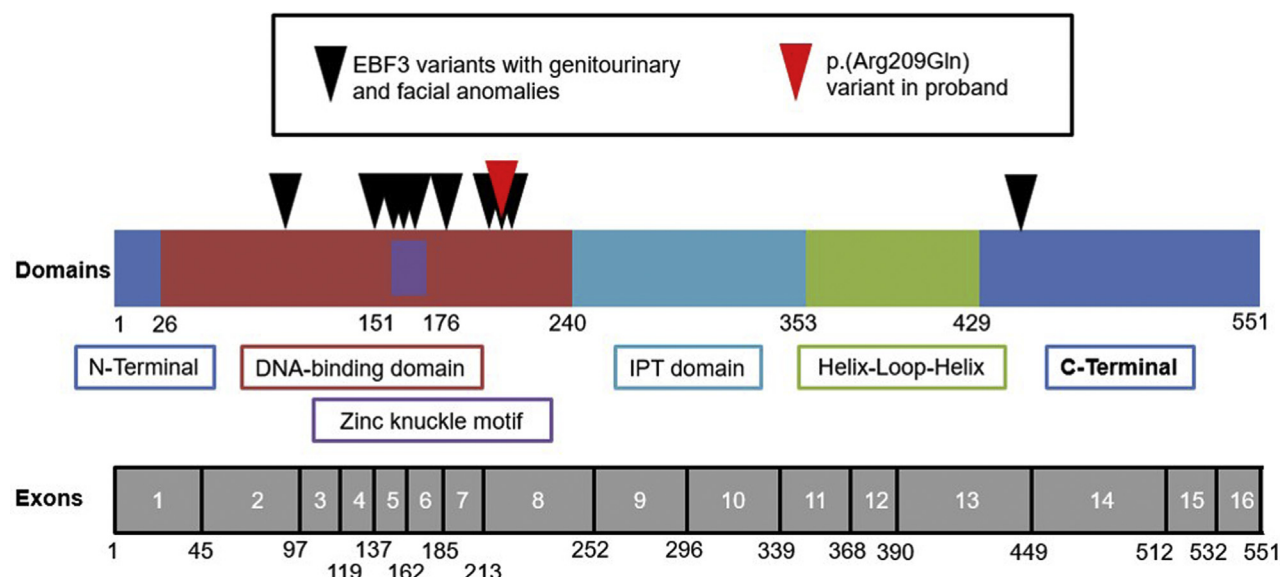
the DNA-binding domain of *EBF3*, as does the variant identified in our proband (Figure 2). Variants in the DNA-binding domain, and more specifically the zinc knuckle motif, have previously been associated with a reduction in *EBF3* binding affinity for DNA,<sup>9</sup> and are therefore likely to alter downstream gene expression, compromising normal development.

As outlined in the introduction, *HPSE2* and *LRIG2*, the genes altered in UFS, are expressed in bladder autonomic nerves, and pathogenic variants in mutant mice lead to abnormal patterns of nerves in the bladder outflow tract and the organ's body.<sup>7</sup> Thus, UFS features a bladder peripheral neuropathy that correlates with the functional outflow obstruction in this disease. Future studies are required to understand how *EBF3* affects the biology of the urinary tract. We note, however, that in the GUDMAP gene expression database,<sup>S12,S13</sup> *Ebf3* transcripts are detected in the embryonic mouse bladder and urethral region, and in the urorectal septum of the nascent organ (<https://www.gudmap.org/id/Q-48NJ@2T7-SRMW-5H12>). It is currently unclear whether they are expressed in particular tissue compartments such as the nerves, muscle, or urothelium.

Aristaless-Related Homeobox (ARX) is an upstream translational repressor of *EBF3*,<sup>S10,S14,S15</sup> and ARX variants have been linked to a wide variety of developmental disorders, including epileptic encephalopathy (MIM 308350), Partington X-linked mental retardation syndrome (MIM 309510), and Proud syndrome (MIM 300004).<sup>8,S16,S17</sup> There is overlap in the range of phenotypes of individuals with ARX variants and those with *EBF3* variants. We speculate that the ARX-*EBF3* interaction may lie upstream of *HPSE2* and *LRIG2*, which are directly implicated in the

**Table 2.** Key teaching points

1. When an individual has vesicoureteric reflux (VUR) and/or dysfunctional urinary voiding plus an abnormal smile, clinicians should consider urofacial syndrome (UFS) and seek biallelic variants in *HPSE2* or *LRIG2*.
2. When a patient has an apparent UFS phenotype plus features such as developmental delay and broader dysmorphology, clinicians should consider hypotonia, ataxia, and developmental delay syndrome (HADDs) syndrome and seek variants in *EBF3*.
3. UFS is an autosomal recessive disease, in contrast to HADDs, which is autosomal dominant.



**Figure 2.** Domains and exons of the transcription factor protein encoded by *EBF3*. Numbers correspond to amino acids. Red arrowhead indicates the variant in the current proband. Black arrowheads point to positions of published variants associated with renal tract and facial anomalies, as detailed in Table 1. Zinc knuckle motif is contained within the DNA-binding domain IPT (Ig-like/plexins/transcription factor domain).

pathobiology of UFS. Perhaps yet other genes associated with UFS-like disease lie downstream of the ARX-*EBF3* pathway, likely with a restricted pattern of expression within key neurological pathways involved in innervation of bladder and facial muscle.

As recently reviewed,<sup>54</sup> the several decades search for genes implicated in primary nonsyndromic VUR, a common condition affecting at least 1% of infants, has yet to yield definitive answers, apart from *TNXB1* in a small subset of familial cases.<sup>518</sup> The condition is most likely genetically heterogeneous, but parametric analyses of 460 affected families comprising 1062 affected individuals showed linkage to rs7907300 located at chromosome 10q24.<sup>58</sup> *EBF3* is 1 of 69 genes that fall within the wider 9 Mb region surrounding rs7907300, making *EBF3* a plausible candidate for conferring VUR susceptibility. Nevertheless, we failed to detect likely pathogenic variants in *EBF3* in 80 index cases from the UK, themselves comprising approximately half of the total group analyzed in the 10q24 linkage study. Previously, we had analyzed the UK cohort by sequencing *HPSE2*<sup>5</sup> and *LRIG2*,<sup>7</sup> with similar negative results. However, it should be noted that patients with VUR associated with neuropathic-type bladders were excluded from this particular collection.<sup>52</sup> In contrast, we have identified potentially pathogenic biallelic *LRIG2* variants in 3 individuals with both severe bladder voiding dysfunction and renal failure.<sup>6,7</sup> Thus, we suggest that the latter population may be enriched for individuals with Mendelian genetic causes of their urinary tract disease.

## CONCLUSION

The association of urinary tract disease, especially a neuropathic bladder and VUR, with an abnormal smile should alert clinicians to the possibility of not only UFS, but also HADDs (Table 2). These insights have implications for genetic counseling given that UFS is autosomal recessive, whereas HADDs is a dominantly inherited disease. *HPSE2* and *LRIG2*, the genes implicated in UFS, pattern bladder nerves. The mechanism whereby *EBF3* affects urinary tract biology requires further study (Table 2).

## DISCLOSURE

All the authors declared no competing interests.

## ACKNOWLEDGMENTS

We thank the family for consent to publish the proband's photo. We thank the UK VUR Study Group for providing index cases with primary nonsyndromic VUR. We acknowledge grant support from: the Wellcome Trust (PhD studentship to JRH); Kidney Research UK (project grant Paed\_RP\_002\_20190925 to WGN and ASW); the Medical Research Council (project grant MR/L002744/1 to ASW and WGN, and MR/T016809/1 to ASW); Newlife Foundation (project grants 15-15/03 and 15-16/06 to WGN and ASW); the NIHR Academic Lecturer scheme (HMS); the Academy of Medical Sciences (HMS); and the Manchester NIHR BRC (IS-BRC-1215-20007 to WGN).



## SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods.

Supplementary References.

**Table S1.** *In silico* tools interrogating the possible pathogenicity of the *EBF3* variant c.626G>A found in the index case.

**Table S2.** Variants detected in *EBF3* sequencing of 80 index cases of the UK VUR cohort.

**Figure S1.** Filtering strategy used for whole exome sequencing data analysis.

## REFERENCES

- Ochoa B. Can a congenital dysfunctional bladder be diagnosed from a smile? The Ochoa syndrome updated. *Pediatr Nephrol.* 2004;19:6–12.
- Newman WG, Woolf AS. Urofacial syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews*®. Seattle (WA): University of Washington, Seattle; 1993. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK154138/>. Accessed March 31, 2020.
- Daly SB, Urquhart JE, Hilton E, et al. Mutations in *HPSE2* cause urofacial syndrome. *Am J Hum Genet.* 2010;86:963–969.
- Pang J, Zhang S, Yang P, et al. Loss-of-function mutations in *HPSE2* cause the autosomal recessive urofacial syndrome. *Am J Hum Genet.* 2010;86:957–962.
- Stuart HM, Roberts NA, Hilton EN, et al. Urinary tract effects of *HPSE2* mutations. *J Am Soc Nephrol.* 2015;26:797–804.
- Stuart HM, Roberts NA, Burgu B, et al. *LRIG2* mutations cause urofacial syndrome. *Am J Hum Genet.* 2013;92:259–264.
- Roberts NA, Hilton EN, Lopes FM, et al. *Lrig2* and *Hpse2*, mutated in urofacial syndrome, pattern nerves in the urinary bladder. *Kidney Int.* 2019;95:1138–1152.
- Tanaka AJ, Cho MT, Willaert R, et al. De novo variants in *EBF3* are associated with hypotonia, developmental delay, intellectual disability, and autism. *Cold Spring Harb Mol Case Stud.* 2017;3:a002097.
- Harms FL, Girisha KM, Hardigan AA, et al. Mutations in *EBF3* disturb transcriptional profiles and cause intellectual disability, ataxia, and facial dysmorphism. *Am J Hum Genet.* 2017;100:117–127.

## **Chapter 4: Clinically diverse and perinatally lethal syndromes with urorectal septum malformation sequence**

American Journal of Medical Genetics Part A, 1–12. DOI: 10.1002/ajmg.a.63067

Shalini S Nayak<sup>1</sup>, J. Robert Harkness<sup>2,3</sup>, Anju Shukla<sup>1</sup>, Siddharth Banka<sup>2,3</sup>, William G Newman<sup>2,3</sup>, Katta M Girisha<sup>1,4</sup>

<sup>1</sup>Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India.






<sup>2</sup>Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Manchester, United Kingdom.

<sup>3</sup>Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom.

<sup>4</sup>Department of Genetics, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman.

## ORIGINAL ARTICLE

# Clinically diverse and perinatally lethal syndromes with urorectal septum malformation sequence

Shalini S. Nayak<sup>1</sup>  | Robert Harkness<sup>2,3</sup>  | Anju Shukla<sup>1</sup>  | Siddharth Banka<sup>2,3</sup>  | William G. Newman<sup>2,3</sup>  | Katta M. Girisha<sup>1,4</sup> 

<sup>1</sup>Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka, India

<sup>2</sup>Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Manchester, UK

<sup>3</sup>Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

<sup>4</sup>Department of Genetics, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman

## Correspondence

Katta M. Girisha, Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India.  
 Email: [girish.katta@manipal.edu](mailto:girish.katta@manipal.edu)

## Funding information

Indian Council of Medical Research, Government of India, Grant/Award Number: 5/13/58/2015/NCD-III; DBT/Wellcome Trust India Alliance Grant, Grant/Award Number: IA/CRC/20/1/600002; Manchester NIHR BRC, Grant/Award Number: IS-BRC-1215-20007

## Abstract

Urorectal septum malformation sequence (URSMS) is characterized by a spectrum of anomalies of the urogenital system, hindgut and perineum. It is presumed to be a constellation of an embryonic defect. Herein, we analyzed the clinically diverse syndromes associated with URSMS in our perinatal evaluation unit. We reviewed fetuses with URSMS in referrals for perinatal autopsy over a period of 3 years. Chromosomal microarray and genome sequencing were performed whenever feasible. Literature was reviewed for syndromes or malformations with URSMS. We ascertained URSMS in 12 of the 215 (5%) fetuses. Nine fetuses (75%) had complete URSMS and remainder had partial/intermediate URSMS. Eleven fetuses had malformations of other systems that included: cerebral ventriculomegaly; right aortic arch with double outlet right ventricle; microcephaly with fetal akinesia deformation sequence; ventricular septal defect and radial ray anomaly; thoraco-abdominoschisis and limb defects; myelomeningocele; spina bifida and fused iliac bones; omphalocele; occipital encephalocele; lower limb amelia and cleft foot. We report on six fetuses with recurrent and five fetuses with unique malformations/patterns where URSMS is a component. Exome sequencing (one family) and genome sequencing (eight families) were performed and were nondiagnostic. Additionally, we review the literature for genetic basis of this condition. URMS is a clinically heterogeneous condition and is a component of several multiple malformation syndromes. We describe several unique and recurrent malformations associated with URSMS.

## KEYWORDS

ambiguous genitalia, imperforate anus, persistent cloaca, urorectal septum malformation sequence

## 1 | INTRODUCTION

Urorectal septum malformation sequence (URSMS) refers to a spectrum of anomalies involving urogenital system, hindgut and perineum. It is hypothesized that URSMS results from deficiency of the caudal end of the mesoderm or defect in the cloacal endoderm, peri-cloacal mesenchyme, and genital ectoderm arising during early embryogenesis (Jain et al., 2008). However, a definitive cause is not yet identified.

This sequence is described as complete, partial, intermediate, and urogenital sinus anomaly based on the persistent cloaca and perineal openings (Huang et al., 2016; Wheeler & Weaver, 2001).

Often, other malformations are also observed with URSMS (Tennant et al., 2014). Recurrence in families (vertical and horizontal), though extremely rare, has also been reported (Aggarwal & Phadke, 2013; Mills & Pergament, 1997) suggesting the existence of yet undetermined monogenic etiology. Hsu et al. provided the first

evidence by identifying de novo variants in Caudal Type Homeo Box Transcription Factor 2 (*CDX2*) in patients with persistent cloaca (Hsu et al., 2018). Herein, we describe 12 fetuses with syndromic and non-syndromic URSMS depicting the clinically diverse perinatal lethal malformations.

## 2 | SUBJECTS AND METHODS

We reviewed fetuses with URSMS in referrals for perinatal autopsy over a period of 3 years at our perinatal genetics services clinic. The study has the approval of the institutional ethics committee and informed consents were taken from the families. Perinatal evaluation was performed (Khong & Malcomson, 2015; Nayak et al., 2015) in all the fetuses. Family history of the couple, antenatal details, exposure to teratogens, consanguinity, three-generation pedigree, findings on antenatal ultrasonography and fetal echocardiography, results of laboratory tests, mode of delivery, and gestational age were systematically documented. Fetal weight, total length, foot length, head circumference, chest circumference, and other measurements were taken and compared with available standard growth charts (Khong & Malcomson, 2015). Anteroposterior and lateral views of the fetus were radiographed to rule out skeletal abnormalities. Additionally, magnetic resonance imaging of the fetal brain/whole body was performed whenever possible. We collected appropriate samples such as skin biopsy, liver, muscle, bone marrow, or umbilical cord from the fetuses. They were used to obtain DNA/RNA/cell lines. An external examination was done from head to toe to document dysmorphism/anomalies. Dissection of the whole body was performed using standard perinatal autopsy procedures (Khong & Malcomson, 2015; Osborn & Lowe, 2017). All internal organs/structures were examined carefully. The normal anatomy and anomalies observed were documented with photographs. Careful examination of placenta was carried out to rule out the cause of fetal demise. Histopathological examination of fetal tissues/organs was performed as appropriate.

Chromosomal microarray, karyotype, exome sequencing, and genome sequencing were performed using fetal genomic DNA whenever possible. For whole genome sequencing (WGS), DNA library preparation and sequencing was performed using DNBSseq platform at Beijing Genomics Institute (BGI), China with standard 30× sequencing depth. Reads were aligned to hg19 human reference genome via Burrows–Wheeler Aligner (BWA v.0.6.2). Variants were called using Genome Analysis Toolkit (GATK) 4.0. Annotation was performed using VarSeqTM v2.2 (Golden Helix Inc., Boseman, MT, USA). Rare variants (<0.01%) were identified from population frequencies observed in the Genome Aggregation Database (gnomAD) and using population frequencies from in-house exome sequencing. In silico variant predictions were modeled using Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Mutation taster (<http://www.mutationtaster.org/>), Mutation Assessor (<http://mutationassessor.org/r3/>), SIFT (Sorting Intolerant From Tolerant) (<http://sift.bii.a-star.edu.sg/>), CADD (GRCh37-v1.4), and REVEL (Ioannidis et al., 2016). Rare intronic and splice site variants were assessed using SpliceAI (Jaganathan et al., 2019). Fetal sex was inferred by non-pseudoautosomal region

X-chromosome genotype analysis using the bcftools guess-ploidy plugin, allowing for genotyping error rate of 0.1.

A careful delineation of the URSMS subtypes and anomalies in other systems was carried out. Literature was reviewed for syndromes or the phenotypes associated with URSMS. We searched PubMed, Online Mendelian Inheritance in Man and other databases with the phrases “urorectal septum malformation sequence,” “persistent cloaca,” “cloacal dysgenesis,” and “cloaca” to look for the associated malformations.

## 3 | RESULTS

We observed URSMS in 12 fetuses (5%) in our cohort of 215 fetal losses. The gestational age ranged 13–36 weeks. Ambiguous genitalia with no perineal openings and persistent cloaca characterizing the complete URSMS was observed in nine (9/12, 75%) fetuses. There was persistent cloaca with a single perineal opening (partial URSMS) in two fetuses and the other fetus had intermediate type (complete partition of the persistent cloaca without perineal openings) of URSMS. Oligohydramnios (4/12, 33%) or renal agenesis/anomaly (4/12, 33%) or megacystis (1/12, 8%) were observed prenatally in the cohort. The clinical details of all the fetuses are provided in Table 1. Overall, we could identify sex in five fetuses based on external and internal examination and histopathology. Recurrence of the condition was not observed in any of the families in our cohort.

Eleven (11/12, 91.6%) fetuses had malformations (Table 1) outside the urinary tract, internal and external genital and hindgut. Anomalies of vertebrae (10/12, 83%), limbs (8/12, 66%), central nervous system (5/12, 41%), umbilical cord (4/11, 36%), heart (3/12, 25%), and ventral body wall (3/12, 25%) were frequently noted in our cohort.

### 3.1 | Unique malformations with URSMS

After reviewing the literature for syndromes or conditions with multiple malformations where URSMS is a component, we note, five fetuses had a rare or unique pattern of multiple malformations in our cohort. These rare presentations are described here.

#### 3.1.1 | Complete URSMS with cerebral ventriculomegaly (Fetus 2)

The fetus was ascertained at 18 weeks of gestation in view of oligohydramnios and bilateral renal agenesis. The fetus had normal anthropometric measurements (weighed 175 g, measured 22 cm in length with head circumference of 15 cm). Depressed nasal bridge, micro-retrognathia, proximally placed left thumb, contractures across elbows, hips and knees, calcaniovalgus of right foot and two vessels umbilical cord were seen. Ambiguous external genitalia with a small phallus like structure and absence of perineal openings was noted. Dilated lateral cerebral ventricles were evident on gross examination



TABLE 1 Clinical features in fetuses with urorectal septum malformation sequence (URSMS)

Features	Fetus 1	Fetus 2	Fetus 3	Fetus 4	Fetus 5	Fetus 6	Fetus 7	Fetus 8	Fetus 9	Fetus 10	Fetus 11	Fetus 12
Gestational age in weeks	13	18	18	24	20	14	36	Not available	14	30	21	12
Gender	Female	Unknown	Male	Female	Female	Unknown	Unknown	Unknown	Unknown	Unknown	Male	Unknown
Consanguinity	No	No	No	No	No	No	No	Not available	No	No	No	No
Gravida	Primi	Second	Third	Primi	Second	Second	Second	Not available	Primi	Second	Primi	Primi
Antenatal events	Fever in first trimester	Nil	Diabetes mellitus and hypothyroidism	Nil	Nil	Nil	Nil	Not available	Twin pregnancy	Nil	Nil	Hypothyroidism and polycystic ovarian disease
Type of URSMS	Complete	Complete	Intermediate	Partial	Complete	Complete	Complete	Complete	Complete	Complete	Partial	Complete
External genitalia	Small phallus-like structure with no urogenital openings	Small phallus-like structure with no urogenital openings	Small phallus-like structure with scrotum with no urethral opening	Ambiguous genitalia with no urogenital openings	Small phallus-like structure with no urogenital openings	Small phallus-like structure with no urogenital openings	Small phallus-like structure with no urogenital openings	Small phallus-like structure with no urogenital openings	Small phallus-like structure with no urogenital openings	Ambiguous genitalia with no urogenital openings	Phallus-like structure with a perineal opening	Small phallus-like structure with no urogenital openings
Anal opening	Absent	Absent	Absent	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Persistent cloaca	Present (blind ending)	Present (blind ending)	Urinary bladder with urethral opening	Vesico-vaginal fistula with common opening	Present (blind ending)	Present (blind ending)	Present (blind ending)	Present (blind ending)	Present (blind ending)	Present (blind ending)	Recto-vesicle fistula with common perineal opening	Present (blind ending)
Internal genitalia	Ovaries, fallopian tube	Indifferent gonads	Testes	Ovaries and uterus	? Ovaries, no uterus	Indifferent gonads	Gonads were not found	Indifferent gonads	Indifferent gonads	Indifferent gonads	Testes	Indifferent gonads
Hindgut	Draining into cloaca	Draining into cloaca	Blind ending	Atresia	Draining into cloaca	Draining into cloaca	Draining into cloaca	Draining into cloaca	Draining into cloaca	Draining into cloaca	Draining through common perineal opening	Draining into cloaca
Renal agenesis (unilateral/bilateral)	No	No	No	No	Left renal agenesis	Left renal agenesis (with adrenal)	Bilateral renal agenesis	Left renal agenesis	No	Bilateral renal agenesis	Bilateral renal agenesis	Right renal agenesis
Cystic kidneys	Present (horse-shoe)	Present (fused)	Hypoplastic	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Umbilical cord	Three vessel	Two vessel	Two vessel	Three vessel	Two vessel	Not available	Three vessel	Three vessel	Three vessel	Two vessel	Three vessel	Three vessel
Sacrum	Absent ossification	Absent ossification	Reduced ossification	Normal	Normal	Absent ossification	Absent ossification	Absent ossification	Absent ossification	Absent ossification	Normal	Absent ossification
Central nervous system anomalies	Absent	Dilated lateral cerebral ventricles	Absent	Small brain	Absent	Absent	Lipomeningomyelocele	Lumbar spina bifida	Absent	Occipital encephalocele	Absent	Absent
Cardiac anomalies	Absent	Absent	Right aortic arch, double outlet right ventricle with ventricular septal defect	Absent	Membranous ventricular septal defect	Dextrocardia	Absent	Absent	Absent	Absent	Absent	Absent

(Continues)

TABLE 1 (Continued)

Features	Fetus 1	Fetus 2	Fetus 3	Fetus 4	Fetus 5	Fetus 6	Fetus 7	Fetus 8	Fetus 9	Fetus 10	Fetus 11	Fetus 12
Limb anomalies	Absent	Proximally placed left thumb, contractures across elbows, hips and knees, talipes equinovagis of right foot	Absent	Bilateral congenital talipes equinovarus	Bilateral radial deviation of hands, bowed fore arms, rudimentary right thumb with absent radius, absent F1 and F2 on left hand with hypoplastic radius	Short right lower limb with contractures, reduction defect of left lower limb below the knee with bowed femora	Right radial club hand, bilateral clubfeet and rhizomelia of left lower limb	Absent	Unilateral club foot	Absent	Proximally placed right thumb, complete absence of left lower limb	Short right lower limb with joint contractures across hip, knee and ankle, right cleft foot with absent two central digits (split foot)
Skeletal abnormalities	Absent	Absent	Absent	Scoliosis, crowding of ribs and hyper mineralization cranium	Absent	Kypho-scoliosis at thoraco-lumbar region	Lumbar scoliosis, segmentation defect of thoracic and lumbar vertebrae	Fused iliac bones, absence of ossification centers for lower lumbar vertebrae	Scoliosis at lumbar region	Segmentation defect in lower lumbar vertebrae	Absent	Kypho-scoliosis at thoraco-lumbar region
Ventral body wall defect	Absent	Absent	Absent	Absent	Absent	Lower thoracic and abdominal wall defect with the herniation	Absent	Absent	Omphalocele	Absent	Absent	Anterior abdominal wall defect with the herniation
Other anomalies	Protuberant abdomen with thin and transparent abdominal wall	Rudimentary allantois	Nonlobulated lungs	Microcephaly, excessive skin on the occipital and neck region, sloping forehead, dysmorphic ears, hypertelorism, midface protrusion, macrostomia, micro-retrognathia, short neck, pterygia, camptodactyly, overriding fingers, craniosynostosis, ankyloglossia, hypoplastic lungs	Facial asymmetry on right side, hypertelorism, short nose, anteverted nares, long philtrum and retrognathia	Absent	Pulmonary hypoplasia	Absent	Absent	Frontal sloping, flattened nasal tip, micrognathia, dysmorphic ears, short neck, bilateral pulmonary hypoplasia	Hypertelorism, retrognathia, persistent vitello-intestinal duct	Absent

TABLE 1 (Continued)

Features	Fetus 1	Fetus 2	Fetus 3	Fetus 4	Fetus 5	Fetus 6	Fetus 7	Fetus 8	Fetus 9	Fetus 10	Fetus 11	Fetus 12
Fetal karyotype	Not done	Not done	Done (Normal)	46, XX	Not done	Not done	Not done	Not done	Not done	Not done	Not done	Not done
Chromosomal microarray	Not done	Not done	Not done	Normal	Not done	Not done	Not done	Not done	Not done	Not done	Not done	Not done
Genome sequencing	Singleton	Trio	Singleton	Trio (Also underwent singleton exome)	Singleton	Trio	Not done	Singleton	Singleton	Not done	Not done	Not done
Inferred genomic sex	Female	Male	Male	Female	Female	Male	Not done	Male	Female	Not done	Not done	Not done

and magnetic resonance imaging of fetal brain (Figure 1). Kidneys were multicystic and fused, with narrow ureters draining into blind-ending cloaca. Rudimentary allantois and indifferent gonads were observed. Radiographs of the fetus showed the absence of ossification of sacral vertebrae.

### 3.1.2 | Intermediate URSMS, right aortic arch with double outlet right ventricle (Fetus 3)

A third gravida with diabetes mellitus and hypothyroidism was noted to have anhydramnios, symmetric early onset of growth retardation and cardiac defect in the fetus antenatally. Evaluation of fetus showed normal anthropometry (corresponding to 18 weeks of gestation) with short nose, long philtrum, retrognathia, imperforate anus, small phallus with small scrotum, and single umbilical artery. Right aortic arch with double outlet right ventricle, sub-aortic ventricular septal defect, hypoplastic left ventricle, and nonlobulated lungs were evident (Figure 2). Both kidneys were hypoplastic and ureters draining into the urinary bladder. The urethral opening was absent and gonads appeared to be testes. There was absence of ossification centers for sacrum.

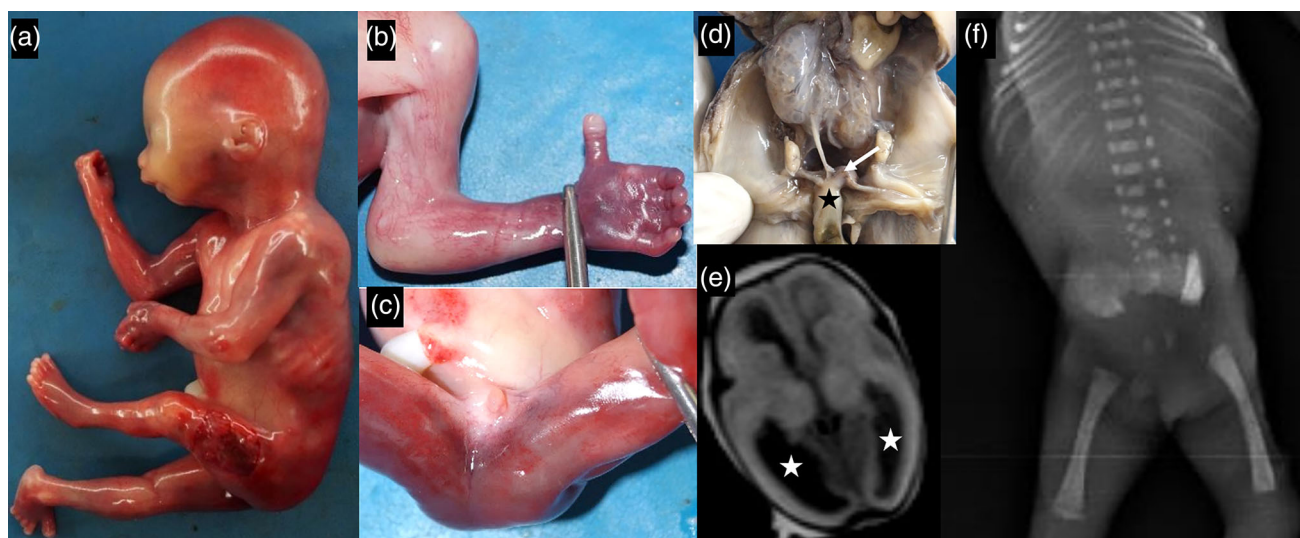
### 3.1.3 | Partial URSMS, microcephaly, and fetal akinesia deformation sequence (Fetus 4)

The fetus weighed 256 g (−3.1 SD) measured 17 cm (−6.8 SD) in length with head circumference of 15 cm (−4.5 SD) at 24 weeks of gestation. Trigonoccephaly was noted with fused anterior fontanel and excessive skin on the occipital and neck region, sloping forehead, dysmorphic ears, hypertelorism, midface protrusion, wide mouth, and micro-retrognathia. The fetus also had short neck with webbing, pterygia across the axillae, elbows, hips and knees, camptodactyly, over-riding of fingers, bilateral congenital talipes equinovarus, ambiguous external genitalia with common urogenital opening and atresia of anal canal (Figure 3).

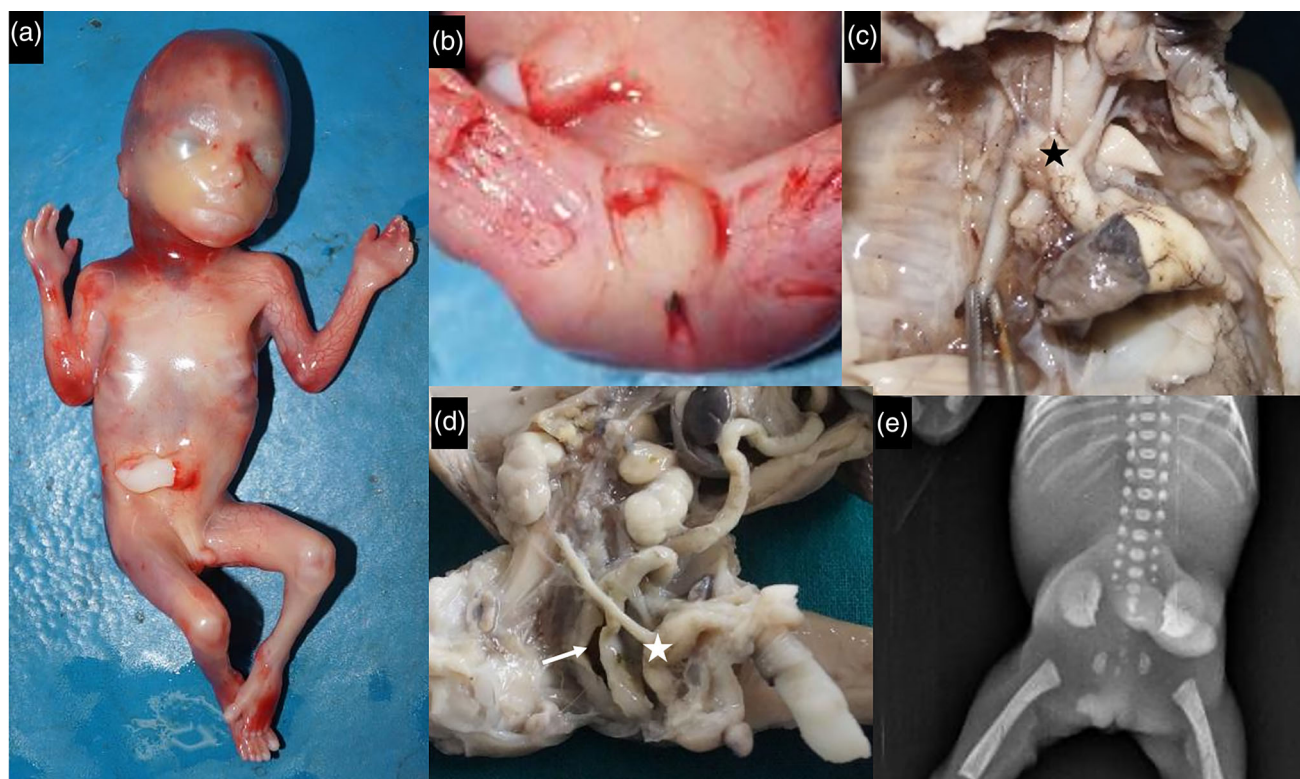
Further examination disclosed underdeveloped brain, ankyloglossia, severely hypoplastic lungs, absence of muscle mass in upper and lower limbs, atresia of terminal portion of hindgut, presence of ovaries and uterus, vesicovaginal fistula with common perineal opening. Scoliosis, crowding of ribs, hyper mineralized cranium, and small head were noted on radiography. Fetal karyotype and chromosomal microarray were normal. Exome sequencing from fetal DNA was nondiagnostic.

### 3.1.4 | Partial URSMS with amelia (Fetus 11)

Prenatal diagnosis disclosed bilateral renal agenesis with absent urinary bladder, malformed left lower limb and intrauterine growth retardation in a fetus. The fetus weighed 223 g (−0.3 SD) measured 19 cm (−2 SD) in length at 19 weeks of gestation. Hypertelorism,



**FIGURE 1** Depressed nasal bridge, micro-retrognathia, contractures in upper and lower limbs (a), proximally placed left thumb (b), ambiguous external genitalia with absent perineal openings (c) and multicystic and fused kidneys with narrow ureters and hindgut (black asterisk, d) draining into common, blind cloaca (arrow, d) were observed in Fetus 2. Imaging revealed cerebral ventriculomegaly (white asterisk, e) and absence of ossification of sacral vertebrae (f).

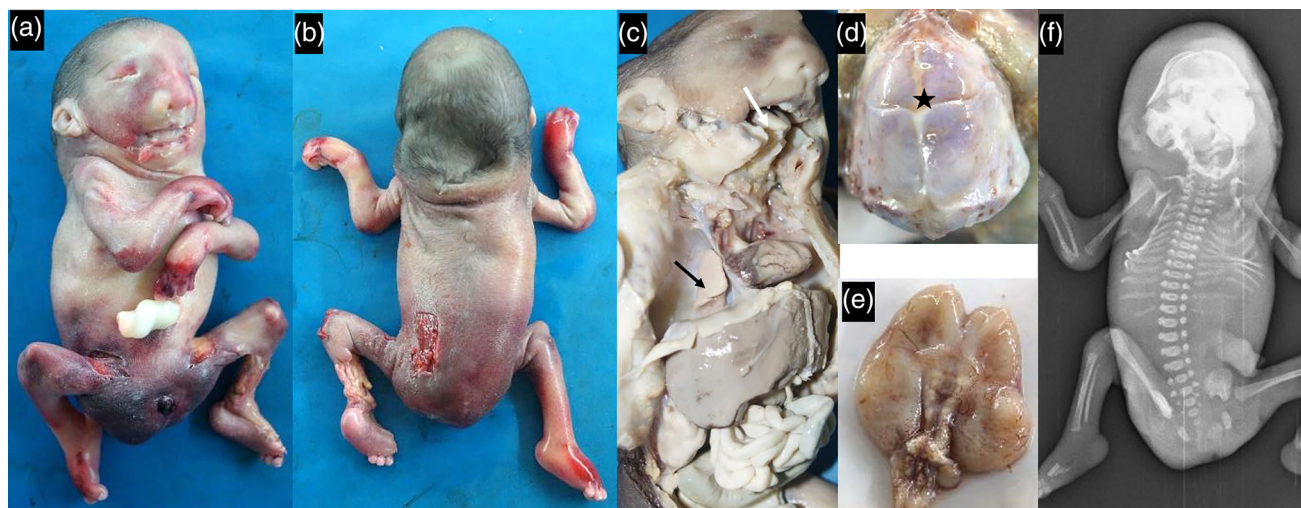


**FIGURE 2** Fetus 3 had short nose, long philtrum, retrognathia (a), small phallus with scrotum and imperforate anus (b), right aortic arch (black asterisk, c), hypoplastic kidneys, separate rectum (arrow) and urinary bladder (white asterisk) with no perineal openings (d) and poorly ossified sacral vertebrae (e).

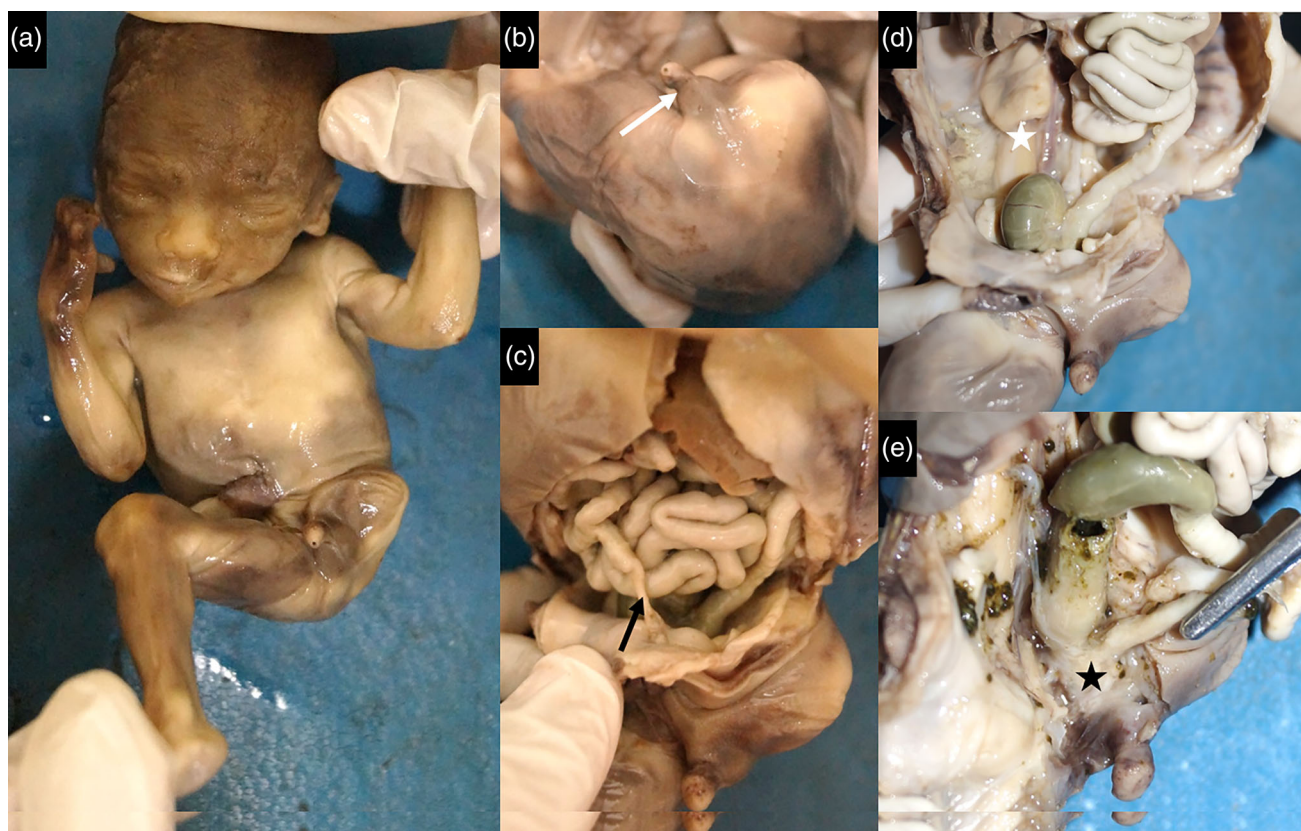
retrognathia, proximally placed right thumb, and complete absence of left lower limb were evident. There was a phallus-like structure with a single perineal opening. Persistent vitellointestinal duct (Meckel's diverticulum), bilateral renal agenesis, and absent ureters with

rudimentary urinary bladder were noted in the fetus (Figure 4). The rectum was communicating with the lower part of bladder and was drained by a common perineal opening. Gonads were confirmed to be testes on histopathology.





**FIGURE 3** Microcephaly, sloping forehead, dysmorphic ears, hypertelorism, midface protrusion, macrostomia and micro-retrognathia, short neck with webbing, pterygia across axillae, elbows, hips and knee joints, ambiguous external genitalia (a, b), ankyloglossia (white arrow), hypoplastic lungs (black arrow, c), premature closure of fontanelle (asterisk, d), small brain (e) and scoliosis, crowding of ribs and hyper-mineralized cranium (f) were evident in Fetus 4.



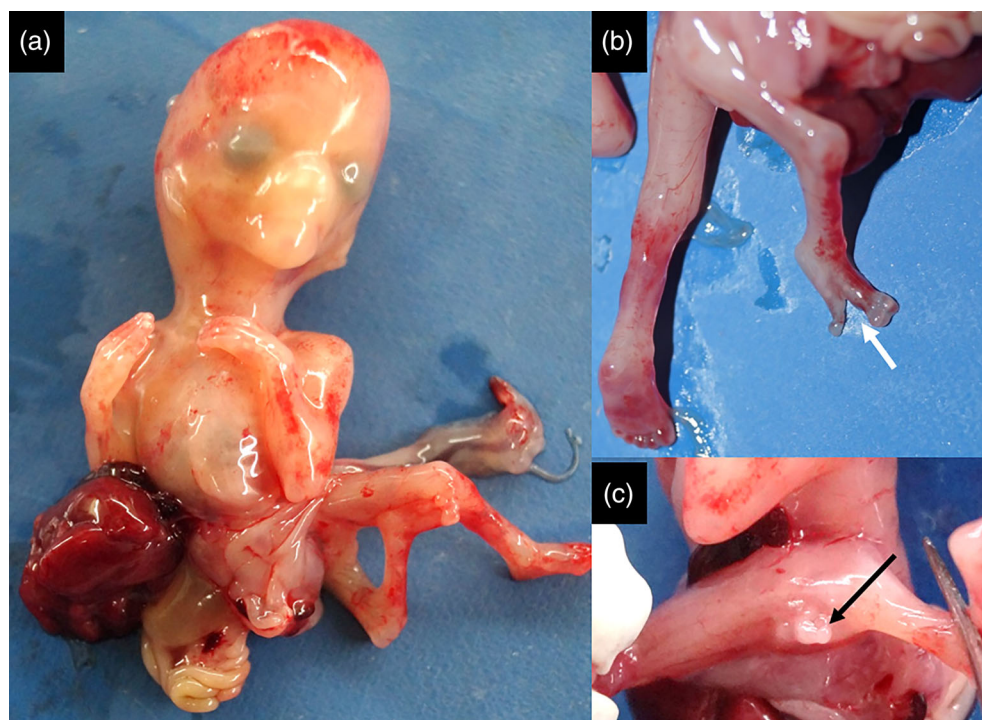
**FIGURE 4** There was amelia of left lower limb (a), phallus-like structure with single perineal opening (white arrow, b), Meckel's diverticulum (black arrow, c), bilateral renal agenesis (white asterisk, d), incomplete separation of rectum, and rudimentary bladder with common outlet (black asterisk, e) in Fetus 11.

### 3.1.5 | URSMS, abdominal wall defect, limb contractures with cleft foot (Fetus 12)

A primigravida with history of laminectomy for scoliosis, polycystic ovarian disease and hypothyroidism underwent pregnancy

interruption at 12 weeks of gestation. Anterior abdominal wall defect with herniation of liver, cystic area in lower portion of abdomen, and posterior bending of lower limbs were observed on ultrasonography.

Growth parameters of the fetus were within normal limits. There was an anterior abdominal wall defect with herniation of liver,



**FIGURE 5** Images depict anterior abdominal wall defect with herniation of abdominal viscera, kyphoscoliosis (a), contractures and spilt right foot (white arrow, b) and ambiguous external genitalia with absent perineal openings (black arrow, c) in Fetus 12.

stomach, spleen, pancreas, and loops of small and large intestine partially covered by peritoneum. Severe kyphoscoliosis at the thoracolumbar region, short right lower limb with joint contractures across hip, knee, and ankle were noted (Figure 5). Right cleft foot was observed with absence of two central digits. Imperforate anus, ambiguous external genitalia, indifferent gonads, a defect in diaphragm, agenesis of right kidney and right adrenal gland, persistent cloaca with hindgut and left ureter draining into it, and atresia of cloacal opening were additional findings.

### 3.2 | Recurrent malformations with URSMS

We observed several recurrent malformations that are already known to occur with URSMS in six fetuses of our cohort. They include radial ray defect, unilateral renal agenesis, and ventricular septal defect suggestive of vertebral defects, anal atresia, tracheoesophageal fistula with esophageal atresia and radial or renal dysplasia (VATER) (Fetus 5); thoraco-abdominoschisis and limb defects (Fetus 6); myelomeningocele/occipital encephalocele, renal agenesis with vertebral segmentation defect (Fetuses 7 and 10); spina bifida and fused iliac bones (Fetus 8) and omphalocele and scoliosis (Fetus 9). The antenatal history, pedigree, clinical findings, and images of the fetuses with recurrent phenotypes are provided in supplementary information S1.

### 3.3 | Genomic testing

Genomic testing was performed in eight (8/12) families. Fetal karyotype was available for Fetus 3 and Fetus 4 following prenatal invasive

testing and showed normal chromosomes. In Fetus 4, chromosomal microarray did not reveal any clinically significant copy number variations. Singleton exome sequencing (nondiagnostic) was carried in Fetus 4. Fetuses 1, 3, 5, 8, and 9 underwent singleton genome sequencing, whereas Fetuses 2, 4, and 6 underwent trio-genome sequencing. Genetic analysis inferred Fetuses 1, 4, 5, and 9 were female, and Fetuses 2, 3, 6, and 8 were male. Data analysis did not identify any pathogenic single nucleotide variants in known genes associated with the URSMS phenotype, but variants of interest have been identified in Fetuses 1, 3, 8, and 9 (Table 2).

## 4 | DISCUSSION

In this study, we report on 12 (5%) fetuses with URSMS and describe the associated pattern of malformations in them. We note complete URSMS in the majority of the fetuses (9/12, 75%), partial URSMS in two fetuses, and intermediate URSMS in one fetus. Eleven fetuses (91.6%) had multiple malformations affecting other systems and one had isolated URSMS. Overall, six fetuses had recurrent and five fetuses had unique malformations in our cohort. We identified variants of interest in four fetuses (Table 2).

There have been reports in the literature on the clinical presentation of fetuses with URSMS (Jain et al., 2008; Patil & Phadke, 2006; Tennant et al., 2014; Wheeler et al., 1997; Wheeler & Weaver, 2001) and a previous cohort from our center (Shah et al., 2016). The rate of associated anomalies of other systems ranges from 65% to 89% (Jain et al., 2008; Patil & Phadke, 2006; Tennant et al., 2014; Wheeler et al., 1997; Wheeler & Weaver, 2001). Limb malformations (e.g., club foot, polydactyly, limb reduction defects, absent/hypoplastic thumb,



**TABLE 2** Variants of interest identified from WGS of eight fetuses with URSMS

Fetus	Gene	Zygosity	GRCh37 coordinate	cDNA	Protein	GnomAD missense z-score (observed/expected)	GnomAD AF	GnomAD south Asian AF	Align GVGD	Mutation taster	Polyphen2	SIFT	CADD	REVEL	ACMG classification	ACMG evidence
1	TMEM132A	Het	11:60701998A > G	NM_017870.2: c.1601A > G	p.Glu534Gly	0.8 (590/647)	0	0	C0	Benign	Possibly Damaging (0.787)	Tolerated (0.27)	23.6	0.13	Uncertain Significance	PM2, PB4
3	NALCN	Het	13:101712309 T > A	NM_052867.4: c.4766A > T	p.Gln1589Leu	4.9% (563/1005.1)	0	0	C0	Deleterious	Possibly Damaging (0.93)	Deleterious (0.01)	23.3	0.864	Uncertain Significance	PM2, PP2, PP3
8	HOXD9	Het	2:176987773A > G	NM_014213.4: c.277A > G	p.Ser93Gly	0.58 (147/168)	0	0	C0	Benign	Possibly Damaging (0.789)	Tolerated (0.10)	22.9	0.383	Uncertain Significance	PM1, PM2, BP4
8	SLIT2	Het	4:20620441G > A	NM_004787.4: c.4399G > A	p.Gly1467Ser	2.1 (687/860)	0	0	C55	Deleterious	Probably Damaging (1.00)	Deleterious (0.00)	29.6	0.51	Uncertain Significance	PM1, PM2, PP3
9	NALCN	Het	13:102051423C > A	NM_052867.4: c.55G > T	p.Gly19Cys	4.9% (563/1005)	0	0	C0	Deleterious	Probably Damaging (0.999)	Deleterious (0.01)	27.7	0.805	Uncertain Significance	PM2, PP3

Abbreviations: AF, allele frequency; URSMS, urorotational septum malformation sequence, WGS, whole genome sequencing.

absent fibula, hip dislocation, cubitus valgus, genu valgum, clinodactyly, and simian creases) and cardiac anomalies (atrial septal defect, ventricular septal defect, tetralogy of Fallot, tricuspid stenosis, pulmonary valve stenosis, truncus arteriosus, atrioventricular canal defect, anomalous venous return, and patent ductus arteriosus) appear to be the most frequent (in 32% each) followed by vertebral anomalies, mainly affecting the lumbar spine in 22.8%; central nervous system anomalies (neural tube defects and holoprosencephaly) in 16.5%; gastrointestinal anomalies (tracheoesophageal fistula, esophageal atresia, small gut atresia or stenosis, Meckel's diverticulum and atresia of bile duct) in 15.7%; ventral body wall defect/omphalocele in 8.6%; single umbilical artery/urachal cyst in 7% and others (congenital diaphragmatic hernia, congenital cystic adenomatoid malformation of lung, cleft palate, perineal hernia) very rarely.

The recurrent malformations or phenotypes observed in our cohort were VATER association, limb body wall defect, neural tube defects and abdominal wall defect. Neural tube defects were more common in our small cohort. Prune-belly sequence and vertebral-anal-cardiac-tracheoesophageal fistula-renal-limb (VACTERL) association were also observed with URSMS earlier (Tennant et al., 2014). Patterns of severe ventral body wall defects were studied by Vauthay et al. and they report cloacal anomalies in 76.9% (Vauthay et al., 2007). Heyroth-Griffis et al. (2007) proposed that exstrophy of cloaca, URSMS and limb body wall complex (LMBW) represented a continuous spectrum of anomalies in view of overlapping features (Heyroth-Griffis et al., 2007). LMBW, Mullerian duct aplasia-renal anomalies-cervicothoracic somite dysplasia (MURCS), oculoauriculo-vertebral spectrum (OAVS), omphalocele-exstrophy-imperforate anus-spinal defects (OEIS) complex, pentalogy of Cantrell (POC) and VATER/VACTERL are developmental disorders occurs during early embryogenesis (Adam et al., 2020). They are termed as recurrent constellation of embryonic malformations (RCEM) and constitute a spectrum and are likely to be causally related. We hypothesize that the URSMS also likely belongs to RCEM based on the overlapping features with LMBW, MURCS, and VACTERL. The highlight of our work is documentation of some of the rare malformations associated with URSMS such as cerebral ventriculomegaly, right aortic arch with double outlet right ventricle, fetal akinesia deformation sequence, amelia and limb contractures with spilt foot that are not described earlier, to the best of our knowledge.

Many dysmorphic features and contractures could be caused by oligohydramnios. Antenatally oligohydramnios or anhydramnios was noted in four fetuses (Fetuses 2, 3, 7, and 10) in our cohort. Fetus 2 had contractures of all major limb joints; Fetus 10 had contractures of limb joint along with short thigh, whereas facial dysmorphism was observed in Fetuses 3 and 10, which may be secondary to oligohydramnios.

One of the twins from a monochorionic diamniotic pregnancy (Fetus 9) had URSMS and associated malformations. Twinning with URSMS was noted previously in the literature (Achiron et al., 2000). The familial occurrence of URSMS is very rare and has been reported in two families (Aggarwal & Phadke, 2013; Mills & Pergament, 1997). One of them had a dominant inheritance, where the mother had mild

manifestations and the child had severe URSMS (Mills & Pergament, 1997), whereas the other family had prune belly in the first sib and URSMS in the second sib suggesting a recessive pattern of inheritance (Aggarwal & Phadke, 2013). These observations suggest the existence of a (mono) genetic etiology, at least for a proportion of patients with URSMS.

The cloaca is an embryonic structure evident by the fourth week of development and is divided to form urogenital sinus and primitive rectum by sixth week of intrauterine life. Hence, we speculate the genes involved in the early development of cloaca—Sonic hedgehog (Shh), SOX2 and CDX2, which express in the cloacal epithelium; Six1 and Ey1 intrinsic regulator of mesenchyme surrounding cloaca; Gli2 and Gli3 downstream mediator of Shh and BMP signaling are likely to play a role in pathophysiology (Runck et al., 2014).

The *TMEM132A* missense variant identified in fetus 1 may be relevant to the persistent cloaca observed in this fetus. *Tmem132a*<sup>−/−</sup> mice have impaired hind limb growth and spina bifida in addition to bladder and kidney defects (Li et al., 2022). Whole embryo staining of *Tmem132a*<sup>−/−</sup> mice revealed *Cdx2* is notably downregulated in the hindgut (Li et al., 2022). The importance of *CDX2* signaling in the normal development of the cloacal epithelium is supported by the identification of damaging de novo variants in *CDX2* in two unrelated patients with persistent cloaca (Hsu et al., 2018). Recently, variants in *CDX2* have also been reported in two familial aggregations with phenotype ranging from sirenomelia with different degrees of urogenital malformations to isolated imperforate anus indicating its role in caudal malformations (Lecoquierre et al., 2020). Moreover, the identification of a missense variant in *HOXD9* in Fetus 8 is also relevant to the *Shh*/*CDX2* pathway. Reduction in *Hoxd9* expression observed in embryonic limb buds of rats supplemented with all-trans-retinoic acid correlated with impaired *Shh*/*Gli3* signaling and reduced *Sox9*/*Col2a11* signaling (Hong et al., 2021). Both *Shh* (Seifert et al., 2009) and *Gli3* (Mo et al., 2001) expression are reduced in rats with anorectal malformations (Liu et al., 2016). Additionally, a disorganization-like syndrome was noted with persistent cloaca, having an incomplete duplication of left lower limb (Al Kaissi et al., 2008), suggesting an overlap in signal defects can cause deficiencies in both tissues. The normal limb development observed in Fetus 8 suggests the *HOXD9* c.277A > G variant would only urorectal development would be affected in this case, despite the association of *HOXD9* with limb defects.

Moreover, The *SLIT2* variant identified in Fetus 8 may relate to the unilateral renal agenesis observed during fetal evaluation. *Slit2*<sup>−/−</sup> mice develop multiple ureters and fused dysplastic kidneys as a result of ureteric buds, which remain inappropriately connected to the nephric duct during development (Grieshammer et al., 2004). Furthermore, heterozygous missense *SLIT2* variants were identified in three unrelated patients from a cohort with congenital anomalies of the kidney and urinary tract (CAKUT), one patient with bilateral subcortical renal cysts, one patient with right multicystic dysplastic kidneys, and one patient with right renal agenesis (Hwang et al., 2015).

In Fetus 8, the identification of concomitant *HOXD9* and *SLIT2* missense variants prompts the hypothesis that the etiology of URSMS in this fetus is oligogenic. URSMS has been identified as a component in a few syndromes/phenotypes, which are listed in Online Mendelian

Inheritance in Man along with their genetic bases: persistent cloaca and prune belly syndrome with tricho-rhino-phalangeal syndrome type II (MIM#150230) associated with 8q interstitial deletion (Ramos et al., 1992); Fanconi anemia, complementation group A (MIM#227650) (Farrell et al., 1994); sacral agenesis with vertebral anomalies (MIM# 615709) due to biallelic variants in *TBXT* (Postma et al., 2014); Omphalocele-Exstrophy-Imperforate Anus-Spinal Defects (OEIS complex, MIM 258040) associated with 1p36 deletion (El-Hattab et al., 2010) and a heterozygous variants in Uroplakin 3A associated with renal dysplasia and persistent cloaca (Jenkins et al., 2005). The London Medical Database has following entities listed with exstrophy of cloaca: caudal duplication syndrome; caudal regression; diphallus plus; femoral duplication; frontonasal dysplasia-exstrophy of bladder or cloaca; Gollop-Wolfgang complex caused by *BHLHA9*; sirenomelia; urorectal septal defects and VATER association.

Several researchers explored the genetic basis of URSMS and related anomalies. Patients with anorectal malformations and VACTERL association or cloacal exstrophy were analyzed for variants in *FGF10* following a reproducible phenotype of urorectal defect in mice; however, the data obtained were not supportive (Krüger et al., 2008). Isochromosome 18q was observed in a fetus with congenital megacystis, cloacal dysgenesis sequence and median cleft lip and palate by karyotype (Chen et al., 1998). Array comparative genomic hybridization in 17 females with cloaca identified novel copy number variations (paternally inherited duplication on 16p13.2 and a de novo deletion on 1q32.1q32.3) in two patients (Harrison et al., 2014). The latter deletion included a plausible candidate, hedgehog acyltransferase (*HHAT*); however, Sanger sequencing of independent patients with cloacal defects did not identify any significant variants in this gene. Prenatal diagnosis by chromosomal microarray in a pregnancy suspected with URSMS detected paternally inherited 111.8Kb deletion at 16p13.3 in the affected fetus (Pei et al., 2016). Similarly, *ZNF157*, *SP8*, *ACOT9*, and *TLL11* are thought to be associated with VATER/VACTERL following exome sequencing in a cohort of 21 families (Kolvenbach et al., 2021). However, further studies are required to explain the causality.

We identified heterozygous missense *NALCN* variants in Fetuses 3 and 9. *NALCN* encodes the nonselective sodium leak channel, and homozygous knockout is incompatible with life in mice (Cochet-Bissuel et al., 2014). Previously, de novo *NALCN* variants were observed in 14 unrelated probands with congenital contractures of the limbs and face, hypotonia and developmental delay (CLIFAHDD, MIM#616266) (Chong et al., 2015). *NALCN* is primarily expressed in the nervous system, suggesting its role is in movement coordination and intellectual development, rather than tissue patterning and development (Kasap et al., 2017). Phenotypes in Fetuses 3 and 9 are do not feature contractures, while hypotonia and developmental delay were not discernible in this cohort. However, scoliosis, clubfoot, and omphalocele observed in Fetus 9 are reminiscent of the additional phenotypes observed among the CLIFAHDD cohort (Chong et al., 2015).

We assume that the presence of various malformations with URSMS represent clinically diverse spectrum of perinatal lethal syndromes with URSMS. Our observations call for a concerted effort for



delineation of clinically and likely etiologically heterogeneous conditions with URSMS. Genome sequencing and analysis did not identify variants in known URSMS genes, but it identifies novel candidate variants that provide the basis for future investigations. However, in some cases, no variant of interest was identified, which suggests that these fetuses may have a more complex oligo/polygenic etiology with environmental influences.

## ACKNOWLEDGMENTS

The authors thank the families for their participation and support. This work was supported by the Indian Council of Medical Research, Government of India (No. 5/13/58/2015/NCD-III) and DBT/Wellcome Trust India Alliance Grant titled 'Center for Rare Disease Diagnosis, Research and Training (Grant number: IA/CRC/20/1/600002). William G. Newman is supported by the Manchester NIHR BRC (IS-BRC-1215-20007).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

## ORCID

Shalini S. Nayak  <https://orcid.org/0000-0001-9521-7031>

Robert Harkness  <https://orcid.org/0000-0001-5951-2199>

Anju Shukla  <https://orcid.org/0000-0003-2471-4094>

William G. Newman  <https://orcid.org/0000-0002-6382-4678>

Katta M. Girisha  <https://orcid.org/0000-0002-0139-8239>

## REFERENCES

- Achiron, R., Frydman, M., Lipitz, S., & Zalel, Y. (2000). Urorectal septum malformation sequence: Prenatal sonographic diagnosis in two sets of discordant twins. *Ultrasound in Obstetrics & Gynecology*, 16(6), 571–574. <https://doi.org/10.1046/j.1469-0705.2000.00233.x>
- Adam, A. P., Curry, C. J., Hall, J. G., Keppler-Noreuil, K. M., Adam, M. P., & Dobyns, W. B. (2020). Recurrent constellations of embryonic malformations re-conceptualized as an overlapping group of disorders with shared pathogenesis. *American Journal of Medical Genetics. Part A*, 182(11), 2646–2661. <https://doi.org/10.1002/ajmg.a.61847>
- Aggarwal, S., & Phadke, S. R. (2013). Recurrence of urorectal septum malformation sequence spectrum anomalies in siblings: Time to explore the genetics. *American Journal of Medical Genetics. Part A*, 161A(7), 1718–1721. <https://doi.org/10.1002/ajmg.a.35950>
- Al Kaissi, A., Strobl, W., Bauer, J., Landauer, F., Klaushofer, K., & Grill, F. (2008). Persistent cloaca associated with a duplicated left leg: A novel disorganization-like syndrome. *Clinical Dysmorphology*, 17(2), 137–139. <https://doi.org/10.1097/MCD.0b013e3282efe244>
- Chen, C. P., Chern, S. R., Lee, C. C., & Town, D. D. (1998). Isochromosome 18q in a fetus with congenital megacystis, intra-uterine growth retardation and cloacal dysgenesis sequence. *Prenatal Diagnosis*, 18(10), 1068–1074.
- Chong, J. X., McMillin, M. J., Shively, K. M., Beck, A. E., Marvin, C. T., Armenteros, J. R., Buckingham, K. J., Nkinsi, N. T., Boyle, E. A., Berry, M. N., Bocian, M., Foulds, N., Uzielli, M. L., Haldeman-Englert, C., Hennekam, R. C., Kaplan, P., Kline, A. D., Mercer, C. L., Nowaczyk, M. J., ... University of Washington Center for Mendelian Genomics. (2015). De novo mutations in NALCN cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. *American Journal of Human Genetics*, 96(3), 462–473. <https://doi.org/10.1016/j.ajhg.2015.01.003>
- Cochet-Bissuel, M., Lory, P., & Monteil, A. (2014). The sodium leak channel, NALCN, in health and disease. *Frontiers in Cellular Neuroscience*, 8, 132. <https://doi.org/10.3389/fncel.2014.00132>
- El-Hattab, A. W., Skorupski, J. C., Hsieh, M. H., Breman, A. M., Patel, A., Cheung, S. W., & Craigen, W. J. (2010). OEIS complex associated with chromosome 1p36 deletion: A case report and review. *American Journal of Medical Genetics. Part A*, 152A(2), 504–511. <https://doi.org/10.1002/ajmg.a.33226>
- Farrell, S. A., Paes, B. A., & Lewis, M. E. (1994). Fanconi anemia in a child previously diagnosed as Baller-Gerold syndrome. *American Journal of Medical Genetics*, 50(1), 98–99. <https://doi.org/10.1002/ajmg.1320500123>
- Grieshammer, U., Le Ma, A. S., Plump, F., Wang, M. T.-L., & Martin, G. R. (2004). SLIT2-mediated ROBO2 signaling restricts kidney induction to a single site. *Developmental Cell*, 6(5), 709–717. [https://doi.org/10.1016/s1534-5807\(04\)00108-x](https://doi.org/10.1016/s1534-5807(04)00108-x)
- Harrison, S. M., Seideman, C., & Baker, L. A. (2014). DNA copy number variations in patients with persistent cloaca. *The Journal of Urology*, 191(5 Suppl), 1543–1546. <https://doi.org/10.1016/j.juro.2013.09.056>
- Heyroth-Griffis, C. A., Weaver, D. D., Faught, P., Bellus, G. A., & Torres-Martinez, W. (2007). On the spectrum of limb-body wall complex, exstrophy of the cloaca, and urorectal septum malformation sequence. *American Journal of Medical Genetics. Part A*, 143A(10), 1025–1031. <https://doi.org/10.1002/ajmg.a.31691>
- Hong, Q., Li, X. D., Xie, P., & Du, S. X. (2021). All-trans-retinoic acid suppresses rat embryo hindlimb bud mesenchymal chondrogenesis by modulating Hox D9 expression. *Bioengineered*, 12(1), 3900–3911. <https://doi.org/10.1080/21655979.2021.1940613>
- Hsu, J. S. J., So, M., Tang, C. S. M., Karim, A., Porsch, R. M., Wong, C., Yu, M., Yeung, F., Xia, H., Zhang, R., Cherny, S. S., Chung, P. H. Y., Wong, K. K. Y., Sham, P. C., Ngo, N. D., Li, M., Tam, P. K. H., Lui, V. C. H., & Garcia-Barcelo, M. M. (2018). De novo mutations in caudal type Homeo box transcription factor 2 (CDX2) in patients with persistent cloaca. *Human Molecular Genetics*, 27(2), 351–358. <https://doi.org/10.1093/hmg/ddx406>
- Huang, K. Y., Kuo, K. T., Li, Y. P., Chen, M., Yu, C. U., & Shih, J. C. (2016). Urorectal septum malformation sequence-fetal series with the description of a new "intermediate" variant. Time to refine the terminology? *American Journal of Medical Genetics. Part A*, 170(9), 2479–2482. <https://doi.org/10.1002/ajmg.a.37788>
- Hwang, D. Y., Kohl, S., Fan, X., Vivante, A., Chan, S., Dworschak, G. C., Schulz, J., van Eerde, A. M., Hilger, A. C., Gee, H. Y., Pennimpede, T., Herrmann, B. G., van de Hoek, G., Renkema, K. Y., Schell, C., Huber, T. B., Reutter, H. M., Soliman, N. A., Stajic, N., ... Hildebrandt, F. (2015). Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1 confer risk for congenital anomalies of the kidney and urinary tract. *Human Genetics*, 134(8), 905–916. <https://doi.org/10.1007/s00439-015-1570-5>
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J., Carpten, J. D., ... Sieh, W. (2016). REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, 99(4), 877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglu, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, 176(3), 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- Jain, D., Sharma, M. C., Kulkarni, K. K., Aggarwal, S., & Karak, A. K. (2008). Urorectal septum malformation sequence: A report of seven cases.

- Congenital Anomalies* (Kyoto), 48(4), 174–179. <https://doi.org/10.1111/j.1741-4520.2008.00200.x>
- Jenkins, D., Bitner-Glindzicz, M., Malcolm, S., Hu, C. C., Allison, J., Winyard, P. J., Gullett, A. M., Thomas, D. F., Belk, R. A., Feather, S. A., Sun, T. T., & Woolf, A. S. (2005). De novo Uroplakin IIIa heterozygous mutations cause human renal adysplasia leading to severe kidney failure. *Journal of the American Society of Nephrology*, 16(7), 2141–2149. <https://doi.org/10.1681/ASN.2004090776>
- Kasap, M., Bonnett, K., Aamodt, E. J., & Dwyer, D. S. (2017). Akinesia and freezing caused by Na. *The Journal of Comparative Neurology*, 525(5), 1109–1121. <https://doi.org/10.1002/cne.24119>
- Khong, T. Y., & Malcomson, R. D. G. (2015). *Keeling's fetal and neonatal pathology*. Springer.
- Kolvenbach, C. M., van der Ven, A. T., Kause, F., Shril, S., Scala, M., Connaughton, D. M., Mann, N., Nakayama, M., Dai, R., Kitzler, T. M., Schneider, R., Schierbaum, L., Schneider, S., Accogli, A., Torella, A., Piatelli, G., Nigro, V., Capra, V., Hoppe, B., ... Hildebrandt, F. (2021). Exome survey of individuals affected by VATER/VACTERL with renal phenotypes identifies phenocopies and novel candidate genes. *American Journal of Medical Genetics. Part A*, 185(12), 3784–3792. <https://doi.org/10.1002/ajmg.a.62447>
- Krüger, V., Khoshvaghti, M., Reutter, H., Vogt, H., Boemers, T. M., & Ludwig, M. (2008). Investigation of FGF10 as a candidate gene in patients with anorectal malformations and exstrophy of the cloaca. *Pediatric Surgery International*, 24(8), 893–897. <https://doi.org/10.1007/s00383-008-2193-x>
- Lecoquiere, F., Brehin, A. C., Coutant, S., Coursimault, J., Bazin, A., Finck, W., Benoist, G., Begorre, M., Beneteau, C., Cailliez, D., Chenal, P., De Jong, M., Degré, S., Devisme, L., Francannet, C., Gérard, B., Jeanne, C., Joubert, M., Journel, H., ... Gerard, M. (2020). Exome sequencing identifies the first genetic determinants of sirenomelia in humans. *Human Mutation*, 41(5), 926–933. <https://doi.org/10.1002/humu.23998>
- Li, B., Brusman, L., Dahlka, J., & Niswander, L. A. (2022). TMEM132A ensures mouse caudal neural tube closure and regulates integrin-based mesodermal migration. *Development*, 149(17). <https://doi.org/10.1242/dev.200442>
- Liu, Z. H., Li, E. H., Xu, D. L., Sun, W. L., Hong, Y., Zhao, W., Xia, S. J., & Jiang, J. T. (2016). Genetic research and structural dysplasia assessment of anorectal malformations in neonatal male rats induced by di(n-butyl) phthalate. *Environmental Toxicology*, 31(3), 261–268. <https://doi.org/10.1002/tox.22040>
- Mills, P. L., & Pergament, E. (1997). Urorectal septal defects in a female and her offspring. *American Journal of Medical Genetics*, 70(3), 250–252.
- Mo, R., Kim, J. H., Zhang, J., Chiang, C., Hui, C. C., & Kim, P. C. (2001). Anorectal malformations caused by defects in sonic hedgehog signaling. *The American Journal of Pathology*, 159(2), 765–774. [https://doi.org/10.1016/S0002-9440\(10\)61747-6](https://doi.org/10.1016/S0002-9440(10)61747-6)
- Nayak, S. S., Shukla, A., Lewis, L., Kadavigere, R., Mathew, M., Adiga, P. K., Vasudeva, A., Kumar, P., Shetty, J., Shah, H., & Girisha, K. M. (2015). Clinical utility of fetal autopsy and its impact on genetic counseling. *Prenatal Diagnosis*, 35(7), 685–691. <https://doi.org/10.1002/pd.4592>
- Osborn, M. I., & Lowe, J. (2017). *Guidelines on autopsy practice: Fetal autopsy (2nd trimester fetal loss and termination of pregnancy for congenital anomaly)*. The Royal College of Pathologists.
- Patil, S. J., & Phadke, S. R. (2006). Urorectal septum malformation sequence: Ultrasound correlation with fetal examination. *Indian Journal of Pediatrics*, 73(4), 287–293. <https://doi.org/10.1007/BF02825821>
- Pei, Y., Wu, Q., Liu, Y., Sun, L., Zhi, W., & Zhang, P. (2016). Prenatal sonographic diagnosis of urorectal septum malformation sequence and chromosomal microarray analysis: A case report and review of the literature. *Medicine* (Baltimore), 95(45), e5326. <https://doi.org/10.1097/MD.00000000000005326>
- Postma, A. V., Alders, M., Sylva, M., Bilardo, C. M., Pajkrt, E., van Rijn, R. R., Schulte-Merker, S., Bulk, S., Stefanovic, S., Ilgun, A., Barnett, P., Mannens, M. M., Moorman, A. F., Oostra, R. J., & van Maarle, M. C. (2014). Mutations in the T (brachyury) gene cause a novel syndrome consisting of sacral agenesis, abnormal ossification of the vertebral bodies and a persistent notochordal canal. *Journal of Medical Genetics*, 51(2), 90–97. <https://doi.org/10.1136/jmedgenet-2013-102001>
- Ramos, F. J., McDonald-McGinn, D. M., Emanuel, B. S., & Zackai, E. H. (1992). Tricho-rhino-phalangeal syndrome type II (Langer-Giedion) with persistent cloaca and prune belly sequence in a girl with 8q interstitial deletion. *American Journal of Medical Genetics*, 44(6), 790–794. <https://doi.org/10.1002/ajmg.1320440614>
- Runck, L. A., Method, A., Bischoff, A., Levitt, M., Peña, A., Collins, M. H., Gupta, A., Shanmukhappa, S., Wells, J. M., & Guasch, G. (2014). Defining the molecular pathologies in cloaca malformation: Similarities between mouse and human. *Disease Models & Mechanisms*, 7(4), 483–493. <https://doi.org/10.1242/dmm.014530>
- Seifert, A. W., Bouldin, C. M., Choi, K. S., Harfe, B. D., & Cohn, M. J. (2009). Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development. *Development*, 136(23), 3949–3957. <https://doi.org/10.1242/dev.042291>
- Shah, K., Nayak, S. S., Shukla, A., & Girisha, K. M. (2016). Spectrum of urorectal septum malformation sequence. *Congenital Anomalies* (Kyoto), 56(3), 119–126. <https://doi.org/10.1111/cga.12149>
- Tennant, P. W., Glinianaia, S. V., Wellesley, D., Draper, E. S., Kurinczuk, J. J., Tonks, A. M., Tucker, D. F., Wreyford, B., & Rankin, J. (2014). Epidemiology of partial urorectal septum malformation sequence (or persistent cloaca): A population-based study in seven regions of England and Wales, 1985–2010. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 99(5), F413–F418. <https://doi.org/10.1136/archdischild-2014-306027>
- Vauthay, L., Mazzitelli, N., & Rittler, M. (2007). Patterns of severe abdominal wall defects: Insights into pathogenesis, delineation, and nomenclature. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 79(3), 211–220. <https://doi.org/10.1002/bdra.20339>
- Wheeler, P. G., & Weaver, D. D. (2001). Partial urorectal septum malformation sequence: A report of 25 cases. *American Journal of Medical Genetics*, 103(2), 99–105. <https://doi.org/10.1002/ajmg.1510>
- Wheeler, P. G., Weaver, D. D., Obeime, M. O., Vance, G. H., Bull, M. J., & Escobar, L. F. (1997). Urorectal septum malformation sequence: Report of thirteen additional cases and review of the literature. *American Journal of Medical Genetics*, 73(4), 456–462.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Nayak, S. S., Harkness, R., Shukla, A., Banka, S., Newman, W. G., & Girisha, K. M. (2022). Clinically diverse and perinatally lethal syndromes with urorectal septum malformation sequence. *American Journal of Medical Genetics Part A*, 1–12. <https://doi.org/10.1002/ajmg.a.63067>

## **Chapter 5: Deep Intronic Variant Causes Aberrant Splicing Of *ATP7A* In A Family With A Variable Occipital Horn Syndrome Phenotype**

J. Robert Harkness<sup>1,2</sup>, Huw B. Thomas<sup>2</sup>, Helen M. Kingston<sup>1</sup>, Jill E. Urquhart<sup>1,2</sup>, Genomics England Research Consortium, Simon Hubbard<sup>2</sup>, Raymond T. O’Keefe<sup>2</sup>, Siddharth Banka<sup>1,2</sup>, William G. Newman<sup>1,2</sup>, Charu Deshpande<sup>1</sup>

<sup>1</sup> Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK.

<sup>2</sup> Division of Evolution, Infection and Genomics, Faculty of Biology, Medicine and Health Sciences, University of Manchester, Manchester, UK.

Correspondence:

Professor William G. Newman

Manchester Centre for Genomic Medicine,  
Manchester University NHS Foundation Trust,  
Manchester, M13 9WL

UK.

Telephone 00 44 161 276 4150

Email [william.newman@manchester.ac.uk](mailto:william.newman@manchester.ac.uk)

Key words. *ATP7A*, deep intronic variant, non-coding, splicing, genome, Menkes disease, occipital horn syndrome, rare disease

## 5.1 Abstract

Rare variants in *ATP7A* are associated with a spectrum of X-linked disorders. In descending order of severity, these are Menkes disease, occipital horn syndrome, and X-linked distal spinal muscular atrophy. After 30 years of diagnostic investigation, we identified a deep intronic *ATP7A* variant in four males from a family affected to variable degrees by a predominantly skeletal phenotype, featuring bowing of long bones, elbow joints with restricted mobility which dislocate frequently, coarse curly hair, chronic diarrhoea, and motor coordination difficulties. Analysis of whole genome sequencing data from the Genomics England 100,000 Genomes Project following clinical re-evaluation identified a deep intronic *ATP7A* variant, which was predicted by SpliceAI to have a modest splicing effect. Using a mini-gene splicing assay, we determined that the intronic variant results in aberrant splicing. Sanger sequencing of patient cDNA revealed *ATP7A* transcripts with exon 5 skipping, or inclusion of a novel intron 4 pseudoexon. In both instances, frameshift leading to premature termination are predicted. Quantification of *ATP7A* mRNA transcripts using a qPCR assay indicated that the majority of transcripts (86.1 %) have non-canonical splicing, with 68.0 % featuring exon 5 skipping, and 18.1 % featuring the novel pseudoexon. We suggest that the variability of the phenotypes within the affected males results from the stochastic effects of splicing. This deep intronic variant, resulting in aberrant *ATP7A* splicing, expands the understanding of intronic variation on the *ATP7A*-related disease spectrum.

## 5.2 Introduction

Menkes disease (MIM 309400) is a rare X-linked copper metabolism disease caused by pathogenic variants in *ATP7A* (MIM 300011). Incidence rates of Menkes disease vary with conservative estimates suggesting 1 in 300,000 affected live births (Tønnesen et al., 1991; Vairo et al., 2019). *ATP7A* encodes the ubiquitously expressed Cu<sup>2+</sup>-transporting ATPase, a transmembrane protein which has dual functions as an ATP-dependent Cu<sup>2+</sup> efflux channel, and a biomolecular loader of copper-dependent enzymes (cuproenzymes) in the trans-Golgi network (TGN) (Møller, 2015).

Menkes disease is the most clinically severe disease on an allelic spectrum of *ATP7A*-related disorders, which includes occipital horn syndrome (OHS; MIM 304150), and X-linked distal

spinal muscular atrophy type 3 (SMA3; MIM 300489) (Fradin et al., 2020; Zlatić et al., 2015). The severity of disease largely correlates to the deleteriousness of the causal *ATP7A* variant (Mhaske et al., 2020). Nonsense variants and large deletions correlate with an absence of functional protein, and cause Menkes disease, which manifest from the first few months of life. Menkes disease is primarily regarded as a neurodevelopmental disorder, featuring moderate to severe intellectual disability, epilepsy, and progressive neurodegeneration (Kaler and DiStasio, 1993; Rangarh and Kohli, 2018). Additionally, connective tissue anomalies, such as restricted bone growth, osteoporosis, and joint hyperflexibility, and ectodermal changes, including pili torti (kinky hair) and cutis laxa, and bladder diverticula are common from childhood. Patients may also experience chronic diarrhoea. If untreated, death usually occurs by three years of age. Early treatment with copper complexes such as copper-histidine offers the possibility to modify the onset of the severest phenotypes and extend life expectancy into adulthood (Christodoulou et al., 1998; Nadal and Baerlocher, 1988; Tümer et al., 2017; Vairo et al., 2019).

Less functionally deleterious variants such as missense variants and small inframe deletions, which permit residual *ATP7A* protein function, result in less-severe phenotypes associated with OHS (Dagenais et al., 2001; Møller, 2015). OHS features are similar to Menkes disease, but with absent or mild neurodegenerative changes, while non-neurological phenotypes can vary in severity (Kaler and DiStasio, 1993; Skjørringe et al., 2017). Typically described are occipital horns, which are wedge-shaped calcifications at the site of attachment of the trapezius and sternocleidomastoid muscles to the occipital bone (Gérard-Blanluet et al., 2004). Connective tissue anomalies are commonly described, alongside dysautonomia, which may cause chronic diarrhoea, syncope, and orthostatic hypotension. Early treatment of OHS patients with copper-histidine can reduce developmental anomalies associated with connective tissues to reduce the burden of disease (Vairo et al., 2019).

Finally, specific missense variants in *ATP7A* gene are associated with late onset distal spinal muscular atrophy (SMA3). This form of the disease features slowly progressive distal limb weakness and muscle wasting (Kaler and DiStasio, 1993; Takata et al., 2004), with occasional description of minor connective tissue anomalies (Gualandi et al., 2019). Although four different *ATP7A* variants (Gualandi et al., 2019; Kennerson et al., 2010; Shibuya et al., 2022)

have been reported in cases of SMAX3, these all cluster in the terminal portion of the gene. Serum copper and caeruloplasmin are usually normal or slightly reduced in SMAX3 patients, suggesting that *ATP7A* variants causing SMAX3 are the mildest on the *ATP7A*-related phenotype spectrum (Fradin et al., 2020).

The function of *ATP7A* as a copper transporter is essential for copper dietary absorption via duodenal enterocytes into the portal circulation, and across the blood-brain barrier into the central nervous system (CNS) (Lutsenko et al., 2007; Maung et al., 2021). Inactivity of the *ATP7A* copper transporter therefore results in severe copper deficiency which particularly affects the brain, and accounts for the severest neurological phenotypes in Menkes disease (Zlatic et al., 2015). *ATP7A*-related copper deficiency impairs the activity of cuproenzymes, including cytochrome c oxidase (respiratory electron transport chain), and superoxide dismutase (maintenance of redox homeostasis), the dysfunction of which contribute to neurodegeneration identified in severely affected Menkes disease patients. At low copper concentrations, *ATP7A* localises to the trans-Golgi network, where it has a secondary function in the direct loading of copper to specific cuproenzymes. These cuproenzymes include lysyl oxidase (LOX) and lysyl oxidase-like (LOXL) proteins involved in collagen and elastin crosslinking (Trackman, 2018), copper-binding monooxygenases which contribute to catecholamine synthesis, and copper-dependent tyrosinases essential for biogenesis of signalling molecules such as melanin (Lutsenko et al., 2007). The reduced activity of LOX/LOXL proteins underlies the connective tissue manifestations of Menkes disease and OHS, such as cutis laxa and joint hyperflexibility (Horn and Wittung-Stafshede, 2021). Dysregulated trafficking of *ATP7A* between the TGN and the plasma membrane during high copper concentrations may also lead to dysfunction in copper homeostasis, rather than direct impairment of *ATP7A* copper transport (Skjørringe et al., 2017).

Here, we present four males from a single family affected by OHS with a novel deep intronic *ATP7A* variant, which leads to mis-splicing in a majority of *ATP7A* transcripts, leading to OHS. Phenotypes vary in severity between individuals, likely owing to the leaky effect of altered splicing.

## 5.3 Results:

### 5.3.1 Clinical features

A three-week old male (Figure 1A, IV:2) was referred to the paediatric clinic due to bilateral inguinal herniae. At the initial assessment, it was noted that he was failing to thrive and had diarrhoea. A sweat test performed at five weeks of age was reported as positive. His brother, IV:1 was assessed at the same time and in view of the similar history of chronic diarrhoea and failure to thrive, was investigated for cystic fibrosis. Both brothers had abnormal fecal elastase levels and although a molecular diagnosis of cystic fibrosis was not made, both patients responded to pancreatic enzyme replacement therapy. Their mother (III:6) had noticed that both boys had similar bony problems and had an unusual appearance of their elbows. Their maternal female cousin, IV:5, has two sons with similar appearances of elbows (V:2 and V:3, Figure 1B-E) and a summary of the clinical features is presented in Table 1 .

The possibility of an underlying skeletal dysplasia was considered in view of the appearance of their elbows, pectus excavatum in IV:2 and V:2, and a skeletal survey noting flared metaphyses and that the bone age was delayed by 2 years compared to chronological age for both boys. In view of the chronic diarrhoea and skeletal anomalies, a differential diagnosis of Schwachman-Diamond syndrome was considered, but was not confirmed by genetic testing.

IV1 and IV:2 both remain on pancreatic enzyme replacement supplements. IV:1 has required revision of his inguinal herniae repair due to recurrence. Joint hypermobility impacts on his fine motor skills as interphalangeal joints lock and subluxate easily. There is a history of recurrent knee dislocation and all affected males have needed multiple surgery of the lower limbs. Notably, none of the affected males have occipital protuberances evident on clinical examination.

The maternal uncle of the proband (III:9) is deceased and did not pursue clinical assessment, so limited clinical details are available, but was reported to have similar problems with diarrhoea and with his joints to affected males in his family. Three females in the family

(III:6, III:12, IV:7) also had prominent ribs or prominent elbows, traits which the family associates with the phenotypes in the affected males.

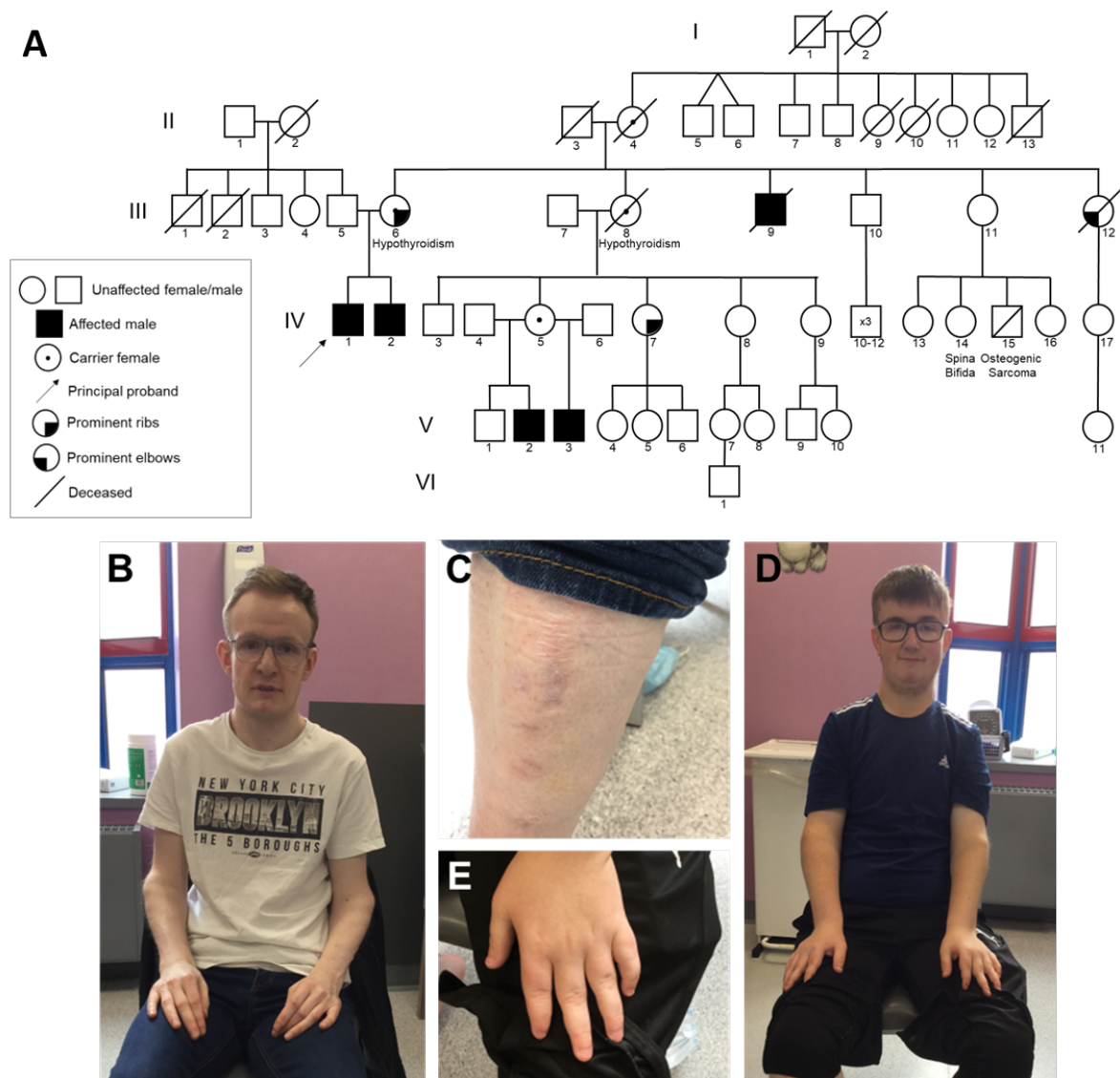


Figure 1: Family with X-linked skeletal dysplasia. (A) Pedigree of large family with five affected males (III:9, IV:1, IV:2, V:2 and V:3). (B) Photograph of V:2 indicating broad forehead with frontal prominences, and prominent elbows unable to pronate/supinate. (C) Cigarette paper scar on lower leg of V:2. (D) Photograph of V:3 indicating narrow thorax, and prominent elbows unable to pronate/supinate. (E) Left hand of V:3 indicating tapered fingers.



Table 1: Phenotypes observed in five related individuals. CPK, Creatine Phosphokinase. EMG, Electromyography. LFT, Liver Function Test. NAD, Nicotinamide Adenine Dinucleotide. NCV, Nerve Conduction Velocity. OFC, Occipital Frontal Circumference. VLCFA, Very Long Chain Fatty Acids.

Pedigree ID	IV:1	IV:2	V:2	V:3
Biometry:				
Age (years) at last assessment	43	39	25	20
Height centile	10 <sup>th</sup>	9-25 <sup>th</sup>	50 <sup>th</sup>	50 <sup>th</sup>
Weight centile	ND	<0.4 <sup>th</sup>	50 <sup>th</sup>	25 <sup>th</sup>
OFC	ND	ND	50 <sup>th</sup>	99 <sup>th</sup>
Birth weight (g)	2637	2098	3629	3289
Ectodermal:				
Skin	Soft, thin, elastic	Soft, thin, elastic	Soft, thin, elastic	Soft, thin, elastic
Papyraceous scars	Yes	Yes	Yes	Yes
Keratosis pilaris	Yes	Yes	Yes	No
Teeth	Deciduous teeth retained	Extra teeth	Normal	Unknown
Hair	Coarse	Sparse, coarse	Sparse, wiry	Coarse, wiry
Inguinal hernia	Bilateral	Bilateral	None	Right
Skeletal:				
Bone age	BA3 CA6	BA3 CA5.5	BA4 CA5	ND

Skull	Bony lumps – temporal bone and over coronal structure	Normal	Broad forehead, 2 frontal prominences	2 calcified cephalo-heamatomas, unilateral skull prominence
Narrow thorax	Yes	Yes	No	Yes
Pectus excavatum	No	Yes	Yes	No
Winged scapulae	Yes	Yes	No	Yes
Prominent elbows	Yes	Yes	Yes	Yes
Dislocated radial head	Yes	Yes	Yes	Yes
Bowed forearm	Yes	Yes	Developing	No
Tapered fingers	Yes	Yes	No	Yes
Genu valgum	Yes	Yes	Yes	Yes
Talipes	Yes	Yes	Yes	Yes
Frequent joint subluxation	Knees, wrists, phalanges	Knees	Knees	Knees
Other:				
Intellectual disability	No	No	Mild	Mild
Thin musculature	Yes	Yes	No	No
Diarrhoea	Yes	Yes	Yes	Yes
Pancreatic insufficiency	Yes	Yes	Unknown	Unknown

Investigations:				
Chromosomes	ND	46, XY	46, XY	ND
Creatinine kinase	ND	Normal	ND	ND
LFTs	ND	Normal	ND	ND
VLCFA	ND	Normal	ND	ND
Serum copper (reference)	ND	ND	10.9 µmol/L (13-26)	ND
Caeruloplasmin (reference)	ND	ND	137 mg/L (200-600)	ND
White cell enzyme	ND	Normal	ND	ND
Pancytopenia	ND	No	ND	ND
Neutropenia	ND	No	ND	ND
Anaemia	ND	No	ND	ND
Thrombocytopenia	ND	No	ND	ND
Muscle biopsy	ND	Normal	ND	ND
EMG/NCV	ND	Normal	ND	ND
Sweat test/CF pathogenic variant	ND	Normal/none	ND	ND
Treatments	Pancreatic enzyme replacement; Copper-histidine supplementation	Pancreatic enzyme replacement		

### 5.3.2 Genetic analysis

Linkage analysis of Affymetrix SNP array data from IV:1 and V:2 revealed three shared regions totalling 25.6 Mb (GRCh37; X:2,694,240-6,691,422 at Xp22.33-p22.31; X:53,887,456-58,107,416 at Xp11.22-p11.21 and X:69,445,677-86,811,737 at Xq13.1-q21.3). No copy number variants of note were identified. Exome sequencing analysis of IV:1 revealed several rare coding variants within the linked regions, but no compelling candidate (Supplemental data, Appendix III). Alongside his affected brother (IV:2), and their mother (III:6), IV:1 was recruited to the 100,000 Genomes Project for whole genome sequencing (WGS) analysis. Review of the clinical phenotype at this time noted the wiry hair and the possibility of an *ATP7A* related phenotype. Rare ( $< 0.01$ ) coding and non-coding variants present at the *ATP7A* locus in the three individuals were filtered according to the expected X-linked mode of inheritance. One deep intronic *ATP7A* variant c.1544-872C>G (GRCh38 X:78002201C>G), within the previously linked region, was identified. Segregation analysis by Sanger sequencing (Figure 2A) confirmed hemizyosity of the c.1544-872C>G variant among the four affected males on whom samples were available (IV:1, IV:2, V2, and V3), and heterozygosity in their respective mothers (III:6 and IV:5). The effect of the variant on splicing was modelled using SpliceAI (Jaganathan et al., 2019), which predicted the introduction of splice acceptor and donor sites 106 bp and 5 bp upstream of the variant, respectively (Figure 2B). This modelling suggests the c.1544-872C>G variant may introduce of a pseudoexon in *ATP7A*.

### 5.3.3 Intronic *ATP7A* variant results in altered splicing

Minigene assays were performed, as described previously (Thomas et al., 2020), to determine the effect of the intronic variant on splicing activity. Minigene vectors were constructed with an *ATP7A* fragment from IV:1 genomic DNA, or the wild type (WT) allele from the proband's mother (IV:6). The *ATP7A* fragment contained the region from exon 4 to exon 5 of *ATP7A*, plus 200 bp flanking sequence. HEK293 cells were transfected with minigene-*ATP7A* vectors and RNA was extracted. Evaluation of reverse transcribe minigene assay products via gel electrophoresis indicated that proband *ATP7A*-minigene vector transcription products were larger compared to the WT vector (Figure 2C). Sanger

sequencing of these products confirmed the inclusion of the predicted novel ~100 bp pseudoexon (Figure 2D).

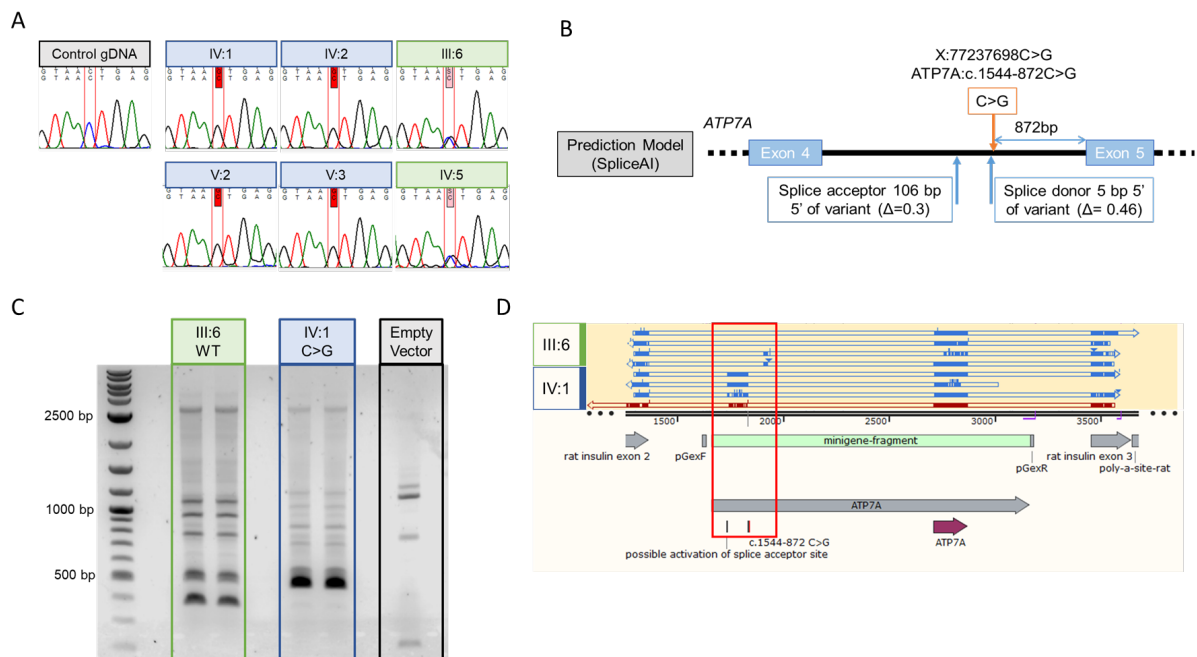


Figure 2: Evaluation of splicing changes from the *ATP7A* c.1544-872C>G intronic variant. (A) Sanger sequencing showing *ATP7A* c.1544-872C>G variant in affected individuals, which confirm the expected X-linked segregation. (B) Diagram of SpliceAI *in silico* prediction. (C) Image of minigene assay reverse transcription product gel electrophoresis indicating larger transcripts are generated from the vector with the *ATP7A* c.1544-872C>G variant. (D) Alignment of minigene product sequencing alignment using SnapGene software, red box highlights alignment of pseudoexon in an assay with the *ATP7A* c.1544-872C>G variant.

#### 5.3.4 Proband *ATP7A* RNA analysis

To validate these splicing changes, RNA was extracted from blood lymphocytes derived from V:2. RNA was reverse transcribed and exon 4-6 of *ATP7A* cDNA was sequenced to confirm presence of the novel *ATP7A* pseudoexon. We identified transcription products of two sizes, indicating the novel splice variant has a leaky effect (Figure 3A). The larger fragment contained the predicted pseudoexon between exons 4 and 5. The smaller fragment did not contain the pseudoexon, but indicated skipping of exon 5 (Figure 3B). Inclusion of the novel pseudoexon is predicted to create a frameshift p.(Ile515Serfs\*26). Similarly, skipping of exon 5 is predicted to create a different frameshift p.(Ile515Glufs\*11) (Figure 3C). In both cases, premature termination at residue 526 or 541 is predicted to result in truncated isoform of the *ATP7A* protein, consisting of only of the copper-binding heavy-metal associated domains within the N-terminal cytoplasmic domain (aa1-653) with absent transmembrane domains (aa654-1406) and C-terminal cytoplasmic domain (aa1407-1500). These truncated products are therefore predicted to be either non-functional for copper transportation across plasma membranes or targeted for nonsense-mediated decay.

In order to establish the proportion of *ATP7A* transcripts disrupted by splicing, we performed quantitative PCR. TaqMan probes specific to *ATP7A* cDNA exon boundaries 1-2, 4-5, and 5-6 were used to quantify the relative proportion of transcripts which contained either the novel pseudoexon, or exon 5 skipping. Probes binding across exon boundaries at exon 4-5 and exon 5-6 were compared to total *ATP7A* transcripts (exon 1-2 boundary) to quantify the relative proportion of transcripts affected by mis-splicing (Figure 4A). In V:2 lymphocyte derived RNA, qPCR analysis determined 13.9 % of all *ATP7A* transcripts were canonical, with 68.0 % of transcripts featuring exon 5 skipping and 18.1 % transcripts featuring the modelled pseudoexon between exons 4 and 5 (Figure 4B and 4C). This suggests the majority of *ATP7A* transcripts in lymphocyte-derived RNA have splicing altered by the *ATP7A* c.1544-872C>G variant, which is likely to reduce global *ATP7A* function.

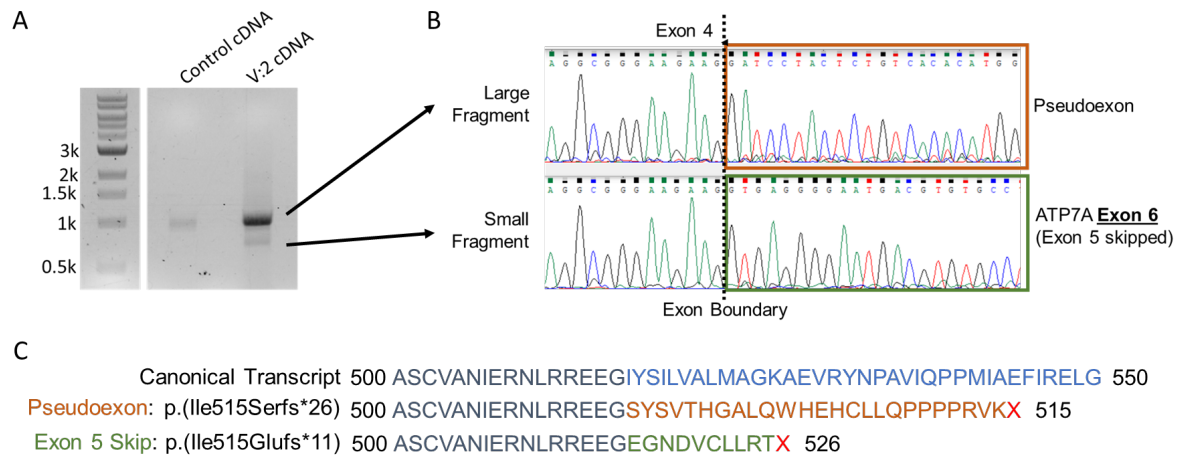


Figure 3: Analysis of *ATP7A* transcripts from V:2 lymphocyte-derived reverse transcribed RNA. (A) Electrophoresis of products from PCR of *ATP7A* exon 4-6 in IV:2 cDNA indicates two products are amplified. (B) Sanger sequencing traces of two PCR products from V:2. (C) Predicted effect of splicing changes on *ATP7A* translation products.

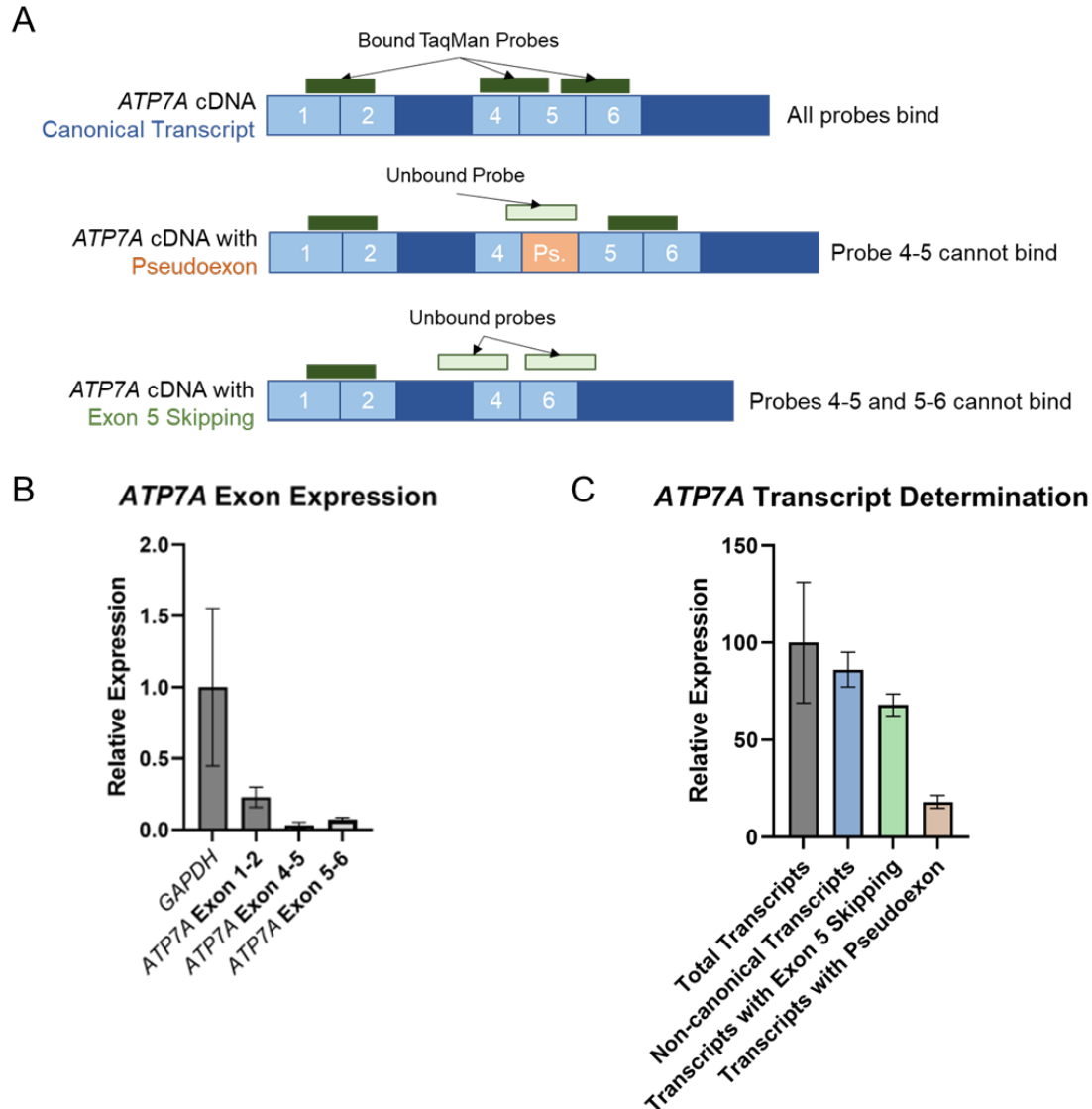


Figure 4: Quantitative PCR of *ATP7A* transcripts from V:2 lymphocyte cDNA. (A) Design of assay using three TaqMan probes specific to *ATP7A* exon boundaries to quantify proportions of transcripts affected by splicing. (B) QPCR of *ATP7A* exons in P3-derived lymphocyte cDNA, relative to *GAPDH*. (C) Determination of proportion of *ATP7A* transcripts with pseudoexon and exon 5 skipping relative to total *ATP7A* transcripts. *ATP7A* and *GAPDH* expression was quantified from a healthy control using a standard curve by plotting log cDNA concentration against Ct, in two experiments each with three triplicate reactions.



## 5.4 Discussion:

Here, we describe four males with a homozygous novel intronic variant in *ATP7A*. Despite the variable phenotypes the four males with normal intellect are each affected by dislocated elbows, genu valgum, soft stretchy and poor healing skin with papyraceous scars and wiry, coarse hair. Some of the males also have a narrow thorax, inguinal herniae, keratosis pilaris, pectus excavatum, and chronic diarrhoea. These clinical features plus biochemical evidence of low serum copper and caeruloplasmin levels are consistent with a diagnosis of OHS. Notably, pancreatic insufficiency is not described as a feature of the *ATP7A*-related disease spectrum. No alternative explanation has been determined for this feature and should be screened for in other individuals determined to have *ATP7A*-related disease.

The majority of variants identified in *ATP7A* have been identified in patients with Menkes disease. Following variant classification analysis based on ACMG guidelines, of the pathogenic and likely pathogenic variants in *ATP7A*, 89% were identified in patients with Menkes disease, with just 4 % in OHS (Mhaske et al., 2020). This difference likely results from a clinical ascertainment bias against the less severe OHS phenotype. There are similar proportions of patients with missense, nonsense, indel, and splice site variants (Møller, 2015), however very few deep intronic variants have been identified in patients consistent with *ATP7A*-related disease to date. We have identified a family affected by OHS with a deep intronic variant in *ATP7A*. This intronic variant introduces a novel splicing donor site and results in mis-splicing of a majority of *ATP7A* transcripts, either by skipping exon 5 or by inclusion of a novel pseudoexon. In both cases, transcripts are predicted to cause frameshift likely results in nonsense mediated decay.

The first patients to receive a molecular diagnosis of OHS due to deep intronic *ATP7A* variants were identified in a cohort in Denmark (Yasmeen et al., 2014). Intronic variants in intron 10 (c.2406+1117A>G), intron 14 (c.2916+2480T>G) and intron 16 (c.2916+2480T>G) were identified by performing overlapping RT-PCR reactions of the entire cDNA *ATP7A* sequence. In each case, the novel intronic variant led to inclusion of a pseudoexon into *ATP7A* transcripts and creation of a premature stop codon. Yasmeen et al., (2014) proposed

that severity of phenotypes correlated with the relative proportion of transcripts containing the pseudoexon. In the most severely affected Menkes disease patient, just 0.2 % of *ATP7A* transcripts were canonical. The proportions of canonical transcripts in the less severely affected patients were 11 % and 2 %, suggesting that the proportion of canonical transcripts required to prevent the onset of the severest phenotypes lies below a 2 % threshold. This hypothesis was supported by the subsequent investigation of five patients with splice-site variants leading to exon skipping (Møller et al., 2021). In one patient with Menkes disease, 0.5 % of transcripts were canonical, which increased to 3.5 % canonical transcripts in patients with OHS. These findings suggest the relative strength of the non-coding variant resulting in splicing changes is important in predicting phenotypic severity.

Analysis of RNA from V:2 showed a higher proportion of canonical *ATP7A* transcripts (13.9 %) compared to the cases reported by Yasmeen et al. (2014) and Moeller et al. (2021), with the majority of the non-canonical transcripts featuring exon 5 skipping, rather than pseudoexon inclusion, as a result of the *ATP7A* c.1544-872C>G variant. The milder *ATP7A* related phenotype in the affected males presented here is consistent with this higher level of the canonical transcript. The analysis of transcripts from lymphocyte-derived RNA is one limitation of our analysis, as lymphocyte expression of *ATP7A* is lower compared other cell types more relevant to the connective tissue phenotype, such as fibroblasts.

The phenotypic variability of patients analysed here can be attributed to the inherent 'leakiness' of splicing, a feature of the stochastic variation resulting from competition between acceptor sites of the pseudoexon and the downstream exon in affinity for binding of the spliceosome (Anna and Monika, 2018; Chin et al., 2022). This phenomenon has been characterized previously in other diseases, such as disease severity in cystic fibrosis in patients correlating with the degree of intron 19 pseudoexon inclusion resulting from the *CFTR* c.3718-2477C>T variant (Sobczyńska-Tomaszewska et al., 2013). The family reported here have a progressive disease, with onset in early childhood. Changes in the proportion of transcripts affected by mis-splicing in different tissue types over time may account for differences in onset and phenotypic variability in this family. Additionally, the phenotypic

differences observed in our patients are biased by age at clinical assessment. The oldest pair of brothers (IV:1 and IV:2) are mature adults, while their cousins (V:2 and V:3) are young adults. The younger patients may develop additional or progressive phenotypes as observed in the older pair of brothers. Age of OHS onset ranges from early childhood to early adulthood, and phenotypes are usually progressive (Tümer and Møller, 2010). Discussion of phenotypic variability between these individuals would be better informed by tracking phenotype progression over time.

Finally, it is important to consider that there are factors aside from splicing which may affect phenotype severity. Previously, analysis of an *ATP7A* missense variant identified in brothers with variable phenotypes indicated that 17 % of wild type activity was maintained (Donsante et al., 2007). Variability in phenotype between the brothers was attributed to greater upregulation of *ATP7A* transcription in fibroblasts from the more mildly affected brother with the OHS phenotype, compared to his more severely affected brother with a Menkes syndrome phenotype. The authors suggested that the increased *ATP7A* expression compensated for lost ATP7A channel function, and that differences in overexpression underlay the phenotypic variability (Donsante et al., 2007). Factors influencing *ATP7A* expression may also facilitate the variation in phenotypes of the family reported here.

The identification of this family with a novel deep intronic *ATP7A* variant contributes to the understanding of OHS as a variable disease and illustrates the value in conducting WGS to identify deep intronic variants. Future work should establish how leaky splicing variants can lead to life-long variability in disease severity in different tissues to better understand disease progression and improve clinical management.

## 5.5 Methods

All participants provided informed written consent in accord with a protocol approved by the South Manchester Ethics committee (11/H1003/3, IRAS 64321). Genomics England has approval from the HRA Committee East of England – Cambridge South (REC Ref 14/EE/1112).

### 5.5.1 SNP Array Genotyping

SNP array analysis was performed with Genome-Wide Human SNP Array v.6.0 (Affymetrix) according to the manufacturer's protocol. Genotypes and copy number data were generated within the Affymetrix Genotyping Console (v.4.1.3.840) via the Birdseed V2 algorithm and SNP 6.0 CN/LOH algorithm, respectively.

### 5.5.2 Exome Sequencing

Exome sequencing was carried out for one affected individual (IV:2) using the SureSelect Human All Exon Kit v4 (Agilent Technologies, Edinburgh UK) for the Illumina HiSeq 2500 system (Illumina, Cambridge, UK). Sequence data were mapped to the hg19 reference human genome using the Burrows–Wheeler aligner software (version 0.6.2; <http://bio-bwa.sourceforge.net>). Genome Analysis Tool Kit software (version 2.4.7; <https://www.broadinstitute.org/gatk>) was used for recalibration of base quality score and for indel realignment before using the unified genotyper (<https://www.broadinstitute.org/gatk>) for variant calling.

### 5.5.3 Whole Genome Sequencing

Whole genome sequencing (WGS) was conducted as part of the 100,000 Genomes Project (Genomics England). Blood samples, consent, clinical indication and HPO terms were collected for patients at their local hospital. Blood samples were sent to the Manchester regional genetic laboratory and DNA was extracted and sent to Illumina for WGS. Library preparation was performed using TruSeq DNA PCR-Free Library Prep, with 250 bp paired-end reads generated from the HiSeqX sequencer. Data were passed through Genomic England's bioinformatics pipeline. All variants within *ATP7A* locus (GRCh38:X:77910690-

78050395) were filtered on the basis of quality control, allele frequency, and hypothesised allele segregation.

#### 5.5.4 Sanger Sequencing

Sanger sequencing was completed using Big Dye Terminator c.3.1 Cycle Sequencing Kit following manufacturer's instructions, and products were analysed with 3730 Genetic Analyser instrument (Thermo Fisher Scientific). Primers for amplifying *ATP7A* intron 4 from genomic DNA were CTGTCACACATGGTGCATTG (forward) and GAGGGCAGATCGCTTGAGTC (reverse).

#### 5.5.5 Minigene Assay

A minigene vector was assembled to assess the effect of the intronic variant on splicing on upstream and downstream *ATP7A* exons. An SK3 vector was used, which is a derivative of the pSpliceExpress vector used previously (Thomas et al., 2020). A fragment of proband genomic *ATP7A* DNA, including the region from exon 4 to exon 5, plus 100 bp flanking sequence, was cloned into the SK3 minigene vector. Additionally, a DNA fragment was amplified from genomic DNA of the mother M1 was used to construct a WT *ATP7A* minigene vector for use as a control. The SK3 minigene backbone was isolated by digestion with restriction enzymes Nhe1 and BamH1. The genomic *ATP7A* DNA fragment was amplified by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) using primers (forward gtacgggtgaccacgcgtccatgggTGGATTCTTTGCTACAATTATGATTTTCATAAGTGCAT, reverse CCGGATCgagctgcatgtgtcagagGAATTTATTCAATTACATGCTACTAGATAGGAATCA) designed to produce fragments which overlap with the digested SK3 minigene backbone. The *ATP7A* fragment and the SK3 minigene backbone were assembled via the Gibson method [Gibson 2011], and transformed into ultra-competent bacteria. Assembled *ATP7A* minigene vectors were isolated from bacterial colonies, and correct vector assembly was validated by Sanger sequencing performed by EuroFins Genomics.

HEK293 cells were cultured to 50 % confluency in 2 ml of Dulbecco's modified Eagle's high-glucose medium (DMEM Sigma-Aldrich), supplemented with 10 % foetal bovine serum

(Sigma-Aldrich) in tissue-culture treated six-well plates at 37°C with 5 % CO<sub>2</sub>. Cells were transfected with 1 µg of proband or maternal WT *ATP7A* minigene vector using Lipofectamine LTX reagent (Thermo Fisher Scientific) following manufacturer's instructions. Following 24 hours incubation (37°C, 5 % CO<sub>2</sub>), RNA was extracted from cells using TRI Reagent (Sigma-Aldrich) according to manufacturer's instructions. Extracted RNA was purified using RNeasy column clean-up kit (Qiagen), including the optional DNase digestion step. 5 µg of RNA was reverse transcribed to cDNA using Superscript Reverse Transcriptase (Thermo Fisher Scientific). cDNA was amplified by PCR using minigene specific primers (forward GCACCTTTGTGGTTCTCACT, reverse GGGCCTAGTTGCAGTAGTTCT), and products were run on 1% agarose gels supplemented with SafeView nucleic acid stain (NGS Biologicals). Gels were visualized using a blue-light transilluminator, and bands of interest were excised. DNA from excised fragments was extracted and purified using QIAquick gel extraction kit (Qiagen). Splicing products were confirmed by Sanger sequencing performed by EuroFins Genomics.

#### 5.5.6 RNA extraction from patient cells

Patient derived lymphocyte cell line cultures, and fibroblast cultures were established by Specialised Cell Culture Services at St Mary's Hospital, Manchester following local procedures. Cells were harvested and centrifuged at 200 xg for 5 mins, and pellets were stored at -80 °C until RNA extraction. To extract RNA, cell pellets were thawed on ice prior to extraction using the Rneasy RNA extraction kit (Qiagen), following manufacturer's instructions, including the optional Dnase digestion step. 1 µg RNA was reverse transcribed to cDNA using High-Capacity RNA-to-cDNA kit (Applied Biosystems), following manufacturer's instructions. Patient cDNA was stored at -20 °C until further use.

#### 5.5.7 Quantitative RT-PCR

Quantitative real time PCR was used to quantify exon junctions in the *ATP7A* transcript. Predesigned *ATP7A* probes Hs00921963\_m1 (exon 1-2 boundary), Hs00921969\_m1 (exon 4-5 boundary), and Hs00921970\_m1 (exon 5-6 boundary) were used. A predesigned probe for *GAPDH* (Hs02786624\_g1) was used as an endogenous control. All probes were 6-carboxy-fluorescein labelled and were purchased from ThermoFisher Scientific. 20 µl reactions

containing cDNA, 20x FAM-labelled probe, and 2x TaqMan™ Gene Expression Master Mix (Applied Biosystems). PCR reactions were quantified using StepOnePlus™ Fast Real-Time PCR machine (Applied Biosystems). Standard curves of  $C_T$  values compared with log cDNA concentration were constructed by measuring tenfold serial dilutions of control cDNA from 100 to 0.01 ng/μl with each probe.

## **5.6 Acknowledgements**

We would like to thank members of the family for their participation. The research was funded through support from the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007) to WGN and the Wellcome Trust (PhD studentship to JRH).

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

## 5.7 References

- Anna, A., Monika, G., 2018. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genetics* 59, 253–268. <https://doi.org/10.1007/s13353-018-0444-7>
- Chin, H.-L., Lin, S., Dalmann, J., Modi, B., Alderman, E., Salman, A., Del Bel, K.L., Lehman, A., Turvey, S.E., Boerkoel, C.F., 2022. Can leaky splicing and evasion of premature termination codon surveillance contribute to the phenotypic variability in Alkuraya-Kucinskas syndrome? *Eur J Med Genet* 65, 104427. <https://doi.org/10.1016/j.ejmg.2022.104427>
- Christodoulou, J., Danks, D.M., Sarkar, B., Baerlocher, K.E., Casey, R., Horn, N., Tümer, Z., Clarke, J.T.R., 1998. Early treatment of Menkes disease with parenteral Cooper-Histidine: Long-term follow-up of four treated patients. *American Journal of Medical Genetics* 76, 154–164. [https://doi.org/10.1002/\(SICI\)1096-8628\(19980305\)76:2<154::AID-AJMG9>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1096-8628(19980305)76:2<154::AID-AJMG9>3.0.CO;2-T)
- Dagenais, S.L., Adam, A.N., Innis, J.W., Glover, T.W., 2001. A novel frameshift mutation in exon 23 of ATP7A (MNK) results in occipital horn syndrome and not in Menkes disease. *Am J Hum Genet* 69, 420–427. <https://doi.org/10.1086/321290>
- Donsante, A., Tang, J., Godwin, S.C., Holmes, C.S., Goldstein, D.S., Bassuk, A., Kaler, S.G., 2007. Differences in ATP7A gene expression underlie intrafamilial variability in Menkes disease/occipital horn syndrome. *J Med Genet* 44, 492–497. <https://doi.org/10.1136/jmg.2007.050013>
- Fradin, M., Lavillaureix, A., Jaillard, S., Quelin, C., Sauleau, P., Minot, M.-C., Menard, D., Edan, G., Ceballos, I., Treguier, C., Proisy, M., Magdelaine, C., Lia, A.-S., Odent, S., Pasquier, L., 2020. ATP7A mutation with occipital horns and distal motor neuropathy: A continuum. *Eur J Med Genet* 63, 104087. <https://doi.org/10.1016/j.ejmg.2020.104087>
- Gérard-Blanluet, M., Birk-Møller, L., Caubel, I., Gélot, A., Billette de Villemeur, T., Horn, N., 2004. Early development of occipital horns in a classical Menkes patient. *Am J Med Genet A* 130A, 211–213. <https://doi.org/10.1002/ajmg.a.30155>
- Gualandi, F., Sette, E., Fortunato, F., Bigoni, S., De Grandis, D., Scotton, C., Selvatici, R., Neri, M., Incensi, A., Liguori, R., Storbeck, M., Karakaya, M., Simioni, V., Squarzoni, S., Timmerman, V., Wirth, B., Donadio, V., Tugnoli, V., Ferlini, A., 2019. Report of a novel ATP7A mutation causing distal motor neuropathy. *Neuromuscular Disorders* 29, 776–785. <https://doi.org/10.1016/j.nmd.2019.08.008>
- Horn, N., Wittung-Stafshede, P., 2021. ATP7A-Regulated Enzyme Metalation and Trafficking in the Menkes Disease Puzzle. *Biomedicines* 9, 391. <https://doi.org/10.3390/biomedicines9040391>
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Batzoglu, S., Sanders, S.J., Farh, K.K.-H., 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- Kaler, S.G., DiStasio, A.T., 1993. ATP7A-Related Copper Transport Disorders, in: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Mirzaa, G.M., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).
- Kennerson, M.L., Nicholson, G.A., Kaler, S.G., Kowalski, B., Mercer, J.F.B., Tang, J., Llanos, R.M., Chu, S., Takata, R.I., Speck-Martins, C.E., Baets, J., Almeida-Souza, L., Fischer, D., Timmerman, V., Taylor, P.E., Scherer, S.S., Ferguson, T.A., Bird, T.D., De Jonghe, P., Feely, S.M.E., Shy, M.E., Garbern, J.Y., 2010. Missense mutations in the copper transporter gene ATP7A cause X-linked distal hereditary motor neuropathy. *Am J Hum Genet* 86, 343–352. <https://doi.org/10.1016/j.ajhg.2010.01.027>
- Lutsenko, S., Barnes, N.L., Bartee, M.Y., Dmitriev, O.Y., 2007. Function and regulation of human copper-transporting ATPases. *Physiol Rev* 87, 1011–1046. <https://doi.org/10.1152/physrev.00004.2006>
- Maung, M.T., Carlson, A., Olea-Flores, M., Elkhadragy, L., Schachtschneider, K.M., Navarro-Tito, N., Padilla-Benavides, T., 2021. The molecular and cellular basis of copper dysregulation and its



- relationship with human pathologies. *FASEB J* 35, e21810.  
<https://doi.org/10.1096/fj.202100273RR>
- Mhaske, A., Dileep, K.V., Kumar, M., Poojary, M., Pandhare, K., Zhang, K.Y.J., Scaria, V., Binukumar, B.K., 2020. ATP7A Clinical Genetics Resource - A comprehensive clinically annotated database and resource for genetic variants in ATP7A gene. *Comput Struct Biotechnol J* 18, 2347–2356. <https://doi.org/10.1016/j.csbj.2020.08.021>
- Møller, L.B., 2015. Small amounts of functional ATP7A protein permit mild phenotype. *J Trace Elem Med Biol* 31, 173–177. <https://doi.org/10.1016/j.jtemb.2014.07.022>
- Møller, L.B., Mogensen, M., Weaver, D.D., Pedersen, P.A., 2021. Occipital Horn Syndrome as a Result of Splice Site Mutations in ATP7A. No Activity of ATP7A Splice Variants Missing Exon 10 or Exon 15. *Frontiers in Molecular Neuroscience* 14.
- Nadal, D., Baerlocher, K., 1988. Menkes' disease: long-term treatment with copper and D-penicillamine. *Eur J Pediatr* 147, 621–625. <https://doi.org/10.1007/BF00442477>
- Rangarh, P., Kohli, N., 2018. Neuroimaging findings in Menkes disease: a rare neurodegenerative disorder. *BMJ Case Rep* 2018, bcr-2017-223858. <https://doi.org/10.1136/bcr-2017-223858>
- Shibuya, M., Yaoita, H., Kodama, K., Okubo, Y., Endo, W., Inui, T., Togashi, N., Takayama, J., Tamiya, G., Kikuchi, A., Kure, S., Haginoya, K., 2022. A patient with early-onset SMAX3 and a novel variant of ATP7A. *Brain Dev* 44, 63–67. <https://doi.org/10.1016/j.braindev.2021.08.004>
- Skjørringe, T., Amstrup Pedersen, P., Salling Thorborg, S., Nissen, P., Gourdon, P., Birk Møller, L., 2017. Characterization of ATP7A missense mutants suggests a correlation between intracellular trafficking and severity of Menkes disease. *Sci Rep* 7, 757.  
<https://doi.org/10.1038/s41598-017-00618-6>
- Sobczyńska-Tomaszewska, A., Ołtarzewski, M., Czerska, K., Wertheim-Tysarowska, K., Sands, D., Walkowiak, J., Bal, J., Mazurczak, T., NBS CF working group, 2013. Newborn screening for cystic fibrosis: Polish 4 years' experience with CFTR sequencing strategy. *Eur J Hum Genet* 21, 391–396. <https://doi.org/10.1038/ejhg.2012.180>
- Takata, R.I., Speck Martins, C.E., Passosbueno, M.R., Abe, K.T., Nishimura, A.L., Da Silva, M.D., Monteiro, A., Lima, M.I., Kok, F., Zatz, M., 2004. A new locus for recessive distal spinal muscular atrophy at Xq13.1-q21. *J Med Genet* 41, 224–229.  
<https://doi.org/10.1136/jmg.2003.013201>
- Thomas, H.B., Wood, K.A., Buczek, W.A., Gordon, C.T., Pingault, V., Attié-Bitach, T., Hentges, K.E., Varghese, V.C., Amiel, J., Newman, W.G., O'Keefe, R.T., 2020. EFTUD2 missense variants disrupt protein function and splicing in mandibulofacial dysostosis Guion-Almeida type. *Human Mutation* 41, 1372–1382. <https://doi.org/10.1002/humu.24027>
- Tønnesen, T., Kleijer, W.J., Horn, N., 1991. Incidence of Menkes disease. *Hum Genet* 86, 408–410.  
<https://doi.org/10.1007/BF00201846>
- Trackman, P.C., 2018. Functional importance of lysyl oxidase family propeptide regions. *J Cell Commun Signal* 12, 45–53. <https://doi.org/10.1007/s12079-017-0424-4>
- Tümer, Z., Møller, L.B., 2010. Menkes disease. *Eur J Hum Genet* 18, 511–518.  
<https://doi.org/10.1038/ejhg.2009.187>
- Tümer, Z., Petris, M., Zhu, S., Mercer, J., Bukrinski, J., Bilz, S., Baerlocher, K., Horn, N., Møller, L.B., 2017. A 37-year-old Menkes disease patient-Residual ATP7A activity and early copper administration as key factors in beneficial treatment. *Clin Genet* 92, 548–553.  
<https://doi.org/10.1111/cge.13083>
- Vairo, F.P. e, Chwal, B.C., Perini, S., Ferreira, M.A.P., de Freitas Lopes, A.C., Saute, J.A.M., 2019. A systematic review and evidence-based guideline for diagnosis and treatment of Menkes disease. *Molecular Genetics and Metabolism* 126, 6–13.  
<https://doi.org/10.1016/j.ymgme.2018.12.005>
- Yasmeen, S., Lund, K., De Paepe, A., De Bie, S., Heiberg, A., Silva, J., Martins, M., Skjørringe, T., Møller, L.B., 2014. Occipital horn syndrome and classical Menkes Syndrome caused by deep

intronic mutations, leading to the activation of ATP7A pseudo-exon. *Eur J Hum Genet* 22, 517–521. <https://doi.org/10.1038/ejhg.2013.191>

Zlatic, S., Comstra, H.S., Gokhale, A., Petris, M.J., Faundez, V., 2015. Molecular basis of neurodegeneration and neurodevelopmental defects in Menkes disease. *Neurobiol Dis* 81, 154–161. <https://doi.org/10.1016/j.nbd.2014.12.024>

## Chapter 6: Biallelic Variants in *RCC1* result in fever-associated axonal neuropathy with encephalopathy

J. Robert Harkness<sup>1,2</sup>, John McDermott<sup>1,2</sup>, Kay Metcalfe<sup>1,2</sup>, William L. Macken<sup>3,4</sup>, Ataf Sabir, Saikat Santra<sup>6</sup>, Jill E. Urquhart<sup>1,2</sup>, Leigh Demain<sup>1,2</sup>, Christopher J. Record<sup>3</sup>, Wyatt W. Yue<sup>7</sup>, Helen Byers<sup>1,2</sup>, Robert D.S. Pitceathly<sup>3,4</sup>, Robert W. Taylor<sup>8,9</sup>, Simon Hubbard<sup>2</sup>, Siddharth Banka<sup>1,2</sup>, Genomics England Research Consortium, Mary M. Reilly<sup>3</sup>, Raymond T. O'Keefe<sup>2</sup>, William G. Newman<sup>1,2</sup>

<sup>1</sup> Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK

<sup>2</sup> Division of Evolution, Infection and Genomics, Faculty of Biology, Medicine and Health Sciences, University of Manchester, Manchester, UK.

<sup>3</sup> Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK.

<sup>4</sup> NHS Highly Specialised Service for Rare Mitochondrial Disorders, Queen Square Centre for Neuromuscular Diseases, The National Hospital for Neurology and Neurosurgery, London, UK.

<sup>5</sup> West Midlands Clinical Genetics Unit, Birmingham Women's and Children's NHS FT and Birmingham Health Partners, Birmingham, United Kingdom.

<sup>6</sup> Department of Clinical Inherited Metabolic Disorders, Birmingham Women's and Children's NHS Foundation Trust, Steelhouse Lane, Birmingham B4 6NH, UK

<sup>7</sup> Newcastle MX Structural Biology Laboratory, Newcastle University, Medical School, NUBI, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK

<sup>8</sup> Wellcome Centre for Mitochondrial Research, Translational and Clinical Research Institute, Faculty of Medical Sciences Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

<sup>9</sup> NHS Highly Specialised Service for Rare Mitochondrial Disorders, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, NE1 4LP, UK

Correspondence to:

Professor William G. Newman

Manchester Centre for Genomic Medicine,  
Manchester University NHS Foundation Trust,  
Manchester, M13 9WL  
UK.  
Telephone 00 44 161 276 4150  
Email [william.newman@manchester.ac.uk](mailto:william.newman@manchester.ac.uk)

Key words. Acute necrotising encephalopathy, axonal neuropathy, *RCC1*, missense, homozygous, recessive, exome, genetics

## 6.1 Abstract

Acute necrotising encephalopathy (ANE) is a rare disease characterised by rapid onset of neurological features, including seizures, cognitive impairment, and in some cases death, following febrile illness. Previously genetic variants in *RANBP2* and *NUP214* have been identified in inherited forms of ANE. We identified a large consanguineous British Pakistani family with four children affected by rapid onset axonal neuropathy following infection with features overlapping ANE. A novel c.127G>A:p.(Gly43Ser) homozygous missense variant in *RCC1* which segregated with the phenotype was identified by autozygosity mapping and exome sequencing. Two further unrelated individuals biallelic for *RCC1* missense variants were identified through reverse phenotyping. Secondary to infection, one individual was affected by encephalopathy and the other had axonal neuropathy similar in presentation to Guillain-Barré syndrome.

Rcc1 protein binds Ran, a small nuclear GTPase which controls nucleocytoplasmic transport and binds to *RANBP2*. Rcc1<sup>G43S</sup> had mildly reduced thermostability compared to Rcc1<sup>WT</sup>, with no effect on GTPase activity when assessed over a physiological temperature gradient. Immunofluorescence staining of patient fibroblasts cultured at 42 °C for 4 hours indicated delocalisation of Rcc1 from chromatin. Further research will explore the effects of biallelic missense variants on nucleocytoplasmic transport. We describe a novel form of autosomal recessive ANE due to biallelic *RCC1* variants, potentially providing an insight into the aetiology of other acute neuropathies.

## 6.2 Introduction

Acute necrotising encephalopathy (ANE) is a rare degenerative brain disorder characterised by the rapid onset of severe neurological symptoms following a viral infection (Levine et al., 2020). After 1-3 days of febrile infection a variable set of neurological symptoms begin to present including: seizures, hallucinations, ataxia, and loss of consciousness (Levine et al., 2020). Radiographically, these symptoms are typically associated with symmetrical multifocal brain lesions affecting the thalami, and may affect the periventricular white matter, brainstem, and cerebellum (Singh et al., 2015). ANE shares significant overlap with similar infection-triggered encephalopathy-related disorders such as acute disseminated encephalomyelitis (ADEM). Both ANE and ADEM are initiated following viral infection, and both result in severe neurological deficiencies (Singh et al., 2015). However, in ADEM multifocal lesions are spread across the central nervous system (CNS) including the brain and spinal cord, and are characterised by demyelination, leading to loss of white matter (Tenenbaum et al., 2007).

Although the aetiology of ANE is unknown in many cases, ANE with a hereditary or genetic basis has been reported (Singh et al., 2015). Heterozygous missense variants in Ran-binding protein 2 (*RANBP2*; MIM: 601181) have been identified in several affected families (Chew and Ngu, 2020; Neilson et al., 2009; Ohashi et al., 2021; Shibata et al., 2021) with autosomal dominant ANE designated ANE1 (MIM: 608033) (Levine et al., 2020; Neilson et al., 2009; Singh et al., 2015). A recent review of 96 ANE1 cases revealed variability in age of onset, with 63.7 % diagnosed below the age of four years (Jiang et al., 2022). The mortality rate among ANE1 cases was 25.4 %, with 52.1 % experiencing long-term neurological and cognitive disability (Jiang et al., 2022). Biallelic variants in *NUP214* (MIM: 114350) have been identified in patients with acute febrile encephalopathy (Fichtman et al., 2019; Shamseldin et al., 2019). Patients with *NUP214* variants presented with mild developmental delay, microcephaly, and transient episodes of febrile illness-related regression featuring encephalopathy and ataxia (Fichtman et al., 2019).

In ANE1, onset of disease is not associated with infection by specific pathogen. Commonly ANE1 is caused by influenza (Mariotti et al., 2010; Martin and Reade, 2010), but other viral pathogens including Human Herpes Virus 6 (HHV6) (Huang et al., 2020; Skelton et al., 2008), varicella (Tran et al., 2001), and coronavirus (Pongpitakmetha et al., 2022; Poyiadji et al., 2020) have been associated with ANE1 symptom onset (Jiang et al., 2022). Additionally, one case was associated with infection by the bacterium *Mycoplasma pneumoniae* (Lee et al., 2017). This implies that host factors are the main driver of the symptom onset, precipitated during the innate response to infection. ANE1 is therefore distinct from other encephalopathic diseases such as herpes-specific acute encephalitis, which is characterised by neuron-specific reaction to herpes virions (Lafaille et al., 2019) and associated with variants in toll-like receptor (TLR) 3 pathway genes (Zhang, 2020) such as *UNC93B1* (Casrouge et al., 2006), *IRF3* (Andersen et al., 2015), and *SNORA31* (Lafaille et al., 2019).

RanBP2 (also called Nup358) and Nup214 both act as nucleoporins in the export of cargoes from the nucleus to the cytoplasm via the Nuclear Pore Complex (NPC) (Jiang et al., 2022; Yokoyama et al., 1995). Nucleocytoplasmic transport (Figure 1) relies on the action of Ran, a small nuclear GTPase. In the nucleus, GDP-bound to Ran is exchanged for GTP by the Ran-specific guanine exchange factor (GEF) Rcc1 (Regulator of chromatin condensation 1) (Chatterjee and Paschal, 2015; Seki et al., 1996). Rcc1 binds to chromatin, effectively anchoring Ran-GTP production in the nucleus (Chen et al., 2007; Clarke and Zhang, 2008; Ohtsubo et al., 1987; Ren et al., 1993). This forms a Ran-GTP gradient which facilitates Ran-GTP-dependent export of nuclear cargoes that contain nuclear export signals to the cytoplasm (Hamed et al., 2021; Hutten and Kehlenbach, 2006).

Here we describe a large consanguineous family with multiple individuals severely affected with ANE. We identify a novel homozygous variant in the *RCC1* gene and report a further unrelated individual with compound heterozygous *RCC1* variants associated with encephalopathy and axonal neuropathy secondary to recurrent viral infection. The aim of this study is to characterise the effect of the homozygous *RCC1* variant on ANE-related cellular processes to establish the mechanism of disease.

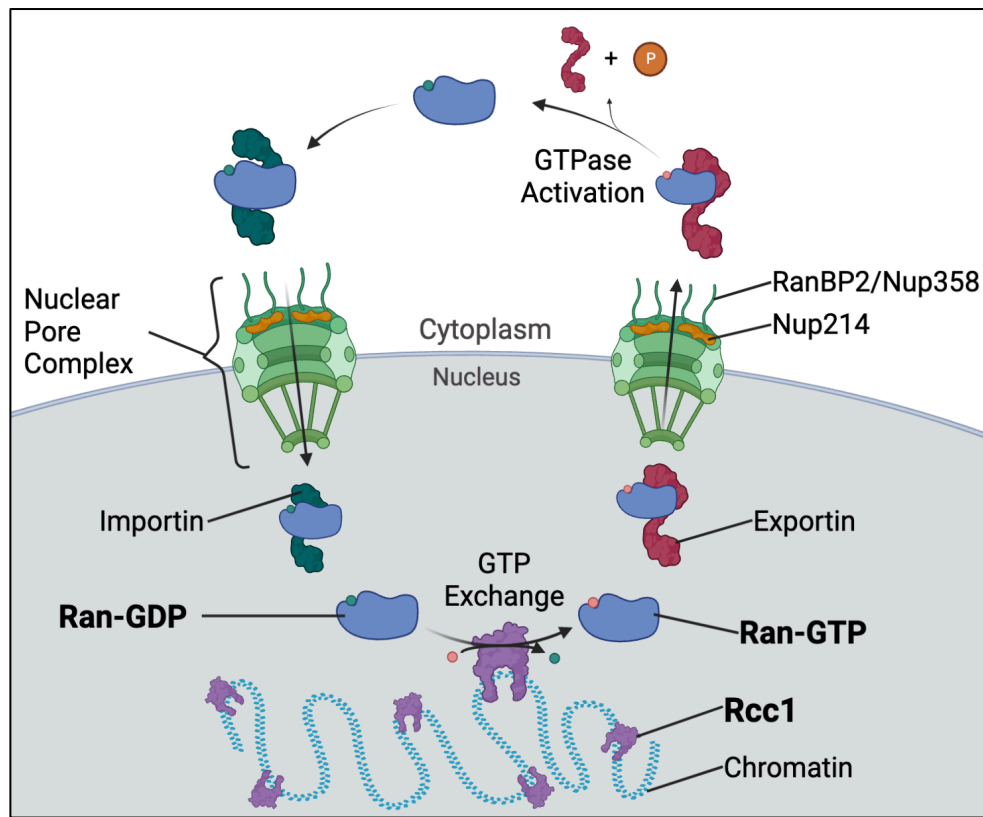


Figure 1. Schematic of Ran GTPase action in nucleocytoplasmic transport. In the nucleus, Rcc1, bound to chromatin acts as a guanine exchange factor (GEF) for Ran, catalysing GDP to GTP exchange. Ran-GTP binds exportins and cargo with nuclear export signals for transport through the nuclear pore complex (NPC), which contains RanBP2/Nup358 and Nup214. In the cytoplasm, RanBP2 binding permits Ran GTPase activation, facilitating exportin complex disassembly. During the GTPase reaction, GTP is converted to GDP and Phosphate. Ran-GDP is reimported to the nucleus by formation of importin complex. Upon re-entry, Ran-GDP undergoes GTP exchange from chromatin-bound Rcc1. Schematic produced using BioRender.com based on schematic by Holzer and Antonin (2022).

## 6.3 Results:

### 6.3.1 Clinical findings

#### Family A

The proband (Family A IV:2, Figure 2A) is a 2-year-old female from a large consanguineous British Pakistani family. She was assessed as having mild motor delay at age 14 months, having sat at one year and crawled at 13 months. At 17 months of age she started to cruise with the aid of furniture but not walking independently. Aged 18 months she developed a fever associated with an upper respiratory tract infection. There was no improvement after four days of treatment with amoxicillin, and she began to develop weakness resulting in attendance at the Emergency Department before admission to the paediatric intensive care unit. Following admission, she developed stridor, had a Glasgow coma scale score of 13/15 (E4,M5,V4), and was assessed as having progressive hypotonia characterised by motor weakness and poor swallow. Four days after admission, she developed respiratory arrest, possibly with loss of cardiac output, and was intubated and ventilated.

Electroencephalogram indicated abnormal background waves suggesting encephalopathy. An MRI soon after admission showed subtle restricted diffusion involving the lateral aspect of the left cerebellum.

In the days following admission, the proband developed profound weakness and hypotonia affecting four limbs, with loss of tendon reflexes, bulbar paralysis, and autonomic dysfunction. She was treated with intravenous immunoglobulin and plasma exchange. She remained on the paediatric intensive care unit for 10 months with little progress during which time brain MRI indicated progressive cerebral and cerebellar atrophy. Muscle biopsy indicated neurogenic atrophy with a group atrophy pattern suggesting anterior horn cell involvement. There was no evidence of respiratory chain defect in patient dermal fibroblasts. Care was stepped down to paediatric high dependency unit, and she was subsequently discharged with home ventilation.

Family history indicated presence of infection-related encephalopathy among three second cousins (IV:3, IV:4 and IV:6). These were all affected by rapid onset of motor weakness



within the first year of life, resulting in death between four (IV:4 and IV:6) and ten months (IV:3) after symptom onset (Table 1).

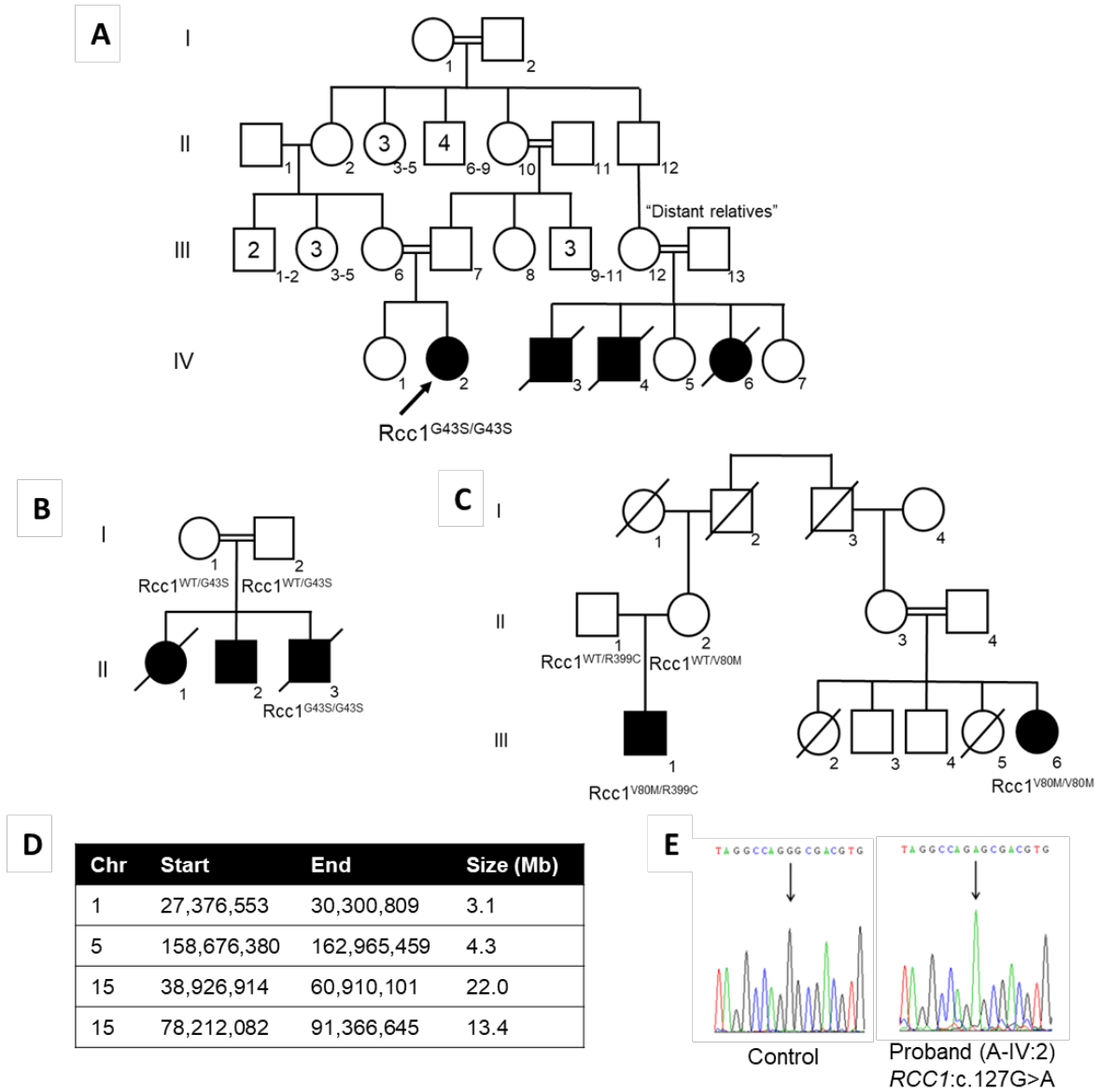


Figure 2: Pedigrees and genetic investigations of individuals with ANE. (A) Pedigree of Family A, a large consanguineous family with 4 individuals affected by ANE (filled shapes). Parents III:12 and III:13 of affected individuals IV:3, IV:4, IV:6 are described by the family as “distant relatives”, but the relationship is unknown. (B-C) Pedigrees of families B and C ascertained following identification of (B) II:3 and (C) III:1 with biallelic *RCC1* variants through the 100,000 genomes project. (D) Regions of autozygosity shared by IV:4 and IV:6 from Family A (coordinates in GRCh38). (E) Sanger sequencing of proband IV:2 from family A with homozygous *RCC1*:c.127G>A variant. ANE, acute necrotising encephalopathy.

Table 1. Summary of clinical findings in family A. CFS, chronic fatigue syndrome. EBV, Epstein-Barr Virus.

Pedigree ID	IV:2	IV:3	IV:4	IV:6
Sex	Female	Male	Male	Female
Initial development	Mild motor delay	Normal	Normal	Normal
Age at onset	18 months	8 months	7 months	7 months
Preceding illness	Upper respiratory tract infection, no improvement with amoxicillin after four days	Diarrhoea for six weeks, febrile illness for three weeks	5 day generally unwell and reduced feeding	Fever and rash (possibly roseola infantum)
Course	Rapid onset weakness and hypotonia. Respiratory arrest with possible loss of cardiac output. Progressive weakness of four limbs with hypotonia and loss of reflexes, bulbar paralysis, and autonomic dysfunction.	Rapid onset ascending flaccid paralysis. Flickering of toes and eyes only. Long term ventilation. Areflexic.	Rapid onset ascending flaccid paralysis. Flickering of toes, fixing, and following. Regained smiling. Areflexic	Rapid onset ascending flaccid paralysis. Some eye opening only. Ventilated. Areflexic
Recurrent episodes.	None	Second febrile illness at 17 months associated with loss of eye movement and respiratory disease.	None	None
EEG finding	Abnormal background waves, encephalopathy			Initially normal, repeat encephalopathic

MRI finding	Subtle restricted diffusion involving lateral aspect of left cerebellum at admission, subsequent progressive cerebral and cerebellar atrophy.	Progressive cerebral and cerebellar atrophy.	Generalised cerebral atrophy.	Not done
Virology	Not done	CSF virology negative. EBV serology consistent with infection more than 8 weeks previously.	Not done	CFS virology negative.
Additional findings	Muscle biopsy: neurogenic atrophy with group atrophy pattern.	Nerve conduction: no response. Brain histology: widespread microvacuolation of neutrophil and neuronal depletion, gliosis, axonal swelling (thalamus, pallidum, nigra, brainstem).	None	None
Current status	Alive, ventilated.	Death at 18 months	Death at 11 months	Death at 11 months

### 6.3.2 Genetic analysis of Family A

The identification multiple individuals and consanguinity in Family A suggested a monogenic basis of disease. Clinical whole exome sequencing (WES) of IV:4 revealed no variants in known encephalopathy-related genes. Autozygosity mapping analysis of affected individuals IV:4 and IV:6 and their unaffected sibling IV:5 identified regions of autozygosity shared by affected individuals on chromosome 1, 5, and two regions on chromosome 15 (Figure 2D). Targeted reanalysis of WES data from IV:4 identified a homozygous *RCC1* c.127G>A:p.(Gly43Ser) missense variant (NM\_001381865.2, GRCh38 1:28531856G>A). Sanger sequencing (Figure 2E) confirmed homozygosity of IV:2 and IV:4 (no sample available for IV:3). Parents III:12 and III:13, and unaffected sibling IV:5 were heterozygous for the *RCC1* c.127G>A variant (unshown). Testing has not been undertaken in either IV:1 or IV:7 at the request of the parents. The *RCC1* variant was not previously identified in gnomAD and had a CADD score of 29.4. This variant is classified as a variant of unknown significance (VUS) according to classification guidelines from the American College of Medical Genetics and Genomics (Richards et al., 2015).

### 6.3.3 Identification of further affected individuals in the 100,000 genomes project

InTo support description of a novel *RCC1*-related pathology, further individuals were sought from data available in the 100,000 genomes project. Analysis of patients with biallelic *RCC1* variants identified two individuals from other centres in the UK, with clinical features overlapping that of the proband (Table 2). Both these individuals were from consanguineous families (Family B and Family C) where there was a history of individuals affected by acute-onset neurological phenotypes.

Individual II:3 in Family B (Figure 2B) was identified with the same novel homozygous *RCC1* c.127G>A:p.(Gly43Ser) variant as the proband (IV:2 from Family A). He died aged 12 months after onset of demyelinating polyneuropathy and severe encephalopathy featuring hypotonia and myopathy following infection with lobular pneumonia. II:3 is the youngest of three siblings from a British Pakistani family. Two older siblings II:1 and II:2 were also affected by rapid onset severe motor neuropathy. The sister II:1 died aged 4 following mild infection and neurological decline, while the brother II:2 acquired four-limb motor

neuropathy following mild infection and is on home ventilation. Given the large consanguineous nature of this family and the identification of the same novel homozygous *RCC1* variant, we suspect that Family B is highly likely to be related to Family A, though it has not yet been possible to establish a definitive connection through consultation with the families.

A further unrelated individual, III:1 in Family C (Figure 2C), was identified with compound heterozygous *RCC1* variants. One heterozygous c.238G>A:p.(Val80Met) variant was inherited from Cypriot mother, while the other heterozygous c.1195C>T:p.(Arg399Cys) variant inherited from a British father. III:1 is a 26-year-old British Cypriot male and has been affected by multiple episodes of recurrent severe axonal polyneuropathy since the age of 18 months. Transient polyneuropathic episodes which coincide with periods of febrile illness in this individual were characterised initially by severe lower limb motor weakness. In subsequent episodes upper limb weakness is also described. Respiratory chain analysis was normal. There is consanguinity among the maternal Cypriot family, and a positive history of severe neurological phenotype onset in the context of infection among maternal second cousins. Sequencing of affected cousin (III:6) identified homozygosity for the c.238G>A:p.(Val80Met) variant.

Table 2: Identification of unrelated probands with biallelic *RCC1* variants from 100,000 Genomes Project Exomiser data. AF allele frequency, HPO human phenotype ontology.

Identifier	Family B II:3	Family C III:1	
Zygosity	Homozygous	Heterozygous	Heterozygous
Variant	c.127G>A:p.(Gly43Ser)	c.238G>A:p.(Val80Met)	c.1195C>T:p.(Arg399Cys)
Mother	Heterozygous	Heterozygous	Reference
Father	Heterozygous	Reference	Heterozygous
gnomAD	Unseen	2 het, AF 0.00092682 %	Unseen
CADD	29.4	28.0	32.0
Conservation	<i>C. elegans</i>	Zebrafish	<i>Drosophila</i>

Development	Jaundice at birth, mild developmental hypotonia/myopathy Congenital laryngomalacia, congenital deformity of feet.	Normal, walking at 9 months
Age at onset	11 months	18 months
Infection	Lobular pneumonia	Episode onset coincides with illness
Disease characteristics	Hypotonia, abnormal involuntary movements, respiratory failure, tachycardia, demyelinating polyneuropathy, encephalopathy.	Recurrent Guillain-Barré-like severe demyelinating axonal polyneuropathy, lower and upper limb weakness, neuromuscular scoliosis, hydronephrosis relating to ureteral obstruction, fatigable weakness of respiratory muscles, sleep apnoea.
Mitochondrial respiratory chain analysis	Mitochondrial complex 1 deficiency. No variants identified from sequencing of mitochondrial genes.	Negative for respiratory chain defects.
Outcome	Death aged 12 months	Alive aged 26.
Family history	Consanguinity	Maternal side consanguinity and history of infection-related polyneuropathy.

#### 6.3.4 Rcc1<sup>G43S</sup> retains Ran GEF activity

As *RCC1* variants have not previously been observed in monogenic disease, we sought to undertake further functional characterisation of Rcc1 protein activity to establish the molecular basis for pathogenicity in this case of acute encephalopathy. Rcc1 is a chromatin-binding protein and acts as a guanine exchange factor (GEF) for Ran GTPase. Assessment of Rcc1 protein structure indicated that the Gly43 residue lies at the interstrand loop within blade 1 of the 7-blade  $\beta$ -propeller structure, on the same face that interacts with Ran (Figure 3A). Gly43 represents a highly conserved residue position: among 84 of 88 Rcc1 orthologues. The adoption of a Glycine residue (the smallest amino acid) at position 48 likely

enables close packing required within the Rcc1 propeller blade and may facilitate interaction with helix 4 of Ran (Figure 3B). A Glycine to Serine substitution may not be tolerated due to the larger sidechain of the Serine residue and could possibly destabilise the propeller blade and impact its interaction with Ran, impairing GTP exchange.

To assess Ran-Rcc1 binding *in vitro*, we performed a GTPase activity assay using purified recombinant wild-type (Rcc1<sup>WT</sup>) and p.Gly43Ser mutant (Rcc1<sup>G43S</sup>). In the GTPase assay a GEF-specific buffer was used to ensure GTP-loading of Ran is Rcc1-dependent. GTPase activity was quantified based on conversion of residual GTP from GTPase reactions to luminescence, where luminescence is inversely proportional to GTPase activity. A mild reduction in residual GTP was observed from reactions containing Rcc1<sup>G43S</sup>, compared to Rcc1<sup>WT</sup> (Figure 3C) which suggests the Rcc1<sup>G43S</sup> has slightly higher GEF activity, albeit not statistically significant.

Due to the thermosensitivity characterised in the Rcc1 mutant tsBN2 cell line (Ohtsubo et al., 1987), we considered that the Gly43Ser variant may induce thermolability in Rcc1. Thermostability of Rcc1 was assessed using a thermal shift assay. Rcc1<sup>WT</sup> displayed a relatively low melting point ( $T_m$ ) at 51.7 °C, however the  $T_m$  of Rcc1<sup>G43S</sup> was 4.9 °C lower, indicating increased thermolability compared to Rcc1<sup>WT</sup> (Figure 3D). Addition of purified Ran GTPase increased  $T_m$  in both reactions, indicating Ran interaction with Rcc1<sup>WT</sup> and Rcc1<sup>G43S</sup>, however the  $T_m$  for the reaction with Rcc1<sup>G43S</sup> remained 3.7 °C lower compared to Rcc1<sup>WT</sup>. In order to assess the impact of reduced thermal stability on GTPase activity, further GTPase assays were performed over a physiological temperature gradient of 35-42 °C. A rise in temperature led to a mild increase in GTP consumption in reactions with either Rcc1<sup>WT</sup> or Rcc1<sup>G43S</sup> (Figure 3E). However, similarly to initial GTPase assays conducted at ambient temperature, Rcc1<sup>G43S</sup> displayed only a mild nonsignificant increase in GTPase activity relative to Rcc1<sup>WT</sup>. The continuous activity of Rcc1<sup>G43S</sup> over a physiological range indicates increase in body temperature associated with febrile infection does not explain the molecular dysfunction resulting from the Gly43Ser variant identified in the proband (IV:2 of Family A).

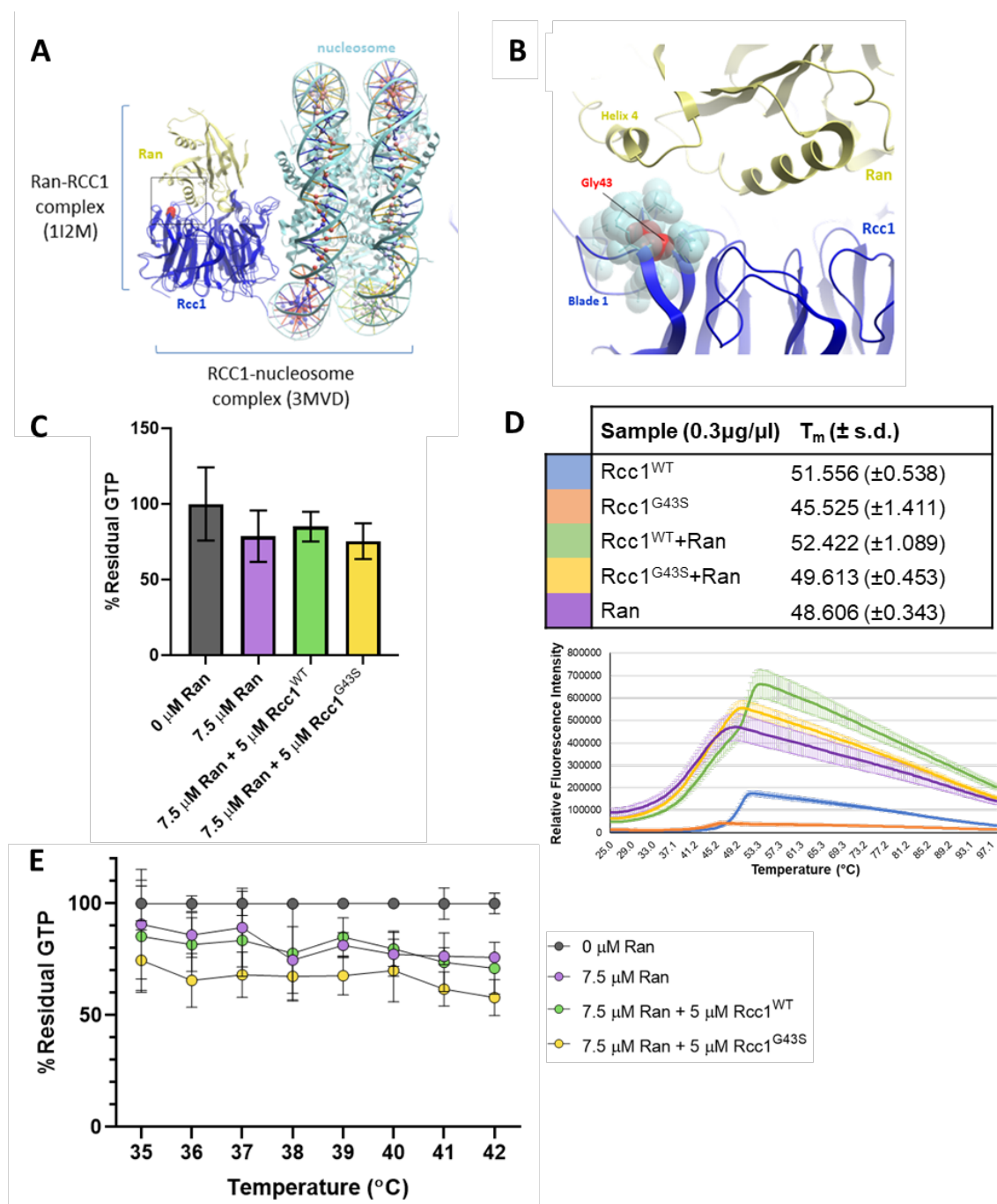


Figure 3: *In silico* and *in vitro* analyses of Rcc1<sup>G43S</sup>. (A) *In silico* modelling of Rcc1-Ran-Nucleosome complex with Rcc1 Gly43 highlighted in red. (B) *In silico* modelling highlighting position of Rcc1 Gly43 in blade 1 of the 7-blade propeller structure of Rcc1, in proximity to Ran during binding. (C) Ran GTPase activity assay with Rcc1<sup>WT</sup> and Rcc1<sup>G43S</sup>, measured by mean luminescence of residual GTP from GTPase reactions  $\pm$  s.d. (D) Table of average melting temperature ( $T_m$ )  $\pm$  s.d. of Rcc1<sup>WT</sup>, Rcc1<sup>G43S</sup>, Ran, and combination of Ran with either Rcc1<sup>WT</sup> or Rcc1<sup>G43S</sup>, derived from the peak in fluorescence from thermal shift assays (graph



represents one triplicate reactions from one assay). (E) Ran GTPase activity assay with Rcc1<sup>WT</sup> and Rcc1<sup>G43S</sup> conducted over a physiological temperature range (35-42 °C). (C,E) Results represent an average of two repeats each containing three individual reactions, Kruskal-Wallis  $p > 0.05$  (ns).

#### 6.3.5 Rcc1 localisation changes in fibroblasts of individuals with ANE

To better understand the effect of the Gly43Ser variant in a cellular context, we obtained dermal fibroblasts from the proband (IV:2 of family A). We hypothesised that Rcc1-dependent functions may be impaired as a result of altered Rcc1 localisation in cells cultured at high temperature. Febrile conditioned were modelled by assaying cells incubated at the nonpermissive temperature of 42 °C for 4 hours.

Using this heat-shock approach, we fixed and stained cells with immunofluorescent markers for Rcc1, chromatin, and active caspase-3 (Figure 4). Imaging of cells following heat-shock stimulation indicated that Rcc1 remains mainly localised to chromatin in control cells, with limited formation of perinuclear foci. However, heat-shock stimulation of proband fibroblasts led to a greater degree delocalisation of Rcc1 from chromatin, forming a more diffuse perinuclear pattern.

Despite the changes in Rcc1 localisation, no predisposition for chromatin condensation was noted in fibroblasts either under standard culture conditions or following heat-shock. Increase in active caspase-3 recognition indicates greater apoptotic activity in proband fibroblasts compared to control cells under standard culture conditions, while exposure to heat-shock leads to activation of caspase-3 to the same extent in control and proband fibroblasts.

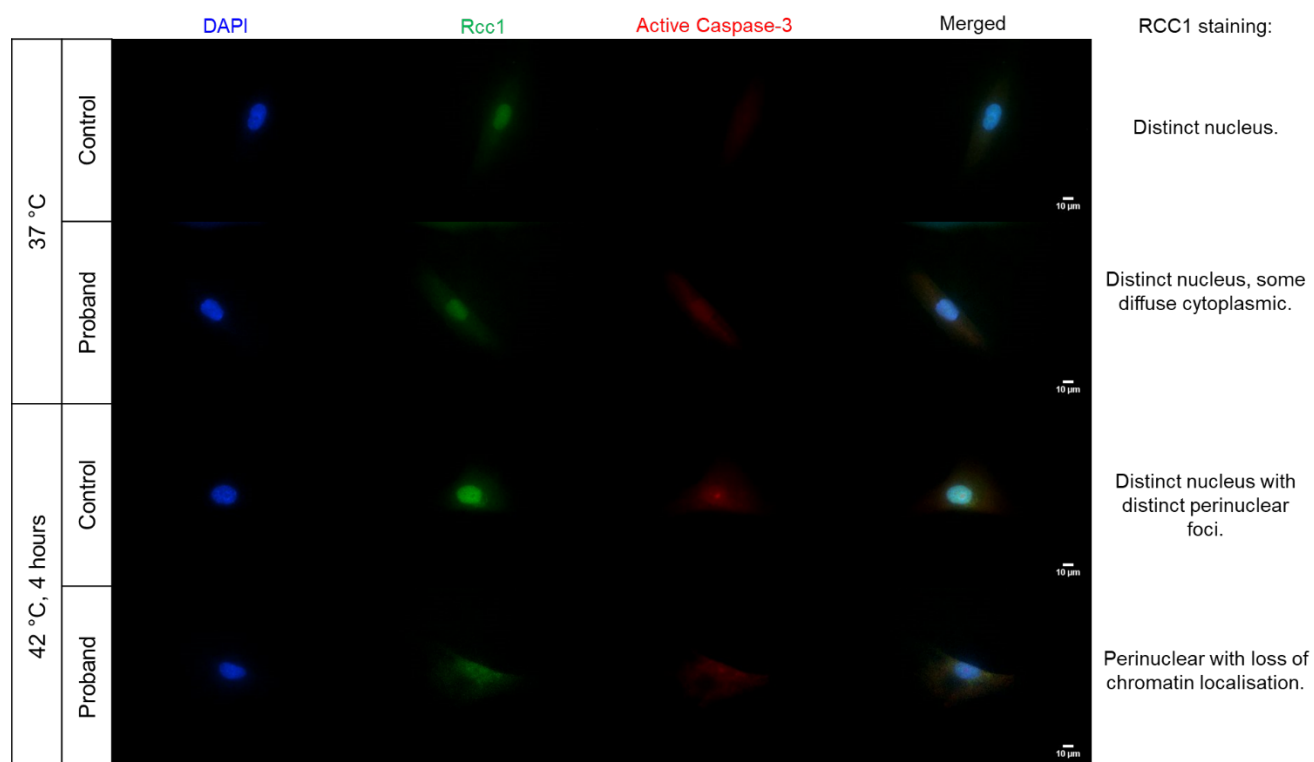


Figure 4: Immunofluorescence of heat-shocked fibroblasts. Proband (Family A, IV:2) dermal fibroblasts and control-derived dermal fibroblasts were fixed and stained after standard culture at 37 °C, or following four hours 42 °C heat-shock. Cells were stained with DAPI (blue), anti-RCC1 (green), and anti-Active Caspase-3 (red).

## 6.4 Discussion

This description of *RCC1*-related autosomal recessive ANE is the first identification of monogenic *RCC1*-related disease. No deleterious homozygous *RCC1* variants are recorded in gnomAD, while homozygous missense variants are only seen in six individuals. *RCC1* is well-constrained according to gnomAD (pLI 0.93, missense Z 1.68), and there is complete absence of homozygous nonsense variants (Karczewski et al., 2019). The identification of further affected families, particularly the affected individuals in family C support the notion that biallelic variants in *RCC1* are deleterious. *RCC1* encodes regulator of chromatin condensation 1, which is a multifunctional nuclear protein, with principal function as a guanine exchange factor (GEF) for Ran, a small GTPase involved in coordinating mitosis and nucleocytoplasmic transport (Seki et al., 1996).

### 6.4.1 Implications of Rcc1 chromatin delocalisation

Our findings indicate that Rcc1-directed Ran GTP exchange is unaffected by the Gly43Ser variant across a range of physiological temperatures (Figure 3E), and that Rcc1 is delocalised from chromatin in heat-shocked proband cells (Figure 4). This suggests that while Ran-GTP exchange continues, Ran-GTP gradients may be disrupted by delocalisation from chromatin, rather than inhibition of GTP-charging of Ran. This is supportive of an acute response to a stimulus rather than a congenital deficiency in Rcc1 activity, which is consistent with the acute nature of the phenotype described in the affected individuals. The segregation of GTP-loading within the nucleus, from Ran GTPase activity in the cytoplasm is essential for cellular processes orchestrated by Ran-GTP activity (Clarke and Zhang, 2008).

The observation of Rcc1 delocalisation from chromatin in fibroblasts from an affected individual allows a hypothesis of Ran-GTP gradient disruption in response to a stimulus of cellular stress. This is consistent with molecular changes in nucleocytoplasmic transport activity observed in patients affected by ANE with *NUP214* or *RANBP2* variants (Levine et al., 2020). Nup214 is a key nucleoporin which facilitates the nucleocytoplasmic export of cargoes complexed with Ran-GTP and CRM1 (chromosomal maintenance 1) from the nucleus (Hamed et al., 2021), while the four Ran-GTPase-binding domains of RanBP2 facilitate activation of Ran GTPase activity, and disassembly of the transport complex

following efflux from the nucleus (Levine et al., 2020). Defects in nucleocytoplasmic export and proteostasis are observed in a murine *Ranbp2* conditional Thy1<sup>+</sup> motor neuron knockout model of amyotrophic lateral sclerosis (Cho et al., 2017). Additionally, fibroblasts from patients with *NUP214* variants are characterised by an increase in dysmorphic nuclei (Shamseldin et al., 2019), increased susceptibility to apoptosis following heat shock exposure, and impaired transport of poly(A)-RNA from the nucleus to the cytoplasm compared to controls (Fichtman et al., 2019). To date, nucleocytoplasmic export changes have not been identified in relation of *Rcc1* defects, however nucleocytoplasmic import was assessed in temperature-sensitive hamster kidney cell line tsBN2, which carry a homozygous *Rcc1* Ser256Phe variant (Nishijima et al., 2000; Ohtsubo et al., 1987). Culture of tsBN2 cells at the non-permissive temperature of 39.5 °C impaired import of a nuclear localisation signal-tagged proteins (Tachibana et al., 1994). Further work is therefore required to compare defects in nucleocytoplasmic transport in *Rcc1*-mutant cells and compare the effects to the changes observed in cells with *RANBP2* and *NUP214* variants to elucidate the mechanism of cellular dysfunction in ANE.

#### 6.4.2 Induction of *Rcc1* dysfunction

Although initial characterisation of *Rcc1* related to temperature-sensitivity, culture at non-permissive temperature is not the only stimulus to induce Ran-related dysfunction in tsBN2 cells. Stimulation of cell stress with ultraviolet laser, or oxidative stress by supplementation of cell cultures with oxidants such as diamide, hydrogen peroxide, or menadione, also generated impaired *Rcc1*-associated Ran-GTP loading and led to a reduction in Ran protein gradient (Chatterjee and Paschal, 2015). This more complex interaction was not considered during our analysis of changes in GTPase activity using recombinant protein (Figure 3D), which may explain why no reduction of GTPase activity was observed with *Rcc1*<sup>G43S</sup>. This therefore suggests that an increase in core body temperature during febrile illness is not a prerequisite for phenotype onset in patients with biallelic *RCC1* variants. In *RANBP2*-related ANE, multiple hypotheses exist for the precipitation of disease, including cytokine-storm (Pongpitakmetha et al., 2022), and mitochondrial dysfunction (Hu et al., 2022; Levine et al., 2020).

Onset of oxidative stress during inflammatory processes, such as during infection, may contribute significantly to the pathology, rather than temperature alone. This is contrary to the thermolability resulting from *CPT2* polymorphisms in acute encephalopathy observed among East Asian patients, where a temperature rise causes structural and functional disruption to the CPT2 protein (Kumakura et al., 2011; Shinohara et al., 2011). The interplay between metabolic stress and ANE1 is observed in *Ranbp2* knockout mice, which display a reduction in ATP, and hexokinase 1 (Aslanukov et al., 2006). Crucially, RanBP2 also regulates mitochondrial metabolism by binding to regulator of ATP production Cox11 (Shibata et al., 2021). RanBP2-Cox11 binding is reduced in lymphoblasts of ANE1 patients with *RANBP2* variants (Shibata et al., 2021). Diagnostic testing of the proband (IV:2 of family A) and individual III:1 of family C revealed no defects in mitochondrial respiratory chain analysis, indicating no innate dysfunction of mitochondrial activity. Although this analysis did not consider the context of exogenous cell stress stimulation necessary for Rcc1-related dysfunction, the mitochondrial actions of RanBP2 are functions independent of RanGTPase activation (Patil et al., 2013). The identification of Rcc1-related ANE suggests the pathogenesis of *RANBP2* variants results from RanGTPase-dependent functions in nucleocytoplasmic exchange.

Induction of disease among individuals diagnosed with ANE1 is usually secondary to febrile infection. The role of cytokine storm, and specifically interleukin-6 (IL-6) is proposed as a key element of disease induction (Levine et al., 2020; Singh et al., 2015), with excessive cytokine secretion observed in 40 % of individuals with ANE1 (Shen et al., 2021). In a case of ANE1 induced following SARS-CoV-2 infection, marked elevation of IL-6 in addition to ferritin, procalcitonin, and high sensitivity C-reactive protein were recorded four days after admission (Pongpitakmetha et al., 2022). This patient carrying a heterozygous *RNABP2* missense variant went on to develop urinary tract infection-related septic shock, acute cerebral infarction, ANE, and multiorgan failure with COVID-19. Although there was no significant pulmonary involvement, the patient died six days after admission (Pongpitakmetha et al., 2022). RanBP2 depletion has implicated in inhibition of translational silencing of IL-6, as a result of RanBP2's role in directing degradation of IL-6 messenger ribonucleoprotein-associated Argonautes (Shen et al., 2021). Despite this interaction, assaying of RanBP2 with ANE1-relevant missense variants remained able to silence IL-6

mRNA (Shen et al., 2021). The absence of the hypothesised mechanism may be explained by the lack of stress stimulus in this model, which suggests that innate IL-6 regulation errors by RanBP2 do entirely explain the pathogenesis.

#### 6.4.3 Ran-related dysfunction in hereditary neuropathies

The episodic demyelinating nature of the polyneuropathy in individual II:1 of family C was similar to the presentation of Guillain-Barré syndrome (GBS). A similar pattern of demyelinating neuropathy was also observed in Family B. GBS is a post-infectious inflammatory demyelinating polyneuropathy typically arising 10-14 days after infection with *Campylobacter jejuni* (Lunn, 2022). In 60 % of GBS cases, molecular mimicry of gangliosides by *C. jejuni* structures results in generation of autoantibodies against axonal glycolipids (Kaida, 2019). Rare occurrences of familial GBS suggests a genetic basis for susceptibility (Bar-Joseph et al., 1991; Korn-Lubetzki et al., 2002), however the genetic aetiology appears either elusive or complex in many cases. Genetic variants affecting myelin proteins (P0, P2, and PMP22), and immunological genes including human leukocyte antigens (HLA), cluster of differentiation (CD) molecules, tumour necrosis factor alpha (TNF- $\alpha$ ), interleukin-17 (IL-17), intracellular adhesion molecule-1 (ICAM1) have been associated with familial forms of GBS (Khanmohammadi et al., 2021).

Dysregulation of myelination was identified in conditional motor neuron *Ranbp2*<sup>-/-</sup> mice, which were used to generate disease resembling amyotrophic lateral sclerosis (ALS) (Cho et al., 2017). This model was characterised by dysregulation of nucleocytoplasmic transport and proteostasis as previously reported in patients with *RANBP2* variants (Cho et al., 2014), but also indicated dysregulated chemokine signalling, likely responsible for a paracrine dysregulation of heterogeneous nuclear ribonuclear protein H3 (hnRNPH3) and matrix metalloproteinase 28 (Mmp28) in Schwann cells (Cho et al., 2017). In ALS, proteostasis defects account for the accumulation of neurotoxic protein substrates relating to *C9orf72* repeat expansions (MIM: 614260) (Westergard et al., 2019). Defects in proteostasis are also implicated in multiple diseases associated with regulation of eukaryotic translation initiation factor 2 subunit A (eIF2A). The eIF2a protein subunit acts as a global repressor of protein translation in response to cellular stress but upregulates translation of neurotoxic dipeptide

repeats which accumulate into stress granules and impair nucleocytoplasmic transport (Guo et al., 2022; Zhang et al., 2018). Variants in eIF2A kinase genes are associated with a group of genetic leukoencephalopathies (Guo et al., 2022). Of specific interest to our work is the onset of severe neurologic decompensation secondary to fever associated with heterozygous missense variants in *EIF2AK2* (MIM: 176871) (Mao et al., 2020).

By considering other genetic causes of acute onset neurological deficits, a molecular framework emerges whereby cellular stress-related disruption of nucleocytoplasmic transport and proteostasis are implicated in *RANBP2*-related ANE1 and ALS-like disease, and *EIF2AK2*-related leukoencephalopathy. The appearance of GBS-like demyelinating disease in individuals identified here with biallelic *RCC1* missense variants suggests pathology underlying these patients could result from dysfunctional nucleocytoplasmic transport and proteostasis, possibly contributing to myelination defects which also underlie the pathogenesis of GBS.

Here we describe a novel form of autosomal recessive acute necrotising encephalopathy associated with homozygous *RCC1* missense variants, which is characterised by delocalisation of Rcc1 from chromatin following exposure to heat-shock. We propose a mechanism whereby cellular stress associated with febrile illness leads to the disruption Ran-dependent processes involved in nucleocytoplasmic transport, ultimately leading to neuronal cell dysfunction. Further research should aim to establish the effect of biallelic *RCC1* variants on nucleocytoplasmic transport, and implications in neuronal signalling to establish the relationship between Ran-related dysfunction and other acute onset neuropathies.

## 6.5 Materials and Methods

Informed written consent was provided by the parents of the affected children and by the parents themselves for their inclusion in the study under the NHS ethics approval reference (11/H1003/3) and the University of Manchester ethics committee.

### 6.5.1 Autozygosity mapping

Autozygosity mapping was performed on DNA from two affected individuals (IV:4 and IV:6) from Family A using the Affymetrix Genome-wide SNP 6.0 microarray (Affymetrix, UK). Genotypes were generated using Birdseed v2 algorithm with a confidence threshold of 0.01 within the Affymetrix Genotyping console. Autozygosity mapping was performed using the AutoSNPa software (<http://dna.leeds.ac.uk/autosnpa/>) (Carr et al., 2006).

### 6.5.2 Exome Sequencing

Whole exome sequencing was performed on DNA from IV:6 of family A. The Agilent SureSelect Human All Exon Kit v4 (Aligent Technologies, UK) was used for library preparation and sequencing was performed on the HiSeq 2500 (Illumina, UK). Sequence data were mapped to the hg19 human reference genome using Burrows-Wheeler aligner software (version 0.6.2; <http://bio-bwa.sourceforge.net>). Genome Analysis Tool Kit software (version 2.4.7; <https://www.broadinstitute.org/gatk>) was used for base quality score recalibration and the unified genotyper (<https://www.broadinstitute.org/gatk>) was used for variant calling. Genetic variants were annotated and filtered using VarSeq<sup>TM</sup> c2.2 software (Golden Helix Inc., MT).

### 6.5.3 Generation of constructs for expression of recombinant Rcc1 and Ran proteins

To conduct *in vitro* assays, constructs were generated for the expression of recombinant Rcc1 and Ran proteins. *RAN* and *RCC1* cDNA was amplified from control cDNA using primers designed to introduce 3' His-tags (*RAN*:

GCCGGGATCCCATATGGCTGCGCAGGGAGAGCCCCAGG and GCGCGGATCC

TCACAGGTCATCATCCTCATCC; *RCC1*:

GGGGGGATCCCATATGTCACCCAAGCGCATAGCTAAAAGAAGG and



GGGGGGATCCTCAGCTCTGTTCTTTGTCCTTGAC). cDNA fragments were ligated into a high copy number pBlueScript (pBS) vector to permit blue/white colony selection following transformation into *E. coli* XL-1 blue ultra-competent bacteria (XL-Gold, Aligent Technologies). Plasmids were extracted from white colonies using the GenElute Plasmid miniprep Kit (Sigma) and sequences were confirmed by DNA sequencing.

#### 6.5.4 Production of single-stranded DNA for site-directed mutagenesis

To produce a construct to express the *RCC1* variant, pBS-*RCC1* ssDNA was produced as a template for site-directed mutagenesis by growing up a single pBS-*RCC1* colony with the VCSM123 helper phage (Aligent Technologies) for 24 hours at 37 °C. 2µg/ml uridine was added to the culture to ensure that WT ssDNA templates would degrade during subsequent amplification of the mutagenized construct. ssDNA was extracted from the aqueous phase by phase separation with TE-buffered Phenol:Chloroform:isoamyl alcohol.

#### 6.5.5 Site-directed mutagenesis of *RCC1*

Site-specific mutagenesis of *RCC1* was facilitated by designing a synthetic oligonucleotide containing the c.127G>A variant (ACACTAGGCCAGAGCGACGTGGGCC). The oligonucleotide was 5'-phosphorylated to ensure unidirectional extension, then hybridised to pBS-*RCC1* ssDNA. The mutagenesis oligonucleotide was then extended with T4 DNA polymerase to produce a pBS-*RCC1*-G43S construct. pBS-*RCC1*-G43S was amplified in *E. coli* XL-1 blue ultra-competent bacteria (XL-Gold, Aligent Technologies). The c.127C>G variant was confirmed by DNA sequencing.

#### 6.5.6 Expression and purification of recombinant Rcc1 and Ran

His-tagged *RAN*, *RCC1*-WT, and *RCC1*-G43S c.DNA fragments from pBS constructs were cloned into low copy pET-28a plasmid suitable for protein expression. Protein was expressed in *E. coli* Rosetta2 (DE3) bacteria (Novagen) in 500ml Overnight Express media (Novagen) for 36 hours at 37 °C (Ran and Rcc1-WT) or 19 °C (Rcc1-G43S). Cultures were pelleted by centrifugation at 7000 rpm for 15 min at 4 °C, and cells lysed in lysis/wash buffer (20 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1 M DTT, 0.02 % Tween-20, 10 mM imidazole, 15 %

glycerol). Cell lysis was completed with 3x sonication for 10 sec at 30 % amplitude. Lysate was cleared by centrifugation at 17000 rpm for 15 min at 4 °C. His-tagged proteins were column-purified using His-Select Nickel Affinity Gel (Sigma). Purified proteins were eluted in 4x 1 mL fractions with elution buffer (20 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1 M DTT, 0.02 % Tween-20, 250 mM imidazole, 15 % glycerol), and purification was evaluated via SDS-PAGE. Once purity was confirmed protein fractions were dialysed in dialysis buffer (20 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1 M DTT, 15 % glycerol) overnight at 4 °C to facilitate removal of imidazole. Proteins were aliquoted and stored at -80 °C until use.

#### 6.5.7 GEF assay

To establish ability of recombinant Rcc1 to facilitate guanine exchange with Ran GTPase, GEF assays were performed using the GTPaseGlo™ Assay kit (Promega). Addition of GEF buffer containing Mg<sup>2+</sup> ensures nucleotide loading is catalysed solely by the GEF (Rcc1). Assays were conducted with 7.5 µg Ran protein, combined with 5 µl Rcc1<sup>WT</sup> or Rcc1<sup>G43S</sup> in 10 µL reactions with 10 µM GTP and 1 mM DTT. GTPase reactions were incubated for one hour at a range of physiological temperatures between 35-42 °C using Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad). Residual GTP was converted to ATP by addition of 10 µL GTPase-Glo Reagent and incubation at room temperature with agitation for 30 min. Bioluminescence was generated from converted ATP by addition of 20 µl Ultra-Glo™ Recombinant Luciferase Detection Reagent. Luminescence was read using Infinite 200 PRO microplate reader (Tecan) 10 min after addition of detection reagent. Luminescence values were normalised to reactions without Ran or Rcc1. Reactions were conducted in triplicate within each assay and results represent mean of three assays ± standard deviation.

#### 6.5.8 Thermal Shift Assay

In order to assess thermal stability of recombinant Rcc1 proteins, thermal shift assays were conducted using Protein Thermal Shift™ Dye Kit (Applied Biosystems). 0.3 µg recombinant Rcc1<sup>WT</sup> or Rcc1<sup>G43S</sup> protein was combined with Protein Thermal Shift™ Dye Kit Buffer and 8x Dye in 20 µL reactions in a 96 well PCR plate. Reactions were also conducted with the addition of 0.3 µg recombinant Ran protein. ROX reporter dye fluorescence was measured

using a StepOnePlus™ Real-Time PCR System (Applied Biosystems) after each 1 % increase in temperature between 25 °C and 99 °C. Melting temperature ( $T_M$ ) was derived from the peak in fluorescence from plotting melt curves of fluorescence against temperature. Reactions were conducted in triplicate within each assay. Results represent the mean of three assays  $\pm$  standard deviation.

#### 6.5.9 Culture of patient fibroblasts

Patient and control fibroblasts were obtained from the Cell Bank at the Genomic Medicine Department, Manchester University HNS Foundation Trust. Cells were maintained under standard conditions (37.0 °C, 5.0 % CO<sub>2</sub>) using Dulbecco's Modified Eagle Medium (DMEM) with high glucose (Sigma) supplemented with 10 % foetal bovine serum (Gibco) and 1 % Penicillin-Streptomycin (Gibco).

#### 6.5.10 Immunofluorescence Imaging

To assess Rcc1 localisation and response to heat shock, immunofluorescence of patient and control fibroblasts was conducted. 10,000 cells were seeded on sterile glass coverslips in 24 well plates. After 24 hours cells were subject either to heat shock by placing in an incubator set to 42.0 °C for 2 or 6 hours or kept under standard culture conditions. Following heat shock, cells were returned to standard 37.0 °C culture conditions for recovery for 2 hours. To stain active mitochondria, cells were stained with 500 nM MitoTracker Red CMXRos probe (Invitrogen) diluted in DMEM for 30 min prior to fixation. Cells were washed twice with warm phosphate buffered saline (PBS, Gibco), fixed with 3.8 % formaldehyde (Sigma) for 15 min at room temperature, then permeabilised with 0.1 % Triton-X 100 solution (Sigma). Cells were washed twice in PBS and were blocked with 1.0 % bovine serum albumin (BSA, Sigma) for 30 min at room temperature. Cells were incubated anti-Rcc1 rabbit monoclonal antibody (ab181155, Abcam) and anti-Active Caspase-3 mouse monoclonal antibody (bsm-33199M, Bioss Antibodies) diluted 1:250 in 1.0 % BSA for 1 hour at room temperature. Cells were washed twice with PBS before applying secondary Alexafluor-488 Goat anti-Rabbit (AB\_2338046, Jackson ImmunoResearch) and AlexaFluor-647 Goat anti-Mouse (AB\_2535804, Invitrogen) antibodies diluted 1:500 and 1:1000 respectively in 1.0 %

BSA for 1 hour at room temperature in the dark. Cover slips were washed twice with PBS to remove unbound antibody, then adhered to glass microscope slides with ProLong Gold Anti-fade Mountant with DAPI (Invitrogen) overnight at room temperature in the dark. Slides were imaged within one week of staining using a Zeiss Axio Imager.D2 upright microscope using a 63x objective and captured using a Coolsnap HQ2 camera (Photometrics) through Micromanager software v1.4.23. Images were processed and analysed using Fiji ImageJ (<http://imagej.net/Fiji/Downloads>).

## 6.6 References

- Andersen, L.L., Mørk, N., Reinert, L.S., Kofod-Olsen, E., Narita, R., Jørgensen, S.E., Skipper, K.A., Höning, K., Gad, H.H., Østergaard, L., Ørntoft, T.F., Hornung, V., Paludan, S.R., Mikkelsen, J.G., Fujita, T., Christiansen, M., Hartmann, R., Mogensen, T.H., 2015. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *Journal of Experimental Medicine* 212, 1371–1379. <https://doi.org/10.1084/jem.20142274>
- Aslanukov, A., Bhowmick, R., Guraju, M., Oswald, J., Raz, D., Bush, R.A., Sieving, P.A., Lu, X., Bock, C.B., Ferreira, P.A., 2006. RanBP2 modulates Cox11 and hexokinase I activities and haploinsufficiency of RanBP2 causes deficits in glucose metabolism. *PLoS Genet* 2, e177. <https://doi.org/10.1371/journal.pgen.0020177>
- Bar-Joseph, G., Etzioni, A., Hemli, J., Gershoni-Baruch, R., 1991. Guillain-Barré syndrome in three siblings less than 2 years old. *Archives of Disease in Childhood* 66, 1078–1079. <https://doi.org/10.1136/adc.66.9.1078>
- Carr, I.M., Flintoff, K.J., Taylor, G.R., Markham, A.F., Bonthron, D.T., 2006. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. *Human Mutation* 27, 1041–1046. <https://doi.org/10.1002/humu.20383>
- Casrouge, A., Zhang, S.-Y., Eidenschenk, C., Jouanguy, E., Puel, A., Yang, K., Alcais, A., Picard, C., Mahfoufi, N., Nicolas, N., Lorenzo, L., Plancoulaine, S., Sénéchal, B., Geissmann, F., Tabeta, K., Hoebe, K., Du, X., Miller, R.L., Héron, B., Mignot, C., de Villemeur, T.B., Lebon, P., Dulac, O., Rozenberg, F., Beutler, B., Tardieu, M., Abel, L., Casanova, J.-L., 2006. Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* 314, 308–312. <https://doi.org/10.1126/science.1128346>
- Chatterjee, M., Paschal, B.M., 2015. Disruption of the ran system by cysteine oxidation of the nucleotide exchange factor RCC1. *Mol Cell Biol* 35, 566–581. <https://doi.org/10.1128/MCB.01133-14>
- Chen, T., Muratore, T.L., Schaner-Tooley, C.E., Shabanowitz, J., Hunt, D.F., Macara, I.G., 2007. N-terminal alpha-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat Cell Biol* 9, 596–603. <https://doi.org/10.1038/ncb1572>
- Chew, H.B., Ngu, L.H., 2020. RANBP2 susceptibility to infection-induced encephalopathy: Clinicoradiologic and molecular description in a Malaysian family. *Mol Genet Metab Rep* 24, 100627. <https://doi.org/10.1016/j.ymgmr.2020.100627>
- Cho, K., Patil, H., Senda, E., Wang, J., Yi, H., Qiu, S., Yoon, D., Yu, M., Orry, A., Peachey, N.S., Ferreira, P.A., 2014. Differential loss of prolyl isomerase or chaperone activity of Ran-binding protein 2 (Ranbp2) unveils distinct physiological roles of its cyclophilin domain in proteostasis. *J Biol Chem* 289, 4600–4625. <https://doi.org/10.1074/jbc.M113.538215>
- Cho, K.-I., Yoon, D., Qiu, S., Danziger, Z., Grill, W.M., Wetsel, W.C., Ferreira, P.A., 2017. Loss of Ranbp2 in motoneurons causes disruption of nucleocytoplasmic and chemokine signaling, proteostasis of hnRNPH3 and Mmp28, and development of amyotrophic lateral sclerosis-like syndromes. *Dis Model Mech* 10, 559–579. <https://doi.org/10.1242/dmm.027730>
- Clarke, P.R., Zhang, C., 2008. Spatial and temporal coordination of mitosis by Ran GTPase. *Nat Rev Mol Cell Biol* 9, 464–477. <https://doi.org/10.1038/nrm2410>
- Fichtman, B., Harel, T., Biran, N., Zagairy, F., Applegate, C.D., Salzberg, Y., Gilboa, T., Salah, S., Shaag, A., Simanovsky, N., Ayoubieh, H., Sobreira, N., Punzi, G., Pierri, C.L., Hamosh, A., Elpeleg, O., Harel, A., Edvardson, S., 2019. Pathogenic Variants in NUP214 Cause “Plugged” Nuclear Pore Channels and Acute Febrile Encephalopathy. *Am J Hum Genet* 105, 48–64. <https://doi.org/10.1016/j.ajhg.2019.05.003>
- Guo, S., Nguyen, L., Ranum, L.P.W., 2022. RAN proteins in neurodegenerative disease: Repeating themes and unifying therapeutic strategies. *Curr Opin Neurobiol* 72, 160–170. <https://doi.org/10.1016/j.conb.2021.11.001>

- Hamed, M., Caspar, B., Port, S.A., Kehlenbach, R.H., 2021. A nuclear export sequence promotes CRM1-dependent targeting of the nucleoporin Nup214 to the nuclear pore complex. *J Cell Sci* 134, jcs258095. <https://doi.org/10.1242/jcs.258095>
- Holzer, G., Antonin, W., 2022. Nup50 plays more than one instrument. *Cell Cycle* 21, 1785–1794. <https://doi.org/10.1080/15384101.2022.2074742>
- Hu, Y., Tian, Z., Zhao, B., Dong, C., Cao, L., 2022. A novel variation in RANBP2 associated with infection-triggered familial acute necrotizing encephalopathy. *Neurol Sci* 43, 3973–3977. <https://doi.org/10.1007/s10072-022-06033-8>
- Huang, X., Li, Y., Yu, Y., Yang, S., Li, M., Li, T., Huang, L., Tao, J., Zhang, M., Delwart, E., Zhang, J., 2020. Human herpesvirus 6-associated acute necrotizing encephalopathy in an infant with a mutation in the RANBP2 gene. *J Paediatr Child Health* 56, 1308–1310. <https://doi.org/10.1111/jpc.14777>
- Hutten, S., Kehlenbach, R.H., 2006. Nup214 is required for CRM1-dependent nuclear protein export in vivo. *Mol Cell Biol* 26, 6772–6785. <https://doi.org/10.1128/MCB.00342-06>
- Jiang, J., Wang, Y.E., Palazzo, A.F., Shen, Q., 2022. Roles of Nucleoporin RanBP2/Nup358 in Acute Necrotizing Encephalopathy Type 1 (ANE1) and Viral Infection. *Int J Mol Sci* 23, 3548. <https://doi.org/10.3390/ijms23073548>
- Kaida, K., 2019. Guillain-Barré Syndrome. *Adv Exp Med Biol* 1190, 323–331. [https://doi.org/10.1007/978-981-32-9636-7\\_20](https://doi.org/10.1007/978-981-32-9636-7_20)
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Consortium, T.G.A.D., Neale, B.M., Daly, M.J., MacArthur, D.G., 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210. <https://doi.org/10.1101/531210>
- Khanmohammadi, S., Malekpour, M., Jabbari, P., Rezaei, N., 2021. Genetic basis of Guillain-Barre syndrome. *J Neuroimmunol* 358, 577651. <https://doi.org/10.1016/j.jneuroim.2021.577651>
- Korn-Lubetzki, I., Argov, Z., Raas-Rothschild, A., Wirguin, I., Steiner, I., 2002. Family with inflammatory demyelinating polyneuropathy and the HNPP 17p12 deletion. *Am J Med Genet* 113, 275–278. <https://doi.org/10.1002/ajmg.10725>
- Kumakura, A., Iida, C., Saito, M., Mizuguchi, M., Hata, D., 2011. Pandemic Influenza A-Associated Acute Necrotizing Encephalopathy Without Neurologic Sequelae. *Pediatric Neurology* 45, 344–346. <https://doi.org/10.1016/j.pediatrneurol.2011.08.004>
- Lafaille, F.G., Harschnitz, O., Lee, Y.S., Zhang, P., Hasek, M.L., Kerner, G., Itan, Y., Ewaleifoh, O., Rapaport, F., Carlile, T.M., Carter-Timofte, M.E., Paquet, D., Dobbs, K., Zimmer, B., Gao, D., Rojas-Duran, M.F., Kwart, D., Rattina, V., Ciancanelli, M.J., McAlpine, J.L., Lorenzo, L., Boucherit, S., Rozenberg, F., Halwani, R., Henry, B., Amenzoui, N., Alsum, Z., Marques, L., Church, J.A., Al-Muhsen, S., Tardieu, M., Bousfiha, A.A., Paludan, S.R., Mogensen, T.H., Quintana-Murci, L., Tessier-Lavigne, M., Smith, G.A., Notarangelo, L.D., Studer, L., Gilbert, W., Abel, L., Casanova, J.-L., Zhang, S.-Y., 2019. Human SNORA31 variations impair cortical neuron-intrinsic immunity to HSV-1 and underlie herpes simplex encephalitis. *Nat Med* 25, 1873–1884. <https://doi.org/10.1038/s41591-019-0672-3>
- Lee, Y.-J., Hwang, S.-K., Lee, S.M., Kwon, S., 2017. Familial acute necrotizing encephalopathy with RANBP2 mutation: The first report in Northeast Asia. *Brain Dev* 39, 625–628. <https://doi.org/10.1016/j.braindev.2017.02.005>

- Levine, J.M., Ahsan, N., Ho, E., Santoro, J.D., 2020. Genetic Acute Necrotizing Encephalopathy Associated with RANBP2: Clinical and Therapeutic Implications in Pediatrics. *Mult Scler Relat Disord* 43, 102194. <https://doi.org/10.1016/j.msard.2020.102194>
- Lunn, M.P., 2022. Guillain-Barré syndrome in an era of global infections and 21st century vaccination. *Curr Opin Neurol* 35, 571–578. <https://doi.org/10.1097/WCO.0000000000001086>
- Mao, D., Reuter, C.M., Ruzhnikov, M.R.Z., Beck, A.E., Farrow, E.G., Emrick, L.T., Rosenfeld, J.A., Mackenzie, K.M., Robak, L., Wheeler, M.T., Burrage, L.C., Jain, M., Liu, P., Calame, D., Küry, S., Sillesen, M., Schmitz-Abe, K., Tonduti, D., Spaccini, L., Iacone, M., Genetti, C.A., Koenig, M.K., Graf, M., Tran, A., Alejandro, M., Undiagnosed Diseases Network, Lee, B.H., Thiffault, I., Agrawal, P.B., Bernstein, J.A., Bellen, H.J., Chao, H.-T., 2020. De novo EIF2AK1 and EIF2AK2 Variants Are Associated with Developmental Delay, Leukoencephalopathy, and Neurologic Decompensation. *Am J Hum Genet* 106, 570–583. <https://doi.org/10.1016/j.ajhg.2020.02.016>
- Mariotti, P., Iorio, R., Frisullo, G., Plantone, D., Colantonio, R., Tartaglione, T., Batocchi, A.P., Valentini, P., 2010. Acute necrotizing encephalopathy during novel influenza A (H1N1) virus infection. *Ann Neurol* 68, 111–114. <https://doi.org/10.1002/ana.21996>
- Martin, A., Reade, E.P., 2010. Acute necrotizing encephalopathy progressing to brain death in a pediatric patient with novel influenza A (H1N1) infection. *Clin Infect Dis* 50, e50–52. <https://doi.org/10.1086/651501>
- Neilson, D.E., Adams, M.D., Orr, C.M.D., Schelling, D.K., Eiben, R.M., Kerr, D.S., Anderson, J., Bassuk, A.G., Bye, A.M., Childs, A.-M., Clarke, A., Crow, Y.J., Di Rocco, M., Dohna-Schwake, C., Dueckers, G., Fasano, A.E., Gika, A.D., Giannis, D., Gorman, M.P., Grattan-Smith, P.J., Hackenberg, A., Kuster, A., Lentschig, M.G., Lopez-Laso, E., Marco, E.J., Mastroianni, S., Perrier, J., Schmitt-Mechelke, T., Servidei, S., Skardoutsou, A., Uldall, P., van der Knaap, M.S., Goglin, K.C., Tefft, D.L., Aubin, C., de Jager, P., Hafler, D., Warman, M.L., 2009. Infection-Triggered Familial or Recurrent Cases of Acute Necrotizing Encephalopathy Caused by Mutations in a Component of the Nuclear Pore, RANBP2. *The American Journal of Human Genetics* 84, 44–51. <https://doi.org/10.1016/j.ajhg.2008.12.009>
- Nishijima, H., Seki, T., Nishitani, H., Nishimoto, T., 2000. Premature chromatin condensation caused by loss of RCC1. *Prog Cell Cycle Res* 4, 145–156. [https://doi.org/10.1007/978-1-4615-4253-7\\_13](https://doi.org/10.1007/978-1-4615-4253-7_13)
- Ohashi, E., Hayakawa, I., Murofushi, Y., Kawai, M., Suzuki-Muromoto, S., Abe, Y., Yoshida, M., Kono, N., Kosaki, R., Hoshino, A., Mizuguchi, M., Kubota, M., 2021. Recurrent acute necrotizing encephalopathy in a boy with RANBP2 mutation and thermolabile CPT2 variant: The first case of ANE1 in Japan. *Brain Dev* 43, 873–878. <https://doi.org/10.1016/j.braindev.2021.04.009>
- Ohtsubo, M., Kai, R., Furuno, N., Sekiguchi, T., Sekiguchi, M., Hayashida, H., Kuma, K., Miyata, T., Fukushima, S., Murotsu, T., 1987. Isolation and characterization of the active cDNA of the human cell cycle gene (RCC1) involved in the regulation of onset of chromosome condensation. *Genes Dev* 1, 585–593. <https://doi.org/10.1101/gad.1.6.585>
- Patil, H., Cho, K., Lee, J., Yang, Y., Orry, A., Ferreira, P.A., 2013. Kinesin-1 and mitochondrial motility control by discrimination of structurally equivalent but distinct subdomains in Ran-GTP-binding domains of Ran-binding protein 2. *Open Biol* 3, 120183. <https://doi.org/10.1098/rsob.120183>
- Pongpitakmetha, T., Hemachudha, P., Rattanawong, W., Thanapornsanguth, P., Viswanathan, A., Hemachudha, T., 2022. COVID-19 related acute necrotizing encephalopathy with extremely high interleukin-6 and RANBP2 mutation in a patient with recently immunized inactivated virus vaccine and no pulmonary involvement. *BMC Infect Dis* 22, 640. <https://doi.org/10.1186/s12879-022-07610-0>

- Poyiadji, N., Shahin, G., Noujaim, D., Stone, M., Patel, S., Griffith, B., 2020. COVID-19-associated Acute Hemorrhagic Necrotizing Encephalopathy: Imaging Features. *Radiology* 296, E119–E120. <https://doi.org/10.1148/radiol.2020201187>
- Ren, M., Drivas, G., D'Eustachio, P., Rush, M.G., 1993. Ran/TC4: a small nuclear GTP-binding protein that regulates DNA synthesis. *J Cell Biol* 120, 313–323. <https://doi.org/10.1083/jcb.120.2.313>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17, 405–424. <https://doi.org/10.1038/gim.2015.30>
- Seki, T., Hayashi, N., Nishimoto, T., 1996. RCC1 in the Ran pathway. *J Biochem* 120, 207–214. <https://doi.org/10.1093/oxfordjournals.jbchem.a021400>
- Shamseldin, H.E., Makhseed, N., Ibrahim, N., Al-Sheddi, T., Alobeid, E., Abdulwahab, F., Alkuraya, F.S., 2019. NUP214 deficiency causes severe encephalopathy and microcephaly in humans. *Hum Genet* 138, 221–229. <https://doi.org/10.1007/s00439-019-01979-w>
- Shen, Q., Wang, Y.E., Truong, M., Mahadevan, K., Wu, J.J., Zhang, H., Li, J., Smith, H.W., Smibert, C.A., Palazzo, A.F., 2021. RanBP2/Nup358 enhances miRNA activity by sumoylating Argonautes. *PLoS Genet* 17, e1009378. <https://doi.org/10.1371/journal.pgen.1009378>
- Shibata, A., Kasai, M., Hoshino, A., Tanaka, T., Mizuguchi, M., 2021. RANBP2 mutation causing autosomal dominant acute necrotizing encephalopathy attenuates its interaction with COX11. *Neurosci Lett* 763, 136173. <https://doi.org/10.1016/j.neulet.2021.136173>
- Shinohara, M., Saitoh, M., Takanashi, J., Yamanouchi, H., Kubota, M., Goto, T., Kikuchi, M., Shiihara, T., Yamanaka, G., Mizuguchi, M., 2011. Carnitine palmitoyl transferase II polymorphism is associated with multiple syndromes of acute encephalopathy with various infectious diseases. *Brain Dev* 33, 512–517. <https://doi.org/10.1016/j.braindev.2010.09.002>
- Singh, R.R., Sedani, S., Lim, M., Wassmer, E., Absoud, M., 2015. RANBP2 mutation and acute necrotizing encephalopathy: 2 cases and a literature review of the expanding clinico-radiological phenotype. *Eur J Paediatr Neurol* 19, 106–113. <https://doi.org/10.1016/j.ejpn.2014.11.010>
- Skelton, B.W., Hollingshead, M.C., Sledd, A.T., Phillips, C.D., Castillo, M., 2008. Acute necrotizing encephalopathy of childhood: typical findings in an atypical disease. *Pediatr Radiol* 38, 810–813. <https://doi.org/10.1007/s00247-008-0823-z>
- Tachibana, T., Imamoto, N., Seino, H., Nishimoto, T., Yoneda, Y., 1994. Loss of RCC1 leads to suppression of nuclear protein import in living cells. *J Biol Chem* 269, 24542–24545.
- Tenembaum, S., Chitnis, T., Ness, J., Hahn, J.S., 2007. Acute disseminated encephalomyelitis. *Neurology* 68, S23–S36. <https://doi.org/10.1212/01.wnl.0000259404.51352.7f>
- Tran, T.D., Kubota, M., Takeshita, K., Yanagisawa, M., Sakakihara, Y., 2001. Varicella-associated acute necrotizing encephalopathy with a good prognosis. *Brain Dev* 23, 54–57. [https://doi.org/10.1016/s0387-7604\(00\)00199-6](https://doi.org/10.1016/s0387-7604(00)00199-6)
- Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., Haeusler, A., 2019. Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Molecular Medicine* 11, e9423. <https://doi.org/10.15252/emmm.201809423>
- Yokoyama, N., Hayashi, N., Seki, T., Panté, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., Aebi, U., 1995. A giant nucleopore protein that binds Ran/TC4. *Nature* 376, 184–188. <https://doi.org/10.1038/376184a0>
- Zhang, K., Daigle, J.G., Cunningham, K.M., Coyne, A.N., Ruan, K., Grima, J.C., Bowen, K.E., Wadhwa, H., Yang, P., Rigo, F., Taylor, J.P., Gitler, A.D., Rothstein, J.D., Lloyd, T.E., 2018. Stress Granule Assembly Disrupts Nucleocytoplasmic Transport. *Cell* 173, 958–971.e17. <https://doi.org/10.1016/j.cell.2018.03.025>



Zhang, S.-Y., 2020. Herpes simplex virus encephalitis of childhood: inborn errors of central nervous system cell-intrinsic immunity. *Hum Genet* 139, 911–918. <https://doi.org/10.1007/s00439-020-02127-5>

## Chapter 7: Discussion

The overall objective of this work was to identify and characterise novel genetic variants in patients with rare disease. By reanalysis of existing clinical genomic data, and by undertaking new genomic sequencing, I identified molecular diagnoses in patients with rare diseases and provided novel genetic and molecular associations relevant to future analyses.

Undertaking WES in a patient with suspected urofacial syndrome provided an interesting insight into the penetrance of urological phenotypes in HADDs. Analysis of WGS in foetuses with URSMS did not yield any variants in previously ascertained genes but did provide candidates for further investigation. Reanalysis of WGS data from patients with variable OHS allowed to identification of a deep intronic variant with a leaky effect on *ATP7A* splicing which likely explains the phenotypic variability of the affected family members. Analysis of WES based on autozygosity mapping and reanalysis of 100kGP data established a novel association between biallelic variants in *RCC1* and autosomal recessive ANE, which may provide insights into Ran-related neuropathies and Guillain-Barré syndrome.

These cases demonstrate both the challenges and opportunities presented during the consideration of missing heritability of rare diseases. Ongoing challenges in variant identification occur at multiple levels. First, if a patient with a suspected genetic condition is identified, eligibility for and participation in genome-wide screening is the first barrier to identifying rare disease-associated genetic variants. Secondly, identifying and adequately describing complex phenotypes can prevent or impede adequate genomic analysis. Thirdly, many genetic disease mechanisms such as non-coding or structural variants, or oligogenic disease models are not widely considered in current bioinformatic analysis pipelines. Finally, characterisation of candidate variants with unknown significance also imposes a limitation in resolving cases of rare disease. This discussion outlines these challenges in relation to the results presented and describes approaches and advancements which will improve identification of novel variants in the future.

## 7.1 Eligibility and participation in genetic screening programmes

The initial barrier to identifying rare variants is access of patients with rare diseases to genetic screening, either as part of a diagnostic pipeline or in a non-diagnostic research-centred programme. For rare diseases with uncertain aetiologies WES and WGS are useful approaches. However, performing NGS analysis often requires the affected individual (and perhaps their parents for trio analysis) to be recruited to a programme offering genomic sequencing. One of the first NGS studies specific to rare diseases with the DDD study in the UK, which recruited affected children from 2011 to 2015 primarily with developmental disorders but also genetic disorders of significant impact where there was no known molecular basis (Firth et al., 2011). This established the feasibility for the integration of large-scale genetic testing into clinical care (Mayor, 2014). Since 2019, implementation of a genomic medicine service in the UK has aimed to provide WGS as part of clinical care, building upon the foundation laid by the 100kGP (Griffin et al., 2017). Two components of the work presented in this thesis relied on genetic analysis conducted entirely on a research basis.

Diagnostic WES or WGS in the UK would have permitted the identification of the *de novo* *EBF3* variant in the individual investigated in chapter 3, however, this patient was from Malaysia, where exome sequencing was unavailable as part of the local diagnostic pipeline. In Malaysia there is no national-scale genomic sequencing programme comparable to 100kGP. However, work is being undertaken to establish a precision medicine initiative (Jamal, 2021). Although the suspicion of UFS was unfounded, a diagnosis of HADDs in this patient would not have been possible without conducting WES as part of a research effort to understand the missing heritability of UFS (Beaman et al., 2022). This raises the value of cross-institution initiatives which permit and promote the sharing of research resources and pipelines to improve definition of molecular mechanisms in rare disease.

The families enrolled in the URSMS study reported on in chapter 4 would not be eligible for diagnostic genetic screening. There are two reasons for this, firstly there was no history of foetal malformations among the families described. Secondly, the lack of monogenic association with URSMS means that genetic testing is not likely to be informative for the

family concerned (Wojcik et al., 2020). The participation of such families in research however is beneficial to identifying mechanisms which may expand the use of prenatal screening (Zhang et al., 2019). Recruitment of families with complex phenotypes is also necessary to probe further into molecular mechanisms underlying complex phenotypes, including structural variation, and complex forms of inheritance such as oligogenic transmission or epigenetic changes.

These examples demonstrate the importance of sharing genetic and phenotype data within research networks. Ongoing projects aim to strengthen and standardise existing research relationships to enable wider engagement of teams with different specialities and improve the connection between research and clinical pipelines to make access better for patients with rare diseases. For example, the benefits of data sharing in projects such as Solve-RD demonstrate the value of improved standardisation of heterogeneous data from different sequencing platforms across data from patients in 15 European countries (Zurek et al., 2021). By combining multiomic approaches with massive data reanalysis, the intended aim of creating a resource similar to gnomAD is to improve variant interpretation within the rare disease community, including clinicians, scientists, and rare disease patient organisations (Zurek et al., 2021). However, the criticism that data from non-Caucasian populations is under-represented in gnomAD can be echoed within the European-centred Solve-RD resource. When considering polygenic or oligogenic variants in complex diseases in individuals from non-European populations, such as our cohort of south Asian foetuses with URSMS, biases in data may hinder the ability to detect and interpret variants that may be rare in European populations, but relatively more common in south Asian populations which are characterised by a high degree of endogamy (Nakatsuka et al., 2017; Pham et al., 2022). Encouraging more diverse participation in genomic screening programmes would help address these challenges and will ensure that power to improve detection and annotation of data is maintained for rare disease patients from non-European populations (Hindorff et al., 2018).

## 7.2 Identifying rare disease aetiology in patients with complex phenotypes

Detection and description of rare phenotypes identified in patients with rare diseases is an essential component in the identification of novel disease-associated genetic variants.

Traditionally, genomic disease research projects have focused on recruiting children with rare diseases because early onset of rare symptoms is suggestive of a congenital condition with a genetic basis (Firth et al., 2011; Sawyer et al., 2016). Complexity in identification and description of rare phenotypes with a likely genetic basis can arise from phenotypes that are progressive in character.

Among the individuals affected by OHS in chapter 5, phenotype onset began in late childhood, with occipital horns typical of OHS only identified among the younger affected individuals in late adolescence and early adulthood. The lack of occipital horns in these individuals led to a preliminary diagnosis of Shwachman-Diamond syndrome (SDS, MIM 260400). Overlapping phenotypes included short stature, bone formation and growth abnormalities, narrow thorax, pancreatic insufficiency, and severe diarrhoea (Nelson and Myers, 1993). Although this phenotypic complexity did permit 100kGP recruitment for WGS, a hypothesis of intronic *ATP7A* variant involvement in disease was only formed following identification of occipital horns in one of the affected family members and confirmed by detection of low serum copper and caeruloplasmin. Alternative hypotheses considered could have related to a blended phenotype, where the presence of two monogenic diseases causes phenotypic heterogeneity (Pérez-Torras et al., 2019). For example, in siblings affected by neurofibromatosis type 1 (*NF1*, MIM 613113) with discordant phenotypes, one sibling was affected by hypotonia and ataxic gait which related to identification of a *de novo* variant in *CASK* (MIM 300172) (Murakami et al., 2019). This demonstrates a challenge of the phenotype-driven hypothesis approach used in this investigation.

Similarly, in chapter 3 phenotyping of the patient with suspected UFS also led to phenotype-biased hypothesis of biallelic variants in *HPSE2* or *LRIG2*. The presence of a hypotonia, developmental delay, short stature, and microcephaly were inconsistent with classical phenotype of UFS which is defined as a dyssynergic bladder and characteristic facial grimace while smiling (Newman and Woolf, 2013; Ochoa, 2004). Identification of a

heterozygous missense variant in *EBF3* which had previously been identified in a patient with HADDs (Tanaka et al., 2017) provided an insight into penetrance of urological phenotypes in HADDs. Our sequencing analysis of individuals with familial vesicoureteral reflux did not detect further pathogenic variants in *EBF3*. This suggests that it is unlikely that *EBF3* explains the missing heritability of UFS.

A phenotype-driven approach was also employed in our analysis of fetuses with URSMS in chapter 4, which occurred as part of a research effort outside of diagnostic pathways. Phenotyping of 12 fetuses with perinatal lethal malformations identified significant phenotypic heterogeneity of URSMS-associated malformations including limb, cardiac, vertebral, central nervous system, and gastroesophageal anomalies. WGS analysis in eight cases did not identify SNVs in genes previously associated with urorectal anomalies. Variants identified in *HOXD9* and *TMEM132A* are considered candidates for *CDX2*-related signalling defects involved in cloacal patterning and development (Hsu et al., 2018), based on findings from rodent knockout (Li et al., 2022) and pharmacological induction (Hong et al., 2021) respectively. Moreover, variants identified in *SLIT2* and *NALCN* relate to URSMS-associated phenotypes identified in each fetus. These variants do not explain the entire phenotype, suggesting that our hypothesis of a monogenic aetiology of URSMS is somewhat weakened. The heterogeneity of the phenotypes identified among this cohort suggests an oligogenic disease model could be considered, where multiple rare variants with variable effect sizes contribute to the different and/or multiple aspects phenotype (Zaghloul and Katsanis, 2010).

While three of the projects in this work featured a phenotype driven approach, the severity and rarity of the ANE phenotype described in chapter 6 was advantageous to that investigation. The novelty of the phenotype, plus history of disease within a consanguineous family permitted undertaking of a family association approach. Autozygosity mapping did not require consideration of the patient phenotype until a locus was identified. Although *RCC1* had not previously been associated with monogenic disease, the clear mechanistic link to *RANBP2* (Levine et al., 2020; Yokoyama et al., 1995) permitted further functional evaluation to be carried out. Our analysis and presentation of the gene-disease association between *RCC1* and autosomal recessive acute necrotising encephalopathy permitted the

identification of subsequent probands with acute onset neurological symptoms from the 100kGP and will facilitate identification of new patients in the future.

The cases described in this thesis demonstrate some of the pitfalls inherent to phenotype-driven approaches to explain the missing heritability of rare disease. Distillation of the complex heterogeneous phenotypes of fetuses with URSMS facilitated a hypothesis of monogenic aetiology, which was not supported in our findings. Genetic investigations can be biased by prioritising certain aspects of an individual's phenotype, demonstrated by the identification of HADDs in a patient with suspected UFS. However, the genetic identification of a deep intronic splicing variant in *ATP7A* was enabled by dialogue with clinicians to construct a clearer picture based on detailed phenotyping and other molecular testing. These findings therefore advocate for nuance when conducting phenotype-driven analyses, as particular aspects of an individual's phenotype can be useful in informing direction of genomic analyses but must not be permitted to bias genetic investigation.

### **7.3 Detection of novel variants associated with rare disease**

Ability to confidently detect pathogenic variants has improved significantly with the advent of NGS. By integrating population-level genomic data, and the implementation of predictive algorithms to assess nucleotide conservation, protein physicochemical changes, and splicing predictions, the ability to identify rare and novel variants in rare disease patients has grown at an astonishing rate (Chen et al., 2019; Ganakammal and Alexov, 2019; Pabinger et al., 2014). However, as more variants are identified in coding regions, in genes with defined functions, and in patients with monogenic disease aetiologies, it is becoming more challenging to identify novel causes of disease. During the course of the projects described in this thesis, there have been several novel and emerging technological and analytical advances which have enabled new approaches which were not possible to consider during the planning stage of this project. These represent an exciting development in the field of genomic medicine and have a huge potential to contribute to the improvement in detection of novel variants for patients with rare disease.

### 7.3.1 Splicing

In chapter 5, by expanding the search for a molecular aetiology of OHS in a family with variable phenotypes to include non-coding variants in *ATP7A*, we detected a novel deep intronic variant strongly predicted to affect splicing of *ATP7A* mRNA. This family was originally referred for genetic testing over 30 years ago, and despite *ATP7A* variants being well-characterised in Menkes disease (Mhaske et al., 2020), the intronic variant was not identifiable through the 100kGP tiering pipeline (100,000 Genomes Project Pilot Investigators et al., 2021). The tiering approach does not consider non-coding variants more than 2 base-pairs away from canonical splice sites (“Rare disease tiering - Genomics England Research Environment User Guide”, 2019). Although the variant identified is novel, the identification of *ATP7A* non-coding variants is not unprecedented in OHS (Mhaske et al., 2020; Yasmeen et al., 2014). Additionally, other instances where deep intronic splice variants have been detected several times include *ATP7B* in Wilsons disease (Woimant et al., 2020) (MIM: 606882), *CFTR* in cystic fibrosis (Sobczyńska-Tomaszewska et al., 2013) (MIM: 602421), and *COL4A5* in X-linked Alport syndrome (Horinouchi et al., 2019) (MIM: 301050). Considering the precedence of deep intronic variants in previous rare disease investigations, and from our analysis, integration of intronic splicing predictions into 100kGP analysis pipelines should be considered.

Guidelines for intronic variant assessment in X-linked Alport syndrome recommend conducting reverse transcriptional PCR analysis in *COL4A5* without undertaking WGS (Yamamura et al., 2022). This is analogous to the initial approach taken to identify *ATP7A* intronic splicing variants in OHS patients following Sanger sequencing of *ATP7A* cDNA (Møller et al., 2021; Yasmeen et al., 2014). The ability to detect deep intronic variants in candidate genes relatively quickly via WGS would allow more informed targeting of downstream approaches such as minigene assays or reverse transcription analysis. This would permit for example the identification of *ATP7A*-spectrum diseases underlying the presentation of cases with complex phenotypes similar to the connective tissue changes observed in OHS. In the case of OHS or SMAX3, identification of patients is vanishingly rare (Mhaske et al., 2020), and often milder non-Menkes *ATP7A*-related disease is not initially considered, as with the family described in chapter 5. However, integrating intronic splice predictions may have the drawback of overlooking variability in disease severity exhibited in



our cohort and in others (Donsante et al., 2007), which identified that variability in patients with the same variant resulted from differences in *ATP7A* overexpression. Mechanisms such as this demonstrate the uncertainty of predicting disease from splicing changes.

In combination with WGS, methods such as RNA sequencing can be useful for determining the effects of intronic variants on the transcriptome. Progress has been made in disease-specific studies, including in Charcot-Marie-Tooth (CMT) disease where RNA sequencing data have been used to identify variants affecting splicing and variation in gene expression (Pipis et al., 2019). Furthermore, *microRNA-140* was shown to act as a modifier of onset age and severity of CMT (Nam et al., 2018). The use of RNA sequencing data to derive the functional impact of genomic variation on RNA transcription is an important advancement. The use of RIVER (RNA-Informed Variant Effect on Regulation) scoring is a good example of this in practice, allowing determination of allele-specific expression, including effects on splicing (Li et al., 2017). One drawback of the implementation of RNA sequencing is the requirement for collection of RNA which is not part of standard clinical NGS pipelines. Depending on the patient and disease concerned, this option may have inherent limitations, as expression patterns in RNA derived from blood cells will differ from expression patterns in other phenotype-relevant organ systems (Li et al., 2017). Until the capabilities in terms of sample collection, storage and analysis permit this multi-omic approach, flagging deep intronic variants with strong predictions to affect *ATP7A* splicing would be useful when considering candidates to characterise in downstream functional assays.

### 7.3.2 Oligogenic disease

Improving assessment of variants that may contribute to an oligogenic disease model is another aspect which could benefit novel variant identification in rare disease. An oligogenic model of disease was initially proposed after identification of digenic inheritance in individuals with Bardet-Biedl syndrome (Dallali et al., 2021; Fauser et al., 2003). However, availability of NGS data acquisition has not greatly expanded the ability to detect which may contribute to oligogenic disease. For example, the NEPTUNE study concerned with identification of the missing heritability among patients with nephrotic syndrome failed to detect an increased burden of protein-altering variants across 21 nephrotic syndrome

related genes in a cohort of 303 patients (Crawford et al., 2017). Our study of fetuses with URSMS aimed to identify a monogenic form of disease so did not explicitly consider oligogenic transmission during analysis of WGS data. Given the number of genes associated with urorectal malformations, and the complexity of the phenotypes observed in this cohort, applying a pipeline to assess oligogenic inheritance should be considered in future analyses. Recently machine learning algorithms including OligoPVP (Boudelloua et al., 2018), VarCoPP (Papadimitriou et al., 2019), and ORVAL (Renaux et al., 2019) have been devised to address the current methodological challenge in assessing the potential effect of oligogenic variants (Rahit and Tarailo-Graovac, 2020; Setty et al., 2022).

### 7.3.3 Structural variation

One aspect of genomic analysis that has not been considered in the work presented is the impact of structural variation on rare disease. SV identification in complex common diseases such as schizophrenia or intellectual disability was enabled by the availability of microarrays useful in detection of CNVs (Stankiewicz and Lupski, 2010). Identification of different classes of SV from short read data is challenging, as different detection algorithms have different sensitivities in the detection of different SVs, such as CNVs, inversions, or translocations, requiring the implementation of several bioinformatic pipelines (Gu et al., 2021).

Improvements in base calling accuracy will facilitate the implementation of long-read sequencing. This technology allows detection of SVs up to 10 kb and offers improved performance in detecting complex SVs (Gu et al., 2021). The use of Pacific Biosciences Sequel I System demonstrated success in a study investigating the implementation of long read sequencing in five family trios affected by intellectual disability [Pauper 2021]. This study describes the benefits of using long-read sequencing in terms of increased genome coverage, and nearly five times greater sensitivity in the detection of SVs, compared to short-read sequencing using the Illumina NovaSeq 6000 platform [Pauper 2021].

As more long-read data sets become available, alignment of long-read sequence data and detection of complex SVs will be improved, allowing long-read sequencing to become the tool of choice to detect structural variants (Jenko Bizjan et al., 2020). One particular accomplishment facilitated by long-read sequencing has been the telomere-to-telomere

sequencing and mapping of the X-chromosome using the Oxford Nanopore MinION sequencing platform to generate ultra-long reads greater than 100 kb (Miga et al., 2020). This allows improvement in genomic mapping and assessment of structural variants in previously poorly or incompletely mapped regions (Miga et al., 2020).

One additional confounder when assessing rare SVs in a gene-agnostic approach is the difficulty in predicting the functional consequences of rare structural rearrangements. Assessment of rare structural variation in human brain tissue has been possible by integrating SV data with transcriptomic analysis from RNA sequencing data, which permitted the quantification of dosage and regulatory effects associated with SVs affecting regulatory regions of the genome (Han et al., 2020). Assessing consequence is essential when classifying SVs in terms of pathogenicity. In the past year recommendations have been made to include classification of non-coding variants into the existing ACMG framework (Richards et al., 2015), allowing detection of variants that affect regulatory regions including non-coding RNAs, promoters, enhancers, and other cis-regulatory elements (Ellingford et al., 2022). Broadening the spectrum of variants that can be classified is a further factor that can improve the ability to interpret genomic data in a clinically useful way.

#### 7.3.4 Non-coding variants

The emphasis on describing the effect of coding variants has led to great improvements, however as more coding variants - both SNV and SV - are described and their effects are characterised, the limitations in the understanding of non-coding variant impact on disease become more apparent. In order to appreciate the functional consequences of rare non-coding variants greater integration of functional genomic technologies is required to provide strong evidence of genotype-phenotype correlations, and ultimately address the missing heritability of rare disease. Functional genomic approaches including transcriptomic, epigenetic, and DNA-RNA-protein interactions (Zhang and Lupski, 2015) can provide useful evidence of non-coding variant involvement in rare disease.

Efforts to understand the effect of non-coding variants affecting regulatory regions of the genome have received attention in research setting but are not currently employed in

clinical diagnostics. Techniques such as Hi-C to map genome-wide interactions between non-contiguous regions of open transcriptionally active chromatin (McArthur and Capra, 2021; Zhang and Lupski, 2015). In the case of one patient with 46,XY gonadal dysgenesis, Hi-C was used to characterise changes in chromatin structure arising from a duplication at locus Xp21.2 (Meinel et al., 2023). Introduction of a new topologically associated domain (TAD) resulted in enhancer hijacking and ectopic expression of NROB1 (Meinel et al., 2023).

Aside from SVs, complementing genotype data from WGS with transcriptomic and epigenomic data can be powerful in detecting the effect of intronic variants (as discussed in relation to splicing in section 7.3.1), but also allows the possibility of detecting the effect of SNVs within regulatory non-coding functional elements of the genome. Changes in gene expression can arise as a result of disrupted transcription factor binding and dysregulation in epigenetic regulation (Martin-Trujillo et al., 2020). For example, a de novo SNV in a CCCTC-binding factor (CTCF) motif has been shown to result in defective DNA methylation in an individual from a cohort of patients with neurodevelopmental disorders and congenital anomalies (Barbosa et al., 2018). Integration of different functional technologies in a coordinated approach will facilitate increased diagnostic yield in the future. However, currently there is insufficient capacity to undertake and integrate different functional data in a diagnostic setting. Applying these technologies to smaller cohorts of unresolved cases, such as the URSMS cohort in Chapter 4 may be an interesting avenue for future study and demonstrate the utility in integrating different approaches to leverage additional diagnoses.

## 7.4 Characterisation of variants of uncertain significance

The current approach to variant classification rightly relies on evidence of pathogenicity, however in many unsolved cases, the current system also hinders the consideration of many variants (Starita et al., 2017). Variants categorised as VUS comprise over 50 % of rare variants that have been interpreted to date, including most novel missense variants (Cooper, 2015). Ensuring models are developed to correctly assess functional effect of VUS and developing tools that can increase the capacity to assess variant effect remain a challenge to identifying and reporting novel variants and disease mechanisms (Cooper, 2015). Furthermore, understanding the biological function of disease genes is also essential in developing targeted therapies (Stoeger et al., 2018; Wangler et al., 2017).

Using an effective model is an important element of characterising functional effect of variant of unknown significance. In this thesis, *in vitro* methods were used to interrogate effects of *ATP7A* splicing, characterise the stability and protein activity of Rcc1 protein, and probe the changes in Rcc1 localisation in proband-derived fibroblasts. In these investigations, the experimental approach taken was determined principally by sample availability. For example, a limitation of the splicing investigation conducted in OHS proband with *ATP7A* variant in chapter 5 was the lack of availability of a cell type with good expression of *ATP7A*. Blood lymphocyte-derived RNA was used to confirm the splicing effect in a patient system, however previous studies evaluated *ATP7A* expression and splicing in patient-derived dermal fibroblasts (Yasmeen et al., 2014). While our findings from lymphocyte-derived RNA are supported by minigene assay findings, optimal determination of *ATP7A* transcript splicing would be ameliorated by using dermal fibroblasts. Although *ATP7A* is expressed in cultured lymphocytes according to the Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/gene/ATP7A>), dermal fibroblasts are a more relevant cell type when considering the cutis laxa phenotype present in OHS patients. The connective tissue phenotype in OHS results from decreased LOXL crosslinking of collagen fibres as a result of impaired Cu<sup>2+</sup> transport and loading by defective *ATP7A* (Horn and Wittung-Stafshede, 2021). Assessment of transcript proportions in different cell types may also inform the basis for phenotype variability in this family.

Expression of recombinant Rcc1 was selected as the model to conduct initial functional evaluation in chapter 6. Thermal stability and GTPase activity were both considered critical to the pathogenesis of ANE. The identification of *RCC1* from thermosensitive hamster cell line tsBN2 (Seki et al., 1996), plus the identification of thermolabile *CPT2* polymorphisms in acute encephalopathy cases (Shinohara et al., 2011) prompted a hypothesis that *RCC1* missense variants may result in a thermolabile product. Expressing recombinant Rcc1 using a bacterial culture permitted the thermal stability of Rcc1 to be assessed, and measurement of Rcc1 GEF activity. Multiple different GTPase reactions occur within the cellular environment, so expressing a recombinant form of Rcc1 permitted a more reliable quantification of activity *in vitro*. However, this reductionist model did not consider interaction of proteins other than Rcc1 and Ran GTPase (Loose et al., 2022). Inclusion of GTPase activating protein RanGAP, or chromatin to facilitate Rcc1 tethering are examples of how this model may be altered to provide a more relevant output. Enzymatic assays such as GTPase assays are a useful approach when assessing multiple variants. The ease by which subsequent variants can be expressed using the same system allows for multiple variants in the same gene to be evaluated in parallel, without requirement for patient-derived material.

Establishing a suitable model for functionally characterising variants identified in genomes of rare disease patients can be resource and time intensive. Determining biological functions of genes associated with rare diseases therefore remains a bottleneck in the field of rare diseases (Boycott et al., 2020). Forming relationships between clinical teams and research scientists with relevant disease models is one approach offered by the rare disease models and mechanisms (RDMM) network active in Canada, Australia, Europe, and Japan (Boycott et al., 2020). This network has enabled functional characterisation of rare disease genes including characterisation of pyridoxine-dependent epilepsy related gene *ALDH7A1* (MIM 107323) in a zebrafish model more than 60 years after the initial disease description (Pena et al., 2017).

However, establishing connections between clinicians and researchers with relevant model organisms is only part of the solution. Analysis of publications relating to functional genetic studies identified an ongoing research bias towards genes whose function has already been described, resulting in suppressed efforts in poorly described genes (Stoeger et al., 2018).

This supports the notion that efforts to identify novel disease-gene relationships are hindered by lack of biological knowledge relating to the function of many genes (Haynes et al., 2018). Additionally, evaluation of variants using *in silico* tools often gives contradictory assessments of pathogenicity (Grimm et al., 2015). Approaches for mass-screening of gene functions using RNAi or model organism knockouts do provide insight into the function of genes but not necessarily the effect of variants which can result in gain or loss of function (Stoeger et al., 2018).

Systematic screening of variants in specific genes can be useful for identifying variants which are likely to be pathogenic, providing the strongest type of evidence for variant classification (Richards et al., 2015). Technological developments allow high throughput deep mutational scans and multiplexing of assays to establish the effect of variants of unknown significance in coding regions, splice sites, and regulatory regions (Gasperini et al., 2016; Starita et al., 2017). Multiplexing of minigene assays is an example of how massively parallel analysis of synthetic sequences can inform alternative splicing (Rosenberg et al., 2015). Similarly, saturation genome editing allows the measurement of functional splicing consequences at specific genic loci (Findlay et al., 2014). Furthermore, conducting parallelised enzyme reporter assays can facilitate an assessment the effect of variants in enhancer regions on expression of reporter genes (Inoue and Ahituv, 2015). Building on these innovations, recent developments permit generation of user-defined variant libraries, facilitating use of these techniques in genes that are poorly defined (Barbon et al., 2021). When combined systematically within a rare disease network of clinicians and scientists, these approaches pave the way for implementation of multiplexing tools to ameliorate functional characterisation of rare and novel genetic variants identified in individuals with rare disease.

## 7.5 Limitations of approach and study

In this thesis, the case-study approach undertaken permitted a broad evaluation of different disease types and mechanisms. The prioritisation of this breadth was intended as a reflection of the varied and stochastic nature of rare disease genetics. The breadth of this work however did prevent pursuing each individual study in greater depth. A limitation of this overall approach was the inability to pursue all possible analytical pipelines available. In chapter 4, no analysis of SVs was attempted despite availability of WGS data. There are two factors which led to the delay in the completion of this work. Firstly, analysis of genes previously associated with anorectal malformations failed to identify a common monogenic aetiology. Identification of a genomic change in more than one family would be necessary to undertake further analysis. Secondly, the overall strategy for this thesis led to prioritisation of elements which were more likely to yield clinically and biologically relevant data. As a result, laboratory work for the staining of proband fibroblasts in Chapter 6 was prioritised over SV analysis to complement the published work in Chapter 4. Undertaking further functional genomic analyses in the URSMS cohort would be of great interest in analysis of non-coding changes relevant to the phenotype.

Prioritisation of immunofluorescence staining of patient fibroblasts in chapter 6 was enabled by availability of patient material. This is in contrast the work carried out in chapter 5, where culturable cell lines were not accessible due to delays in elective surgery during which fibroblasts were to be collected from the patient. Availability of patient fibroblasts would have improved analysis of ATP7A splicing defects by allowing the evaluation of Atp7a protein expression. Protein expression would confirm if mis-spliced transcripts are stable or are degraded.

In Chapter 6, the identification of additional affected families and the Rcc1V80M and Rcc1R399C variants were made after undertaking the in vitro stability and functional analyses using recombinant Rcc1G43S. As a result, analyses were not extended to include additional variants identified in new patients. This will be a consideration during future analyses. Additionally, the in vitro approach taken in Chapter 6 could have been improved by considering complementary approaches to the GTPase luminescence assay. One example is the radioactive nucleotide exchange assay demonstrated by Chatterjee et al., (2015).



Alternatively, use of fluorescent N-Methylanthraniloyl (MANT)-Guanine nucleotides (Holzer et al., 2021; Kanie and Jackson, 2018) could be used to quantify changes in Ran GTPase rate constants in reactions with different *Rcc1* mutant proteins. This approach would enable quantification changes in reaction kinetics and permit a more specific characterisation of the functionality of *Rcc1* with different missense variants.

## 7.6 Further work

This work has highlighted opportunities for conducting future studies. Our description of autosomal recessive ANE associated with *RCC1* variants requires further evaluation to define the molecular mechanism underlying Ran disruption following fever. Specifically, interrogating the effects of *Rcc1* dysfunction on nucleocytoplasmic transport, which is known to be involved in *RANBP2*- and *NUP214*-related pathologies (Fichtman et al., 2019; Shamseldin et al., 2019). We plan to establish model to assess the effect of *RCC1* variants on nucleocytoplasmic import and export. The use of digitonin to permeabilise the plasma membrane, but not the nuclear envelope, will allow for the delivery of exogenous fluorescently labelled cargoes which can be examined using fluorescence microscopy (Cassany and Gerace, 2009). Export of poly(A)<sup>+</sup> RNA from the nucleus can also be assessed by hybridisation (Fichtman et al., 2019). We also plan to use tsBN2 cell lines to investigate the ability of variants identified in several patients to rescue their innate temperature-sensitive phenotype. This will enable us to screen future variants identified in patients with acute-onset neurological symptoms.

Moreover, undertaking further analysis of WGS data from the fetuses with URSMS would allow for novel tools to be assessed in a cohort with complex phenotypes. Focusing on identifying possible oligogenic candidates using machine learning approaches such as ORVAL (Renaux et al., 2019), or identification of complex structural variants using long-read sequencing and additional functional genomic approaches may reveal variants which can be associated with these complex cases and may provide further molecular insight.

## 7.7 Final Summary

In this thesis, the phenotype-driven approach used to identify variants in individuals with rare diseases has had mixed success. Although this approach was less successful in identifying a monogenic cause of URSMS, we have identified candidates for future studies, and have the opportunity to conduct further genomic analyses to assess the role of non-monogenic disease mechanisms. This work also highlights the current challenges posed by the incomplete functional annotation of many genes and non-coding regions which hinders the ability to assign a pathogenic classification to rare variants identified in patients.

Identification of causal variants in *EBF3*, *ATP7A*, and *RCC1* highlight the value and utility of genomic approaches in the identification of a molecular aetiology in patients with complex phenotypes. Further research triggered by our identification of autosomal recessive *RCC1*-related encephalopathy will provide a novel insight into the molecular cause of Ran-related neuropathies and facilitate screening of *RCC1* variants identified in subsequent individuals.

## References

- 100,000 Genomes Project Pilot Investigators, Smedley, D., Smith, K.R., Martin, A., Thomas, E.A., McDonagh, E.M., Cipriani, V., Ellingford, J.M., Arno, G., Tucci, A., Vandrovcova, J., Chan, G., Williams, H.J., Ratnaïke, T., Wei, W., Stirrups, K., Ibanez, K., Moutsianas, L., Wielscher, M., Need, A., Barnes, M.R., Vestito, L., Buchanan, J., Wordsworth, S., Ashford, S., Rehmström, K., Li, E., Fuller, G., Twiss, P., Spasic-Boskovic, O., Halsall, S., Floto, R.A., Poole, K., Wagner, A., Mehta, S.G., Gurnell, M., Burrows, N., James, R., Penkett, C., Dewhurst, E., Gräf, S., Mapeta, R., Kasanicki, M., Haworth, A., Savage, H., Babcock, M., Reese, M.G., Bale, M., Baple, E., Boustred, C., Brittain, H., de Burca, A., Bleda, M., Devereau, A., Halai, D., Haraldsdottir, E., Hyder, Z., Kasperaviciute, D., Patch, C., Polychronopoulos, D., Matchan, A., Sultana, R., Ryten, M., Tavares, A.L.T., Tregidgo, C., Turnbull, C., Welland, M., Wood, S., Snow, C., Williams, E., Leigh, S., Foulger, R.E., Daugherty, L.C., Niblock, O., Leong, I.U.S., Wright, C.F., Davies, J., Crichton, C., Welch, J., Woods, K., Abulhoul, L., Aurora, P., Bockenhauer, D., Broomfield, A., Cleary, M.A., Lam, T., Dattani, M., Footitt, E., Ganesan, V., Grunewald, S., Compeyrot-Lacassagne, S., Muntoni, F., Pilkington, C., Quinlivan, R., Thapar, N., Wallis, C., Wedderburn, L.R., Worth, A., Bueser, T., Compton, C., Deshpande, C., Fassihi, H., Haque, E., Izatt, L., Josifova, D., Mohammed, S., Robert, L., Rose, S., Ruddy, D., Sarkany, R., Say, G., Shaw, A.C., Wolejko, A., Habib, B., Burns, G., Hunter, S., Grocock, R.J., Humphray, S.J., Robinson, P.N., Haendel, M., Simpson, M.A., Banka, S., Clayton-Smith, J., Douzgou, S., Hall, G., Thomas, H.B., O'Keefe, R.T., Michaelides, M., Moore, A.T., Malka, S., Pontikos, N., Browning, A.C., Straub, V., Gorman, G.S., Horvath, R., Quinton, R., Schaefer, A.M., Yu-Wai-Man, P., Turnbull, D.M., McFarland, R., Taylor, R.W., O'Connor, E., Yip, J., Newland, K., Morris, H.R., Polke, J., Wood, N.W., Campbell, C., Camps, C., Gibson, K., Koelling, N., Lester, T., Németh, A.H., Palles, C., Patel, S., Roy, N.B.A., Sen, A., Taylor, J., Cacheiro, P., Jacobsen, J.O., Seaby, E.G., Davison, V., Chitty, L., Douglas, A., Naresh, K., McMullan, D., Ellard, S., Temple, I.K., Mumford, A.D., Wilson, G., Beales, P., Bitner-Glindzicz, M., Black, G., Bradley, J.R., Brennan, P., Burn, J., Chinnery, P.F., Elliott, P., Flinter, F., Houlden, H., Irving, M., Newman, W., Rahman, S., Sayer, J.A., Taylor, J.C., Webster, A.R., Wilkie, A.O.M., Ouwehand, W.H., Raymond, F.L., Chisholm, J., Hill, S., Bentley, D., Scott, R.H., Fowler, T., Rendon, A., Caulfield, M., 2021. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* 385, 1868–1880. <https://doi.org/10.1056/NEJMoa2035790>
- 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., 2010. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. <https://doi.org/10.1038/nature09534>
- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. <https://doi.org/10.1038/nature11632>
- 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2015. A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>

- Abacan, M., Alsubaie, L., Barlow-Stewart, K., Caanen, B., Cordier, C., Courtney, E., Davoine, E., Edwards, J., Elackatt, N.J., Gardiner, K., Guan, Y., Huang, L.-H., Malmgren, C.I., Kejriwal, S., Kim, H.J., Lambert, D., Lantigua-Cruz, P.A., Lee, J.M.H., Lodahl, M., Lunde, Å., Macaulay, S., Macciocca, I., Margarit, S., Middleton, A., Moldovan, R., Ngeow, J., Obregon-Tito, A.J., Ormond, K.E., Paneque, M., Powell, K., Sanghavi, K., Scotcher, D., Scott, J., Juhé, C.S., Shkedi-Rafid, S., Wessels, T.-M., Yoon, S.-Y., Wicklund, C., 2019. The Global State of the Genetic Counseling Profession. *Eur J Hum Genet* 27, 183–197. <https://doi.org/10.1038/s41431-018-0252-x>
- Abyzov, A., Urban, A.E., Snyder, M., Gerstein, M., 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21, 974–984. <https://doi.org/10.1101/gr.114876.110>
- Adusumalli, S., Mohd Omar, M.F., Soong, R., Benoukraf, T., 2015. Methodological aspects of whole-genome bisulfite sequencing analysis. *Brief Bioinform* 16, 369–379. <https://doi.org/10.1093/bib/bbu016>
- Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7, Unit7.20. <https://doi.org/10.1002/0471142905.hg0720s76>
- Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, YoSon, Parsana, P., Segrè, A.V., Strober, B.J., Zappala, Z., Cummings, B.B., Gelfand, E.T., Hadley, K., Huang, K.H., Lek, M., Li, Xiao, Nedzel, J.L., Nguyen, D.Y., Noble, M.S., Sullivan, T.J., Tukiainen, T., MacArthur, D.G., Getz, G., Addington, A., Guan, P., Koester, S., Little, A.R., Lockhart, N.C., Moore, H.M., Rao, A., Struewing, J.P., Volpi, S., Brigham, L.E., Hasz, R., Hunter, M., Johns, C., Johnson, M., Kopen, G., Leinweber, W.F., Lonsdale, J.T., McDonald, A., Mestichelli, B., Myer, K., Roe, B., Salvatore, M., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Bridge, J., Foster, B.A., Gillard, B.M., Karasik, E., Kumar, R., Miklos, M., Moser, M.T., Jewell, S.D., Montroy, R.G., Rohrer, D.C., Valley, D., Mash, D.C., Davis, D.A., Sobin, L., Barcus, M.E., Branton, P.A., Abell, N.S., Balliu, B., Delaneau, O., Frésard, L., Gamazon, E.R., Garrido-Martín, D., Gewirtz, A.D.H., Gliner, G., Gloudemans, M.J., Han, B., He, A.Z., Hormozdiari, F., Li, Xin, Liu, B., Kang, E.Y., McDowell, I.C., Ongen, H., Palowitch, J.J., Peterson, C.B., Quon, G., Ripke, S., Saha, A., Shabalin, A.A., Shimko, T.C., Sul, J.H., Teran, N.A., Tsang, E.K., Zhang, H., Zhou, Y.-H., Bustamante, C.D., Cox, N.J., Guigó, R., Kellis, M., McCarthy, M.I., Conrad, D.F., Eskin, E., Li, G., Nobel, A.B., Sabatti, C., Stranger, B.E., Wen, X., Wright, F.A., Ardlie, K.G., Dermitzakis, E.T., Lappalainen, T., Aguet, F., Ardlie, K.G., Cummings, B.B., Gelfand, E.T., Getz, G., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., Karczewski, K.J., Lek, M., Li, Xiao, MacArthur, D.G., Nedzel, J.L., Nguyen, D.T., Noble, M.S., Segrè, A.V., Trowbridge, C.A., Tukiainen, T., Abell, N.S., Balliu, B., Barshir, R., Basha, O., Battle, A., Bogu, G.K., Brown, A., Brown, C.D., Castel, S.E., Chen, L.S., Chiang, C., Conrad, D.F., Cox, N.J., Damani, F.N., Davis, J.R., Delaneau, O., Dermitzakis, E.T., Engelhardt, B.E., Eskin, E., Ferreira, P.G., Frésard, L., Gamazon, E.R., Garrido-Martín, D., Gewirtz, A.D.H., Gliner, G., Gloudemans, M.J., Guigo, R., Hall, I.M., Han, B., He, Y., Hormozdiari, F., Howald, C., Kyung Im, H., Jo, B., Yong Kang, E., Kim, Y., Kim-Hellmuth, S., Lappalainen, T., Li, G., Li, Xin, Liu, B., Mangul, S., McCarthy, M.I., McDowell, I.C., Mohammadi, P., Monlong, J., Montgomery, S.B., Muñoz-Aguirre, M., Ndungu, A.W., Nicolae, D.L., Nobel, A.B., Oliva, M., Ongen, H., Palowitch, J.J., Panousis, N., Papasaikas, P., Park, YoSon,

- Parsana, P., Payne, A.J., Peterson, C.B., Quan, J., Reverter, F., Sabatti, C., Saha, A., Sammeth, M., Scott, A.J., Shabalín, A.A., Sodaei, R., Stephens, M., Stranger, B.E., Strober, B.J., Sul, J.H., Tsang, E.K., Urbut, S., van de Bunt, M., Wang, G., Wen, X., Wright, F.A., Xi, H.S., Yeger-Lotem, E., Zappala, Z., Zaugg, J.B., Zhou, Y.-H., Akey, J.M., Bates, D., Chan, J., Chen, L.S., Claussnitzer, M., Demanelis, K., Diegel, M., Doherty, J.A., Feinberg, A.P., Fernando, M.S., Halow, J., Hansen, K.D., Haugen, E., Hickey, P.F., Hou, L., Jasmine, F., Jian, R., Jiang, L., Johnson, A., Kaul, R., Kellis, M., Kibriya, M.G., Lee, K., Billy Li, J., Li, Q., Li, Xiao, Lin, J., Lin, S., Linder, S., Linke, C., Liu, Y., Maurano, M.T., Moliníe, B., Montgomery, S.B., Nelson, J., Neri, F.J., Oliva, M., Park, Yongjin, Pierce, B.L., Rinaldi, N.J., Rizzardi, L.F., Sandstrom, R., Skol, A., Smith, K.S., Snyder, M.P., Stamatoyannopoulos, J., Stranger, B.E., Tang, H., Tsang, E.K., Wang, L., Wang, M., Van Wittenberghe, N., Wu, F., Zhang, R., Nierras, C.R., Branton, P.A., Carithers, L.J., Guan, P., Moore, H.M., Rao, A., Vaught, J.B., Gould, S.E., Lockart, N.C., Martin, C., Struwing, J.P., Volpi, S., Addington, A.M., Koester, S.E., Little, A.R., GTEx Consortium, Lead analysts: Laboratory, D.A. & C.C. (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group; Laboratory, D.A. & C.C. (LDACC)—Analysis W.G., Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, 2017. Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. <https://doi.org/10.1038/nature24277>
- Albader, N., Zou, M., BinEssa, H.A., Abdi, S., Al-Enezi, A.F., Meyer, B.F., Alzahrani, A.S., Shi, Y., 2022. Insights of Noncanonical Splice-site Variants on RNA Splicing in Patients With Congenital Hypothyroidism. *The Journal of Clinical Endocrinology & Metabolism* 107, e1263–e1276. <https://doi.org/10.1210/clinem/dgab737>
- Alkan, C., Coe, B.P., Eichler, E.E., 2011. Genome structural variation discovery and genotyping. *Nat Rev Genet* 12, 363–376. <https://doi.org/10.1038/nrg2958>
- Álvaro-Sánchez, S., Abreu-Rodríguez, I., Abulí, A., Serra-Juhe, C., Garrido-Navas, M. del C., 2021. Current Status of Genetic Counselling for Rare Diseases in Spain. *Diagnostics (Basel)* 11, 2320. <https://doi.org/10.3390/diagnostics11122320>
- Amberger, J., Bocchini, C.A., Scott, A.F., Hamosh, A., 2009. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* 37, D793–796. <https://doi.org/10.1093/nar/gkn665>
- Angelis, A., Tordrup, D., Kanavos, P., 2015. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. *Health Policy* 119, 964–979. <https://doi.org/10.1016/j.healthpol.2014.12.016>
- Anna, A., Monika, G., 2018. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genetics* 59, 253–268. <https://doi.org/10.1007/s13353-018-0444-7>
- Antonarakis, S.E., Kazazian, H.H., Tuddenham, E.G., 1995. Molecular etiology of factor VIII deficiency in hemophilia A. *Hum Mutat* 5, 1–22. <https://doi.org/10.1002/humu.1380050102>
- Asim, A., Agarwal, S., Panigrahi, I., Saiyed, N., Bakshi, S., 2017. MTHFR promoter hypermethylation may lead to congenital heart defects in Down syndrome. *Intractable Rare Dis Res* 6, 295–298. <https://doi.org/10.5582/irdr.2017.01068>
- Austin, C.P., Cutillo, C.M., Lau, L.P.L., Jonker, A.H., Rath, A., Julkowska, D., Thomson, D., Terry, S.F., de Montleau, B., Ardigò, D., Hivert, V., Boycott, K.M., Baynam, G.,

Kaufmann, P., Taruscio, D., Lochmüller, H., Suematsu, M., Incerti, C., Draghia-Akli, R., Norstedt, I., Wang, L., Dawkins, H.J.S., International Rare Diseases Research Consortium (IRDiRC), 2018. Future of Rare Diseases Research 2017-2027: An IRDiRC Perspective. Clin Transl Sci 11, 21–27. <https://doi.org/10.1111/cts.12500>

Bajaj, A., Mathew, S., Vellarikkal, S.K., Sivadas, A., Bhoyar, R.C., Joshi, K., Jain, A., Mishra, A., Verma, A., Jayarajan, R., Nalini, A., Kumar, A.R., Seeralar, A.T.A., Gupta, Aayush, Srivastava, A.K., Joshi, A., Sinha, A., Jandial, A., Khan, A., Sonakar, A.K., Chandy, A., Sharma, A., Roy, A., Rawat, A., Biswas, A., Vanlalawma, A., Chaudhary, A., Chopra, A., Panday, A., Sabharwal, A., Mitra, A., Narang, A., Rajab, A., Kumar, A., Gurjar, A.S., Ranawat, A.S., Anu, R.I., Tiwary, A.K., Anuradha, Kalanad, A., Mathur, A., Lakshman, A., Batra, A., Bagga, A., Aggarwal, A., Gupta, Ashok, Rastogi, A., Aslam, P.K., Astha, V., Nair, A., Athulya, E.P., Chatterjee, A., Jindal, A., Kashyap, A.K., Priyadarshini, B., Thapa, B.R., Bhargava, B., Sharma, Balram, Jolly, B., Uppilli, B.R., Balachander, B., Shankar, B., Kar, B., Binukumar, B.K., Lalchandama, C., Datar, C., Sachidanandan, C., Master, D.C., Khera, D., Chowdhury, D., Danda, D., Kumar, D., Pandhi, D., Siddharthan, D., Sharma, D., Pachat, D., Sharma, Brijesh, Vegulada, D.R., Naidu, G.S.R.S.N.K., Padma, G., Vishnu Priya, G., Sharma, G., Gauthamen, R., Govindaraj, G., Varghese, G.M., Gireesh, S., Unnikrishnan, G.K., Hafiz, S.A., Hazeena, K.R., Dhiman, H., Singh, H., Sarkar, H., Ahmed, I., Menon, J., Goraya, J., Mathew, J., Thottath, J., Sahu, J.K., Oswal, J., Menachery, J., Hariprakash, J.M., Bhargava, K., Talwar, K.K., Cherian, K.M., Aravindan, K.P., Pramila, K., Saroja, K., Shantaraman, K., Pandhare, K., Mandapati, K.K., Kiran, P., Rakesh, K., Shah, Krati, Krishnan, C., Shah, Kriti, Singh, K., Anand, K., Pachuau, L., Chandrashekar, L., Rajasekhar, L., Mishra, L., Padma, M.V., Kabra, M., Chowdhary, M.R., Seth, M., Rai, M., Kumar, Manish, Parakh, M., Goyal, M., Gurjar, M., Sahay, M., Rophina, M., Mukerji, M., Ali, M., Faruq, M., Kariparth, M.N., Divakar, M.K., Jayakrishnan, M.P., Kumar, Mukesh, Poojary, M., Prabhu, M.A., Kumar, N.S., Rais, N., Bhaskaranand, N., Bagri, N.K., Sankhyani, N., Awasthy, N., Gupta, Neeraj, Parakh, N., Gupta, Neeraj, Bhari, N., Kushwaha, N., Sharma, N., Virmani, N., Kundu, N., Plakkal, N., Tyagi, N., Radhakrishnan, N., Naik, N., Rai, N., Mondal, N., Bhargava, N., Hari, P., Sehgal, P., Kumar, Piyush, Chauhan, P., Mailankody, P., Sharma, P., Parakh, P., Nair, P.A., Chakraborty, P., Shirol, P.K., Singh, P., Gangadhar, P., Kumar, Prawin, Chandra, P., Krishnan, R., Srilakshmi, R., Lakshmi, R.S., Anantharaman, R., Mahadevan, R., Mahajan, R., Shanmugam, R., Sharma, R., Rajendran, V.R., Dhamija, R.K., Ramanan, R.P., Kumar, R., Rajneesh, A.R., Juneja, R., Aggarwal, R., Sahay, R., Ramakrishnan, S., Narayanan, R., Shukla, R., Koshy, R., Kumari, R., Chaudhary, R., Jain, R., Arakkal, R., Rajan, R., Ravi, R., Baruah, S., Sitaraman, S., Chandra, S.R., Chenkual, S., Sailaja, V., Ambawat, S., Panda, Samhita, Zahra, S., Kumar, Sanchit, Arora, S., Mathur, S., Seth, S., Sandhya, P., Goswami, S., Paul, S., Pandey, S., Kalyanaraman, S., Patnaik, S., Wadhwa, S., Venu, S., Nanda, S., Panda, Saumya, Chopra, S., Singh, S., Savinitha, P., Kapoor, S., Sivadasan, S., Sethuraman, G., Khan, S.P., Shaji, C.V., Gurusamy, S., Gulati, S., Gandhi, S., Ramalingam, S., Nath, S., Kumar, Somesh, Sathian, S., Lakhani, S., Nair, Soumya S., Sundaram, S., Ghosh, S., Raju, S.B., Valappil, S., Nair, S., Puttaiah, S.K., Nair, Sruthi S., Geevarghese, S.K., Mohanty, S., Khandpur, S., Jain, Suman, Sumeet, Sharma, Sumit, Trehan, S., Sharma, Suvasini, Jain, S., Jain, Swetha, Badam, T.K., Umamaheswari, S., Gaharwar, U., Shamim, U., Rao, V.R., Krishna, V., Jain, V., Suroliya, V., Vyas, V., Vedartham, V., Venketesh, S., Senthivel, V., Bhavi, V., Jadhav, V., Gera, V., Dixit, V.,

- Gupta, Vishal, Agarwal, V., Vishnu, V.Y., Gupta, Vishu, Vysakha, K.V., Sharma, Y.K., Brahmachari, S.K., Scaria, V., Sivasubbu, S., Sivasubbu, S., Scaria, V., The GUARDIAN Consortium, 2019. Genomics of rare genetic diseases—experiences from India. *Human Genomics* 13, 52. <https://doi.org/10.1186/s40246-019-0215-5>
- Barbon, L., Offord, V., Radford, E.J., Butler, A.P., Gerety, S.S., Adams, D.J., Tan, H.K., Waters, A.J., 2021. Variant Library Annotation Tool (VaLiAnT): an oligonucleotide library design and annotation tool for Saturation Genome Editing and other Deep Mutational Scanning experiments. *Bioinformatics* 38, 892–899. <https://doi.org/10.1093/bioinformatics/btab776>
- Barbosa, M., Joshi, R.S., Garg, P., Martin-Trujillo, A., Patel, N., Jadhav, B., Watson, C.T., Gibson, W., Chetnik, K., Tessereau, C., Mei, H., De Rubeis, S., Reichert, J., Lopes, F., Vissers, L.E.L.M., Kleefstra, T., Grice, D.E., Edelmann, L., Soares, G., Maciel, P., Brunner, H.G., Buxbaum, J.D., Gelb, B.D., Sharp, A.J., 2018. Identification of rare de novo epigenetic variations in congenital disorders. *Nat Commun* 9, 2064. <https://doi.org/10.1038/s41467-018-04540-x>
- Bateson, W., Mendel, G., 2013. *Mendel's Principles of Heredity*. Courier Corporation.
- Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., De Zaeytijd, J., Khan, M., Sadler, F., Balikova, I., Van Cauwenbergh, C., Rosseel, T., Bauwens, J., De Leeneer, K., De Jaegere, S., Van Laethem, T., De Vries, M., Carss, K., Arno, G., Fakin, A., Webster, A.R., de Ravel de l'Argentièrre, T.J.L., Sznajder, Y., Vuylsteke, M., Kohl, S., Wissinger, B., Cherry, T., Collin, R.W.J., Cremers, F.P.M., Leroy, B.P., De Baere, E., 2019. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel non-coding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet Med* 21, 1761–1771. <https://doi.org/10.1038/s41436-018-0420-y>
- Beaman, G.M., Woolf, A.S., Lopes, F.M., Guo, S.A., Harkness, J.R., Cervellione, R.M., Keene, D., Mushtaq, I., Clatworthy, M.R., Newman, W.G., 2022. Narrowing the chromosome 22q11.2 locus duplicated in bladder exstrophy-epispadias complex. *J Pediatr Urol* 18, 362.e1–362.e8. <https://doi.org/10.1016/j.jpuro.2022.04.006>
- Bernatsky, S., Duffy, C., Malleon, P., Feldman, D.E., Pierre, Y.S., Clarke, A.E., 2007. Economic impact of juvenile idiopathic arthritis. *Arthritis Care & Research* 57, 44–48. <https://doi.org/10.1002/art.22463>
- Boudelloua, I., Kulmanov, M., Schofield, P.N., Gkoutos, G.V., Hoehndorf, R., 2018. OligoPVP: Phenotype-driven analysis of individual genomic information to prioritize oligogenic disease variants. *Sci Rep* 8, 14681. <https://doi.org/10.1038/s41598-018-32876-3>
- Boycott, K.M., Campeau, P.M., Howley, H.E., Pavlidis, P., Rogic, S., Oriel, C., Berman, J.N., Hamilton, R.M., Hicks, G.G., Lipshitz, H.D., Masson, J.-Y., Shoubridge, E.A., Junker, A., Leroux, M.R., McMaster, C.R., Michaud, J.L., Turvey, S.E., Dymont, D., Innes, A.M., van Karnebeek, C.D., Lehman, A., Cohn, R.D., MacDonald, I.M., Rachubinski, R.A., Frosk, P., Vandersteen, A., Wozniak, R.W., Pena, I.A., Wen, X.-Y., Lacaze-Masmonteil, T., Rankin, C., Hieter, P., 2020. The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. *Am J Hum Genet* 106, 143–152. <https://doi.org/10.1016/j.ajhg.2020.01.009>
- Brittain, H.K., Scott, R., Thomas, E., 2017. The rise of the genome and personalised medicine. *Clin Med (Lond)* 17, 545–551. <https://doi.org/10.7861/clinmedicine.17-6-545>

- Butkiewicz, M., Bush, W.S., 2016. In Silico Functional Annotation of Genomic Variation. *Curr Protoc Hum Genet* 88, 6.15.1-6.15.17. <https://doi.org/10.1002/0471142905.hg0615s88>
- Carr, I.M., Bhaskar, S., O'Sullivan, J., Aldahmesh, M.A., Shamseldin, H.E., Markham, A.F., Bonthron, D.T., Black, G., Alkuraya, F.S., 2013. Autozygosity mapping with exome sequence data. *Hum Mutat* 34, 50–56. <https://doi.org/10.1002/humu.22220>
- Carr, I.M., Camm, N., Taylor, G.R., Charlton, R., Ellard, S., Sheridan, E.G., Markham, A.F., Bonthron, D.T., 2011. GeneScreen: a program for high-throughput mutation detection in DNA sequence electropherograms. *J Med Genet* 48, 123–130. <https://doi.org/10.1136/jmg.2010.082081>
- Cassany, A., Gerace, L., 2009. Reconstitution of nuclear import in permeabilized cells. *Methods Mol Biol* 464, 181–205. [https://doi.org/10.1007/978-1-60327-461-6\\_11](https://doi.org/10.1007/978-1-60327-461-6_11)
- Chen, H.Y., Welby, E., Li, T., Swaroop, A., 2019. Retinal disease in ciliopathies: Recent advances with a focus on stem cell-based therapies. *Transl Sci Rare Dis* 4, 97–115. <https://doi.org/10.3233/TRD-190038>
- Chen, J., Li, X., Zhong, H., Meng, Y., Du, H., 2019. Systematic comparison of germline variant calling pipelines cross multiple next-generation sequencers. *Sci Rep* 9, 9345. <https://doi.org/10.1038/s41598-019-45835-3>
- Conrad, D.F., Keebler, J.E.M., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., Zilversmit, M., Cartwright, R., Rouleau, G.A., Daly, M., Stone, E.A., Hurles, M.E., Awadalla, P., 1000 Genomes Project, 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43, 712–714. <https://doi.org/10.1038/ng.862>
- Cooper, G.M., 2015. Parlez-vous VUS? *Genome Res* 25, 1423–1426. <https://doi.org/10.1101/gr.190116.115>
- Cormier, M.J., Pedersen, B.S., Bayrak-Toydemir, P., Quinlan, A.R., 2022. Combining genetic constraint with predictions of alternative splicing to prioritize deleterious splicing in rare disease studies. *BMC Bioinformatics* 23, 482. <https://doi.org/10.1186/s12859-022-05041-x>
- Cox, D.M., Nelson, K.L., Clytone, M., Collins, D.L., 2018. Hereditary cancer screening: Case reports and review of literature on ten Ashkenazi Jewish founder mutations. *Mol Genet Genomic Med* 6, 1236–1242. <https://doi.org/10.1002/mgg3.460>
- Crawford, B.D., Gillies, C.E., Robertson, C.C., Kretzler, M., Otto, E.A., Vega-Warner, V., Sampson, M.G., 2017. Evaluating Mendelian nephrotic syndrome genes for evidence for risk alleles or oligogenicity that explain heritability. *Pediatr Nephrol* 32, 467–476. <https://doi.org/10.1007/s00467-016-3513-3>
- Cunha, K.S., Oliveira, N.S., Fausto, A.K., de Souza, C.C., Gros, A., Bandres, T., Idrissi, Y., Merlio, J.-P., de Moura Neto, R.S., Silva, R., Geller, M., Cappellen, D., 2016. Hybridization Capture-Based Next-Generation Sequencing to Evaluate Coding Sequence and Deep Intronic Mutations in the NF1 Gene. *Genes (Basel)* 7, 133. <https://doi.org/10.3390/genes7120133>
- Dallali, H., Kheriji, N., Kammoun, W., Mrad, M., Soltani, M., Trabelsi, H., Hamdi, W., Bahlous, A., Ben Ahmed, M., Mahjoub, F., Jamoussi, H., Abdelhak, S., Kefi, R., 2021. Multiallelic Rare Variants in BBS Genes Support an Oligogenic Ciliopathy in a Non-obese Juvenile-Onset Syndromic Diabetic Patient: A Case Report. *Front Genet* 12, 664963. <https://doi.org/10.3389/fgene.2021.664963>



- Dapas, M., Dunaif, A., 2020. The contribution of rare genetic variants to the pathogenesis of polycystic ovary syndrome. *Curr Opin Endocr Metab Res* 12, 26–32. <https://doi.org/10.1016/j.coemr.2020.02.011>
- D'Aurelio, M., Vives-Bauza, C., Davidson, M.M., Manfredi, G., 2010. Mitochondrial DNA background modifies the bioenergetics of NARP/MILS ATP6 mutant cells. *Hum Mol Genet* 19, 374–386. <https://doi.org/10.1093/hmg/ddp503>
- de Bruijn, S.E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U.S., Corral-Serrano, J.C., Mulders, T., Georgiou, M., Rivolta, C., Pontikos, N., Arno, G., Roberts, L., Greenberg, J., Albert, S., Gilissen, C., Aben, M., Rebello, G., Mead, S., Raymond, F.L., Corominas, J., Smith, C.E.L., Kremer, H., Downes, S., Black, G.C., Webster, A.R., Inglehearn, C.F., van den Born, L.I., Koenekoop, R.K., Michaelides, M., Ramesar, R.S., Hoyng, C.B., Mundlos, S., Mhlanga, M.M., Cremers, F.P.M., Cheetham, M.E., Roosing, S., Hardcastle, A.J., 2020. Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am J Hum Genet* 107, 802–814. <https://doi.org/10.1016/j.ajhg.2020.09.002>
- Deciphering Developmental Disorders Study, 2015. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 519, 223–228. <https://doi.org/10.1038/nature14135>
- Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., Bérout, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37, e67. <https://doi.org/10.1093/nar/gkp215>
- DiStefano, M.T., Goehringer, S., Babb, L., Alkuraya, F.S., Amberger, J., Amin, M., Austin-Tse, C., Balzotti, M., Berg, J.S., Birney, E., Bocchini, C., Bruford, E.A., Coffey, A.J., Collins, H., Cunningham, F., Daugherty, L.C., Einhorn, Y., Firth, H.V., Fitzpatrick, D.R., Foulger, R.E., Goldstein, J., Hamosh, A., Hurler, M.R., Leigh, S.E., Leong, I.U.S., Maddirevula, S., Martin, C.L., McDonagh, E.M., Olry, A., Puzriakova, A., Radtke, K., Ramos, E.M., Rath, A., Riggs, E.R., Roberts, A.M., Rodwell, C., Snow, C., Stark, Z., Tahiliani, J., Tweedie, S., Ware, J.S., Weller, P., Williams, E., Wright, C.F., Yates, T.M., Rehm, H.L., 2022. The Gene Curation Coalition: A global effort to harmonize gene-disease evidence resources. *Genet Med* 24, 1732–1742. <https://doi.org/10.1016/j.gim.2022.04.017>
- Donsante, A., Tang, J., Godwin, S.C., Holmes, C.S., Goldstein, D.S., Bassuk, A., Kaler, S.G., 2007. Differences in ATP7A gene expression underlie intrafamilial variability in Menkes disease/occipital horn syndrome. *J Med Genet* 44, 492–497. <https://doi.org/10.1136/jmg.2007.050013>
- Eichler, E.E., 2019. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *N Engl J Med* 381, 64–74. <https://doi.org/10.1056/NEJMr1809315>
- Ejigu, G.F., Jung, J., 2020. Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology (Basel)* 9, 295. <https://doi.org/10.3390/biology9090295>
- Ellingford, J.M., Ahn, J.W., Bagnall, R.D., Baralle, D., Barton, S., Campbell, C., Downes, K., Ellard, S., Duff-Farrier, C., FitzPatrick, D.R., Grealley, J.M., Ingles, J., Krishnan, N., Lord, J., Martin, H.C., Newman, W.G., O'Donnell-Luria, A., Ramsden, S.C., Rehm, H.L., Richardson, E., Singer-Berk, M., Taylor, J.C., Williams, M., Wood, J.C., Wright, C.F., Harrison, S.M., Whiffin, N., 2022. Recommendations for clinical interpretation of

- variants found in non-coding regions of the genome. *Genome Med* 14, 73.  
<https://doi.org/10.1186/s13073-022-01073-3>
- Ellingford, J.M., Campbell, C., Barton, S., Bhaskar, S., Gupta, S., Taylor, R.L., Sergouniotis, P.I., Horn, B., Lamb, J.A., Michaelides, M., Webster, A.R., Newman, W.G., Panda, B., Ramsden, S.C., Black, G.C., 2017. Validation of copy number variation analysis for next-generation sequencing diagnostics. *Eur J Hum Genet* 25, 719–724.  
<https://doi.org/10.1038/ejhg.2017.42>
- ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>
- Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R.K., Hauke, J., 2018. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Medical Genomics* 11, 35.  
<https://doi.org/10.1186/s12920-018-0353-y>
- EURORDIS. 2020. About Rare Diseases [Online]. Available at:  
<https://www.eurordis.org/about-rare-diseases> (accessed 1/6/20).
- Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* 8, 610–618. <https://doi.org/10.1038/nrg2146>
- Fan, Y., Silber, S.J., 1993. Y Chromosome Infertility, in: Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).
- Farwell, K.D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E.C., Tippin Davis, B., Baxter, R.M., Zeng, W., Mroske, C., Parra, M.C., Gandomi, S.K., Lu, I., Li, X., Lu, H., Lu, H.-M., Salvador, D., Ruble, D., Lao, M., Fischbach, S., Wen, J., Lee, S., Elliott, A., Dunlop, C.L.M., Tang, S., 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* 17, 578–586.  
<https://doi.org/10.1038/gim.2014.154>
- Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., Horton, R., Kerr, B., Kumar, D., Lehman, A., McKee, S., Morton, J., Parker, M.J., Rankin, J., Robertson, L., Temple, I.K., Study, C.A. of the U. of S. and E. as a S. (CAUSES), Study, T.D.D.D. (DDD), Banka, S., 2018. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. *American Journal of Human Genetics* 102, 175.  
<https://doi.org/10.1016/j.ajhg.2017.11.013>
- Fausser, S., Munz, M., Besch, D., 2003. Further support for digenic inheritance in Bardet-Biedl syndrome. *J Med Genet* 40, e104. <https://doi.org/10.1136/jmg.40.8.e104>
- Fernandez-Marmiesse, A., Gouveia, S., Couce, M.L., 2018. NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. *Curr. Med. Chem.* 25, 404–432. <https://doi.org/10.2174/0929867324666170718101946>
- Fichtman, B., Harel, T., Biran, N., Zagairy, F., Applegate, C.D., Salzberg, Y., Gilboa, T., Salah, S., Shaag, A., Simanovsky, N., Ayoubieh, H., Sobreira, N., Punzi, G., Pierri, C.L., Hamosh, A., Elpeleg, O., Harel, A., Edvardson, S., 2019. Pathogenic Variants in NUP214 Cause “Plugged” Nuclear Pore Channels and Acute Febrile Encephalopathy. *Am J Hum Genet* 105, 48–64. <https://doi.org/10.1016/j.ajhg.2019.05.003>
- Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., Shendure, J., 2014. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* 513, 120–123.  
<https://doi.org/10.1038/nature13695>

- Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Vooren, S.V., Moreau, Y., Pettett, R.M., Carter, N.P., 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *The American Journal of Human Genetics* 84, 524–533. <https://doi.org/10.1016/j.ajhg.2009.03.010>
- Firth, H.V., Wright, C.F., DDD Study, 2011. The Deciphering Developmental Disorders (DDD) study. *Dev Med Child Neurol* 53, 702–703. <https://doi.org/10.1111/j.1469-8749.2011.04032.x>
- Ganakammal, S.R., Alexov, E., 2019. Evaluation of performance of leading algorithms for variant pathogenicity predictions and designing a combinatory predictor method: application to Rett syndrome variants. *PeerJ* 7, e8106. <https://doi.org/10.7717/peerj.8106>
- Gando, I., Yang, H.-Q., Coetzee, W.A., 2019. Functional significance of channelopathy gene variants in unexplained death. *Forensic Sci Med Pathol* 15, 437–444. <https://doi.org/10.1007/s12024-018-0063-y>
- Gasparini, M., Starita, L., Shendure, J., 2016. The power of multiplexed functional analysis of genetic variants. *Nat Protoc* 11, 1782–1787. <https://doi.org/10.1038/nprot.2016.135>
- Gene Ontology Consortium, 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 43, D1049–1056. <https://doi.org/10.1093/nar/gku1179>
- Gifford, C.A., Ranade, S.S., Samarakoon, R., Salunga, H.T., de Soysa, T.Y., Huang, Y., Zhou, P., Elfenbein, A., Wyman, S.K., Bui, Y.K., Cordes Metzler, K.R., Ursell, P., Ivey, K.N., Srivastava, D., 2019. Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science* 364, 865–870. <https://doi.org/10.1126/science.aat5056>
- Gilissen, C., Hoischen, A., Brunner, H.G., Veltman, J.A., 2011. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 12, 228. <https://doi.org/10.1186/gb-2011-12-9-228>
- Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., de Vries, B.B.A., Kleefstra, T., Brunner, H.G., Vissers, L.E.L.M., Veltman, J.A., 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347.
- Grantham, R., 1974. Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. <https://doi.org/10.1126/science.185.4154.862>
- Green, D.J., Lenassi, E., Manning, C.S., McGaughey, D., Sharma, V., Black, G.C., Ellingford, J.M., Sergouniotis, P.I., 2021. North Carolina Macular Dystrophy: Phenotypic Variability and Computational Analysis of Disease-Associated Non-coding Variants. *Invest Ophthalmol Vis Sci* 62, 16. <https://doi.org/10.1167/iovs.62.7.16>
- Griffin, B.H., Chitty, L.S., Bitner-Glindzicz, M., 2017. The 100 000 Genomes Project: What it means for paediatrics. *Arch Dis Child Educ Pract Ed* 102, 105–107. <https://doi.org/10.1136/archdischild-2016-311029>
- Grimm, D.G., Azencott, C.-A., Aicheler, F., Gieraths, U., MacArthur, D.G., Samocha, K.E., Cooper, D.N., Stenson, P.D., Daly, M.J., Smoller, J.W., Duncan, L.E., Borgwardt, K.M., 2015. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Hum Mutat* 36, 513–523. <https://doi.org/10.1002/humu.22768>
- Gu, W., Zhou, A., Wang, L., Sun, S., Cui, X., Zhu, D., 2021. SVLR: Genome Structural Variant Detection Using Long-Read Sequencing Data. *J Comput Biol* 28, 774–788. <https://doi.org/10.1089/cmb.2021.0048>

- Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., Consortium, G.A.D., Rehm, H.L., MacArthur, D.G., O'Donnell-Luria, A., 2022. Variant interpretation using population databases: lessons from gnomAD. *Human Mutation* 43, 1012–1030. <https://doi.org/10.1002/humu.24309>
- Guimarães, J.R., Coêlho, M. de C., de Oliveira, N.F.P., 2022. Contribution of DNA methylation to the pathogenesis of Sjögren's syndrome: A review. *Autoimmunity* 55, 215–222. <https://doi.org/10.1080/08916934.2022.2062593>
- Gurbich, T.A., Ilinsky, V.V., 2020. ClassifyCNV: a tool for clinical annotation of copy-number variants. *Sci Rep* 10, 20375. <https://doi.org/10.1038/s41598-020-76425-3>
- Han, L., Zhao, X., Benton, M.L., Perumal, T., Collins, R.L., Hoffman, G.E., Johnson, J.S., Sloofman, L., Wang, H.Z., Stone, M.R., CommonMind Consortium, Brennand, K.J., Brand, H., Sieberts, S.K., Marengo, S., Peters, M.A., Lipska, B.K., Roussos, P., Capra, J.A., Talkowski, M., Ruderfer, D.M., 2020. Functional annotation of rare structural variation in the human brain. *Nat Commun* 11, 2990. <https://doi.org/10.1038/s41467-020-16736-1>
- Haynes, W.A., Tomczak, A., Khatri, P., 2018. Gene annotation bias impedes biomedical research. *Sci Rep* 8, 1362. <https://doi.org/10.1038/s41598-018-19333-x>
- Herlin, M.K., Le, V.Q., Højland, A.T., Ernst, A., Okkels, H., Petersen, A.C., Petersen, M.B., Pedersen, I.S., 2019. Whole-exome sequencing identifies a GREB1L variant in a three-generation family with Müllerian and renal agenesis: a novel candidate gene in Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. A case report. *Hum. Reprod.* 34, 1838–1846. <https://doi.org/10.1093/humrep/dez126>
- Heuer, B., Seibert, D.C., 2022. Mitochondrial disorders: Understanding mitochondrial DNA point mutations and deletion syndromes. *J Am Assoc Nurse Pract* 34, 954–956. <https://doi.org/10.1097/JXX.0000000000000755>
- Hindorff, L.A., Bonham, V.L., Brody, L.C., Ginoza, M.E.C., Hutter, C.M., Manolio, T.A., Green, E.D., 2018. Prioritizing diversity in human genomics research. *Nat Rev Genet* 19, 175–185. <https://doi.org/10.1038/nrg.2017.89>
- Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., Manolio, T.A., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106, 9362–9367. <https://doi.org/10.1073/pnas.0903103106>
- Holzer, G., De Magistris, P., Gramminger, C., Sachdev, R., Magalska, A., Schooley, A., Scheufen, A., Lennartz, B., Tatarek-Nossol, M., Lue, H., Linder, M.I., Kutay, U., Preisinger, C., Moreno-Andres, D., Antonin, W., 2021. The nucleoporin Nup50 activates the Ran guanine nucleotide exchange factor RCC1 to promote NPC assembly at the end of mitosis. *EMBO J* 40, e108788. <https://doi.org/10.15252/embj.2021108788>
- Homburger, J.R., Neben, C.L., Mishne, G., Zhou, A.Y., Kathiresan, S., Khera, A.V., 2019. Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. *Genome Med* 11, 74. <https://doi.org/10.1186/s13073-019-0682-2>
- Hong, Q., Li, X.-D., Xie, P., Du, S.-X., 2021. All-trans-retinoic acid suppresses rat embryo hindlimb bud mesenchymal chondrogenesis by modulating HoxD9 expression. *Bioengineered* 12, 3900–3911. <https://doi.org/10.1080/21655979.2021.1940613>
- Horinouchi, T., Nozu, K., Yamamura, T., Minamikawa, S., Nagano, C., Sakakibara, N., Nakanishi, K., Shima, Y., Morisada, N., Ishiko, S., Aoto, Y., Nagase, H., Takeda, H.,

- Rossanti, R., Kaito, H., Matsuo, M., Iijima, K., 2019. Determination of the pathogenicity of known COL4A5 intronic variants by in vitro splicing assay. *Sci Rep* 9, 12696. <https://doi.org/10.1038/s41598-019-48990-9>
- Horn, N., Wittung-Stafshede, P., 2021. ATP7A-Regulated Enzyme Metalation and Trafficking in the Menkes Disease Puzzle. *Biomedicines* 9, 391. <https://doi.org/10.3390/biomedicines9040391>
- Hsu, J.S.J., So, M., Tang, C.S.M., Karim, A., Porsch, R.M., Wong, C., Yu, M., Yeung, F., Xia, H., Zhang, R., Cherny, S.S., Chung, P.H.Y., Wong, K.K.Y., Sham, P.C., Ngo, N.D., Li, M., Tam, P.K.H., Lui, V.C.H., Garcia-Barcelo, M.-M., 2018. De novo mutations in Caudal Type Homeo Box transcription Factor 2 (CDX2) in patients with persistent cloaca. *Hum Mol Genet* 27, 351–358. <https://doi.org/10.1093/hmg/ddx406>
- Iglesias, A., Anyane-Yeboah, K., Wynn, J., Wilson, A., Truitt Cho, M., Guzman, E., Sisson, R., Egan, C., Chung, W.K., 2014. The usefulness of whole-exome sequencing in routine clinical practice. *Genet. Med.* 16, 922–931. <https://doi.org/10.1038/gim.2014.58>
- Inoue, F., Ahituv, N., 2015. Decoding enhancers using massively parallel reporter assays. *Genomics* 106, 159–164. <https://doi.org/10.1016/j.ygeno.2015.06.005>
- Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., Sieh, W., 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* 99, 877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., McVean, G., 2012. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* 44, 226–232. <https://doi.org/10.1038/ng.1028>
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S.J., Farh, K.K.-H., 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>
- Jamal, A.R.A., 2021. Precision Medicine: Making It Happen for Malaysia. *Malays J Med Sci* 28, 1–4. <https://doi.org/10.21315/mjms2021.28.3.1>
- Jenko Bizjan, B., Katsila, T., Tesovnik, T., Šket, R., Debeljak, M., Matsoukas, M.T., Kovač, J., 2020. Challenges in identifying large germline structural variants for clinical use by long read sequencing. *Comput Struct Biotechnol J* 18, 83–92. <https://doi.org/10.1016/j.csbj.2019.11.008>
- Jiang, Y., Wang, Y., Brudno, M., 2012. PRISM: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics* 28, 2576–2583. <https://doi.org/10.1093/bioinformatics/bts484>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44, D457–462. <https://doi.org/10.1093/nar/gkv1070>
- Kanie, T., Jackson, P.K., 2018. Guanine Nucleotide Exchange Assay Using Fluorescent MANT-GDP. *Bio Protoc* 8, e2795. <https://doi.org/10.21769/BioProtoc.2795>

- Kannan, T.P., Zilfalil, B.A., 2009. Cytogenetics: past, present and future. *Malays J Med Sci* 16, 4–9.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Consortium, T.G.A.D., Neale, B.M., Daly, M.J., MacArthur, D.G., 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210. <https://doi.org/10.1101/531210>
- Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. <https://doi.org/10.1038/ng.2892>
- Kitts, A., Sherry, S., 2011. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation, The NCBI Handbook [Internet]. National Center for Biotechnology Information (US).
- Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., FitzPatrick, D.R., Eppig, J.T., Jackson, A.P., Freson, K., Girdea, M., Helbig, I., Hurst, J.A., Jähn, J., Jackson, L.G., Kelly, A.M., Ledbetter, D.H., Mansour, S., Martin, C.L., Moss, C., Mumford, A., Ouwehand, W.H., Park, S.-M., Riggs, E.R., Scott, R.H., Sisodiya, S., Van Vooren, S., Wapner, R.J., Wilkie, A.O.M., Wright, C.F., Vulto-van Silfhout, A.T., de Leeuw, N., de Vries, B.B.A., Washington, N.L., Smith, C.L., Westerfield, M., Schofield, P., Ruef, B.J., Gkoutos, G.V., Haendel, M., Smedley, D., Lewis, S.E., Robinson, P.N., 2014. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 42, D966-974. <https://doi.org/10.1093/nar/gkt1026>

- Konkle, B.A., Nakaya Fletcher, S., 1993. Hemophilia A, in: Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).
- Krumm, N., Sudmant, P.H., Ko, A., O’Roak, B.J., Malig, M., Coe, B.P., NHLBI Exome Sequencing Project, Quinlan, A.R., Nickerson, D.A., Eichler, E.E., 2012. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 22, 1525–1532. <https://doi.org/10.1101/gr.138115.112>
- Kumar, K.R., Davis, R.L., Tchan, M.C., Wali, G.M., Mahant, N., Ng, K., Kotschet, K., Siow, S.-F., Gu, J., Walls, Z., Kang, C., Wali, G., Levy, S., Phua, C.S., Yiannikas, C., Darveniza, P., Chang, F.C.F., Morales-Briceño, H., Rowe, D.B., Drew, A., Gayevskiy, V., Cowley, M.J., Minoche, A.E., Tisch, S., Hayes, M., Kummerfeld, S., Fung, V.S.C., Sue, C.M., 2019. Whole genome sequencing for the genetic diagnosis of heterogenous dystonia phenotypes. *Parkinsonism Relat. Disord.* 69, 111–118. <https://doi.org/10.1016/j.parkreldis.2019.11.004>
- 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., Lehoczy, 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, Y., 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., Raymond, C., 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., Blöcker, 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., Kasprzyk, 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., Szustakowki, 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., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. <https://doi.org/10.1038/35057062>
- Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., Maglott, D.R., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42, D980-985. <https://doi.org/10.1093/nar/gkt1113>
- Lau, Y.-F.C., 2020. Y chromosome in health and diseases. *Cell & Bioscience* 10, 97. <https://doi.org/10.1186/s13578-020-00452-w>
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. <https://doi.org/10.1038/nature19057>
- Leman, R., Tubeuf, H., Raad, S., Tournier, I., Derambure, C., Lanos, R., Gaildrat, P., Castelain, G., Hauchard, J., Killian, A., Baert-Desurmont, S., Legros, A., Goardon, N., Quesnelle, C., Ricou, A., Castera, L., Vaur, D., Le Gac, G., Ka, C., Fichou, Y., Bonnet-Dorion, F., Sevenet, N., Guillaud-Bataille, M., Boutry-Kryza, N., Schultz, I., Caux-Moncoutier, V., Rossing, M., Walker, L.C., Spurdle, A.B., Houdayer, C., Martins, A., Krieger, S., 2020. Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. *BMC Genomics* 21, 86. <https://doi.org/10.1186/s12864-020-6484-5>
- Levine, J.M., Ahsan, N., Ho, E., Santoro, J.D., 2020. Genetic Acute Necrotizing Encephalopathy Associated with RANBP2: Clinical and Therapeutic Implications in Pediatrics. *Mult Scler Relat Disord* 43, 102194. <https://doi.org/10.1016/j.msard.2020.102194>
- Li, B., Brusman, L., Dahlka, J., Niswander, L.A., 2022. TMEM132A ensures mouse caudal neural tube closure and regulates integrin-based mesodermal migration. *Development* 149, dev200442. <https://doi.org/10.1242/dev.200442>
- Li, M.-X., Kwan, J.S.H., Bao, S.-Y., Yang, W., Ho, S.-L., Song, Y.-Q., Sham, P.C., 2013. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet* 9, e1003143. <https://doi.org/10.1371/journal.pgen.1003143>



- Li, X., Kim, Y., Tsang, E.K., Davis, J.R., Damani, F.N., Chiang, C., Hess, G.T., Zappala, Z., Strober, B.J., Scott, A.J., Li, A., Ganna, A., Bassik, M.C., Merker, J.D., GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration & Visualization—EBI, Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Hall, I.M., Battle, A., Montgomery, S.B., 2017. The impact of rare variation on gene expression across tissues. *Nature* 550, 239–243. <https://doi.org/10.1038/nature24267>
- Li, Y., Jia, X., Wu, H., Xun, G., Ou, J., Zhang, Q., Li, H., Bai, T., Hu, Z., Zou, X., Xia, K., Guo, H., 2018. Genotype and phenotype correlations for SHANK3 de novo mutations in neurodevelopmental disorders. *Am J Med Genet A* 176, 2668–2676. <https://doi.org/10.1002/ajmg.a.40666>
- Lim, E.T., Würtz, P., Havulinna, A.S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi, R., Inouye, M., Lappalainen, T., Chan, Y., Salem, R.M., Lek, M., Flannick, J., Sim, X., Manning, A., Ladenvall, C., Bumpstead, S., Hämäläinen, E., Aalto, K., Maksimow, M., Salmi, M., Blankenberg, S., Ardisson, D., Shah, S., Horne, B., McPherson, R., Hovingh, G.K., Reilly, M.P., Watkins, H., Goel, A., Farrall, M., Girelli, D., Reiner, A.P., Stitzel, N.O., Kathiresan, S., Gabriel, S., Barrett, J.C., Lehtimäki, T., Laakso, M., Groop, L., Kaprio, J., Perola, M., McCarthy, M.I., Boehnke, M., Altshuler, D.M., Lindgren, C.M., Hirschhorn, J.N., Metspalu, A., Freimer, N.B., Zeller, T., Jalkanen, S., Koskinen, S., Raitakari, O., Durbin, R., MacArthur, D.G., Salomaa, V., Ripatti, S., Daly, M.J., Palotie, A., Sequencing Initiative Suomi (SISu) Project, 2014. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* 10, e1004494. <https://doi.org/10.1371/journal.pgen.1004494>
- Liu, P., Meng, L., Normand, E.A., Xia, F., Song, X., Ghazi, A., Rosenfeld, J., Magoulas, P.L., Braxton, A., Ward, P., Dai, H., Yuan, B., Bi, W., Xiao, R., Wang, X., Chiang, T., Vetrini, F., He, W., Cheng, H., Dong, J., Gijavanekar, C., Benke, P.J., Bernstein, J.A., Eble, T., Eroglu, Y., Erwin, D., Escobar, L., Gibson, J.B., Gripp, K., Kleppe, S., Koenig, M.K., Lewis, A.M., Natowicz, M., Mancias, P., Minor, L., Scaglia, F., Schaaf, C.P., Streff, H., Vernon, H., Uhles, C.L., Zackai, E.H., Wu, N., Sutton, V.R., Beaudet, A.L., Muzny, D., Gibbs, R.A., Posey, J.E., Lalani, S., Shaw, C., Eng, C.M., Lupski, J.R., Yang, Y., 2019. Reanalysis of Clinical Exome Sequencing Data. *New England Journal of Medicine* 380, 2478–2480. <https://doi.org/10.1056/NEJMc1812033>
- Loose, M., Auer, A., Brognara, G., Budiman, H.R., Kowalski, L., Matijević, I., 2022. In vitro reconstitution of small GTPase regulation. *FEBS Lett.* <https://doi.org/10.1002/1873-3468.14540>
- MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., Barrett, J.C., Biesecker, L.G., Conrad, D.F., Cooper, G.M., Cox, N.J., Daly, M.J., Gerstein, M.B., Goldstein, D.B., Hirschhorn, J.N., Leal, S.M., Pennacchio, L.A., Stamatoyannopoulos, J.A., Sunyaev, S.R., Valle, D., Voight, B.F., Winckler, W., Gunter, C., 2014. Guidelines for

- investigating causality of sequence variants in human disease. *Nature* 508, 469–476. <https://doi.org/10.1038/nature13127>
- MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42, D986–992. <https://doi.org/10.1093/nar/gkt958>
- Majumder, P.P., Basu, A., 2014. A genomic view of the peopling and population structure of India. *Cold Spring Harb Perspect Biol* 7, a008540. <https://doi.org/10.1101/cshperspect.a008540>
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. *Nature* 461, 747–753. <https://doi.org/10.1038/nature08494>
- Martin-Trujillo, A., Patel, N., Richter, F., Jadhav, B., Garg, P., Morton, S.U., McKean, D.M., DePalma, S.R., Goldmuntz, E., Gruber, D., Kim, R., Newburger, J.W., Porter, G.A., Giardini, A., Bernstein, D., Tristani-Firouzi, M., Seidman, J.G., Seidman, C.E., Chung, W.K., Gelb, B.D., Sharp, A.J., 2020. Rare genetic variation at transcription factor binding sites modulates local DNA methylation profiles. *PLoS Genet* 16, e1009189. <https://doi.org/10.1371/journal.pgen.1009189>
- Mastana, S.S., 2014. Unity in diversity: an overview of the genomic anthropology of India. *Annals of Human Biology* 41, 287–299. <https://doi.org/10.3109/03014460.2014.922615>
- Matalonga, L., Hernández-Ferrer, C., Piscia, D., Solve-RD SNV-indel working group, Schüle, R., Synofzik, M., Töpf, A., Vissers, L.E.L.M., de Voer, R., Solve-RD DITF-GENTURIS, Solve-RD DITF-ITHACA, Solve-RD DITF-euroNMD, Solve-RD DITF-RND, Tonda, R., Laurie, S., Fernandez-Callejo, M., Picó, D., Garcia-Linares, C., Papakonstantinou, A., Corvó, A., Joshi, R., Diez, H., Gut, I., Hoischen, A., Graessner, H., Beltran, S., Solve-RD Consortia, 2021. Solving patients with rare diseases through programmatic reanalysis of genome-phenome data. *Eur J Hum Genet* 29, 1337–1347. <https://doi.org/10.1038/s41431-021-00852-7>
- Mayor, S., 2014. Programme for genetic diagnosis of rare diseases proves feasible for routine clinical practice. *BMJ* 349, g7781. <https://doi.org/10.1136/bmj.g7781>
- McArthur, E., Capra, J.A., 2021. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* 108, 269–283. <https://doi.org/10.1016/j.ajhg.2021.01.001>
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., Hirschhorn, J.N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369. <https://doi.org/10.1038/nrg2344>
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. *Genome Biology* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>
- McNally, E., MacLeod, H., Dellefave-Castillo, L., 1993. Arrhythmogenic Right Ventricular Cardiomyopathy, in: Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), *GeneReviews*®. University of Washington, Seattle, Seattle (WA).

- Meinel, J.A., Yumiceba, V., Künstner, A., Schultz, K., Kruse, N., Kaiser, F.J., Holterhus, P.-M., Claviez, A., Hiort, O., Busch, H., Spielmann, M., Werner, R., 2023. Disruption of the topologically associated domain at Xp21.2 is related to 46,XY gonadal dysgenesis. *J Med Genet* 60, 469–476. <https://doi.org/10.1136/jmg-2022-108635>
- Mhaske, A., Dileep, K.V., Kumar, M., Poojary, M., Pandhare, K., Zhang, K.Y.J., Scaria, V., Binukumar, B.K., 2020. ATP7A Clinical Genetics Resource - A comprehensive clinically annotated database and resource for genetic variants in ATP7A gene. *Comput Struct Biotechnol J* 18, 2347–2356. <https://doi.org/10.1016/j.csbj.2020.08.021>
- Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.-M., Selvaraj, S., Dennis, M.Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R. ana, Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E., Phillippy, A.M., 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585, 79–84. <https://doi.org/10.1038/s41586-020-2547-7>
- Minoche, A.E., Horvat, C., Johnson, R., Gayevskiy, V., Morton, S.U., Drew, A.P., Woo, K., Statham, A.L., Lundie, B., Bagnall, R.D., Ingles, J., Semsarian, C., Seidman, J.G., Seidman, C.E., Dinger, M.E., Cowley, M.J., Fatkin, D., 2019. Genome sequencing as a first-line genetic test in familial dilated cardiomyopathy. *Genet Med* 21, 650–662. <https://doi.org/10.1038/s41436-018-0084-7>
- Mitsuhashi, S., Matsumoto, N., 2020. Long-read sequencing for rare human genetic diseases. *J Hum Genet* 65, 11–19. <https://doi.org/10.1038/s10038-019-0671-8>
- Møller, L.B., Mogensen, M., Weaver, D.D., Pedersen, P.A., 2021. Occipital Horn Syndrome as a Result of Splice Site Mutations in ATP7A. No Activity of ATP7A Splice Variants Missing Exon 10 or Exon 15. *Frontiers in Molecular Neuroscience* 14.
- Moore, L.D., Le, T., Fan, G., 2013. DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38. <https://doi.org/10.1038/npp.2012.112>
- Murakami, H., Kimura, Y., Enomoto, Y., Tsurusaki, Y., Akahira-Azuma, M., Kuroda, Y., Tsuji, M., Goto, T., Kurosawa, K., 2019. Discordant phenotype caused by CASK mutation in siblings with NF1. *Hum Genome Var* 6, 20. <https://doi.org/10.1038/s41439-019-0051-0>
- Nakatsuka, N., Moorjani, P., Rai, N., Sarkar, B., Tandon, A., Patterson, N., Bhavani, G.S., Girisha, K.M., Mustak, M.S., Srinivasan, S., Kaushik, A., Vahab, S.A., Jagadeesh, S.M., Satyamoorthy, K., Singh, L., Reich, D., Thangaraj, K., 2017. The promise of discovering population-specific disease-associated genes in South Asia. *Nat Genet* 49, 1403–1407. <https://doi.org/10.1038/ng.3917>
- Nam, S.H., Kanwal, S., Nam, D.E., Lee, M.H., Kang, T.H., Jung, S.-C., Choi, B.-O., Chung, K.W., 2018. Association of miR-149 polymorphism with onset age and severity in Charcot-Marie-Tooth disease type 1A. *Neuromuscul Disord* 28, 502–507. <https://doi.org/10.1016/j.nmd.2018.04.002>
- Nelson, A., Myers, K., 1993. Shwachman-Diamond Syndrome, in: Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).

- Neumeyer, S., Hemani, G., Zeggini, E., 2019. Strengthening Causal Inference for Complex Disease Using Molecular Quantitative Trait Loci. *Trends Mol Med*. <https://doi.org/10.1016/j.molmed.2019.10.004>
- Newman, W.G., Woolf, A.S., 1993. Urofacial Syndrome, in: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Stephens, K., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).
- Nguengang Wakap, S., Lambert, D.M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., Rath, A., 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 28, 165–173. <https://doi.org/10.1038/s41431-019-0508-0>
- Norgett, E.E., Hatsell, S.J., Carvajal-Huerta, L., Cabezas, J.C., Common, J., Purkis, P.E., Whittock, N., Leigh, I.M., Stevens, H.P., Kelsell, D.P., 2000. Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. *Hum Mol Genet* 9, 2761–2766. <https://doi.org/10.1093/hmg/9.18.2761>
- Ochoa, B., 2004. Can a congenital dysfunctional bladder be diagnosed from a smile? The Ochoa syndrome updated. *Pediatr. Nephrol.* 19, 6–12. <https://doi.org/10.1007/s00467-003-1291-1>
- Ohno, K., Takeda, J.-I., Masuda, A., 2018. Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA* 9. <https://doi.org/10.1002/wrna.1451>
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M.R., Zschocke, J., Trajanoski, Z., 2014. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 15, 256–278. <https://doi.org/10.1093/bib/bbs086>
- Pant, S., Weiner, R., Marton, M.J., 2014. Navigating the Rapids: The Development of Regulated Next-Generation Sequencing-Based Clinical Trial Assays and Companion Diagnostics. *Front Oncol* 4. <https://doi.org/10.3389/fonc.2014.00078>
- Papadimitriou, S., Gazzo, A., Versbraegen, N., Nachtegaal, C., Aerts, J., Moreau, Y., Van Dooren, S., Nowé, A., Smits, G., Lenaerts, T., 2019. Predicting disease-causing variant combinations. *Proc Natl Acad Sci U S A* 116, 11878–11887. <https://doi.org/10.1073/pnas.1815601116>
- Parthasarathy, S., Ruggiero, S.M., Gelot, A., Soardi, F.C., Ribeiro, B.F.R., Pires, D.E.V., Ascher, D.B., Schmitt, A., Rambaud, C., Represa, A., Xie, H.M., Lusk, L., Wilmarth, O., McDonnell, P.P., Juarez, O.A., Grace, A.N., Buratti, J., Mignot, C., Gras, D., Nava, C., Pierce, S.R., Keren, B., Kennedy, B.C., Pena, S.D.J., Helbig, I., Cuddapah, V.A., 2022. A recurrent de novo splice site variant involving DNMT1 exon 10a causes developmental and epileptic encephalopathy through a dominant-negative mechanism. *Am J Hum Genet* S0002-9297(22)00493–1. <https://doi.org/10.1016/j.ajhg.2022.11.002>
- Pauper, M., Kucuk, E., Wenger, A.M., Chakraborty, S., Baybayan, P., Kwint, M., van der Sanden, B., Nelen, M.R., Derks, R., Brunner, H.G., Hoischen, A., Vissers, L.E.L.M., Gilissen, C., 2021. Long-read trio sequencing of individuals with unsolved intellectual disability. *Eur J Hum Genet* 29, 637–648. <https://doi.org/10.1038/s41431-020-00770-0>
- Pena, I.A., Roussel, Y., Daniel, K., Mongeon, K., Johnstone, D., Weinschutz Mendes, H., Bosma, M., Saxena, V., Lepage, N., Chakraborty, P., Dymont, D.A., van Karnebeek, C.D.M., Verhoeven-Duif, N., Bui, T.V., Boycott, K.M., Ekker, M., MacKenzie, A., 2017.

- Pyridoxine-Dependent Epilepsy in Zebrafish Caused by Aldh7a1 Deficiency. *Genetics* 207, 1501–1518. <https://doi.org/10.1534/genetics.117.300137>
- Pereira, S.V.-N., Ribeiro, J.D., Ribeiro, A.F., Bertuzzo, C.S., Marson, F.A.L., 2019. Novel, rare and common pathogenic variants in the CFTR gene screened by high-throughput sequencing technology and predicted by in silico tools. *Sci Rep* 9. <https://doi.org/10.1038/s41598-019-42404-6>
- Pérez-Torras, S., Mata-Ventosa, A., Drögemöller, B., Tarailo-Graovac, M., Meijer, J., Meisma, R., van Cruchten, A.G., Kulik, W., Viel-Oliva, A., Bidon-Chanal, A., Ross, C.J., Wassermann, W.W., van Karnebeek, C.D.M., Pastor-Anglada, M., van Kuilenburg, A.B.P., 2019. Deficiency of perforin and hCNT1, a novel inborn error of pyrimidine metabolism, associated with a rapidly developing lethal phenotype due to multi-organ failure. *Biochim Biophys Acta Mol Basis Dis* 1865, 1182–1191. <https://doi.org/10.1016/j.bbadis.2019.01.013>
- Peterson, T.A., Doughty, E., Kann, M.G., 2013. Towards precision medicine: advances in computational approaches for the analysis of human variants. *J Mol Biol* 425, 4047–4063. <https://doi.org/10.1016/j.jmb.2013.08.008>
- Pham, D., Truong, B., Tran, K., Ni, G., Nguyen, D., Tran, T.T.H., Tran, M.H., Nguyen Thuy, D., Vo, N.S., Nguyen, Q., 2022. Assessing polygenic risk score models for applications in populations with under-represented genomics data: an example of Vietnam. *Brief Bioinform* 23, bbac459. <https://doi.org/10.1093/bib/bbac459>
- Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., Dumitriu, S., Dyke, S.O.M., den Dunnen, J.T., Firth, H.V., Gibbs, R.A., Girdea, M., Gonzalez, M., Haendel, M.A., Hamosh, A., Holm, I.A., Huang, L., Hurles, M.E., Hutton, B., Krier, J.B., Misyura, A., Mungall, C.J., Paschall, J., Paten, B., Robinson, P.N., Schiettecatte, F., Sobreira, N.L., Swaminathan, G.J., Taschner, P.E., Terry, S.F., Washington, N.L., Züchner, S., Boycott, K.M., Rehm, H.L., 2015. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* 36, 915–921. <https://doi.org/10.1002/humu.22858>
- Pipis, M., Rossor, A.M., Laura, M., Reilly, M.M., 2019. Next-generation sequencing in Charcot-Marie-Tooth disease: opportunities and challenges. *Nat Rev Neurol* 15, 644–656. <https://doi.org/10.1038/s41582-019-0254-5>
- Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., Kumararatne, D., Doffinger, R., Nejentsev, S., 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 28, 2747–2754. <https://doi.org/10.1093/bioinformatics/bts526>
- Plon, S.E., Eccles, D.M., Easton, D., Foulkes, W.D., Genuardi, M., Greenblatt, M.S., Hogervorst, F.B.L., Hoogerbrugge, N., Spurdle, A.B., Tavtigian, S.V., IARC Unclassified Genetic Variants Working Group, 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 29, 1282–1291. <https://doi.org/10.1002/humu.20880>
- Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35, D61–65. <https://doi.org/10.1093/nar/gkl842>
- Rahit, K.M.T.H., Tarailo-Graovac, M., 2020. Genetic Modifiers and Rare Mendelian Disease. *Genes (Basel)* 11, 239. <https://doi.org/10.3390/genes11030239>

- Rajasimha, H.K., Shirol, P.B., Ramamoorthy, P., Hegde, M., Barde, S., Chandru, V., Ravinandan, M.E., Ramchandran, R., Haldar, K., Lin, J.C., Babar, I.A., Girisha, K.M., Srinivasan, S., Navaneetham, D., Battu, R., Devarakonda, R., Kini, U., Vijayachandra, K., Verma, I.C., 2014. Organization for rare diseases India (ORDI) - addressing the challenges and opportunities for the Indian rare diseases' community. *Genet Res (Camb)* 96, e009. <https://doi.org/10.1017/S0016672314000111>
- Rang, F.J., Kloosterman, W.P., de Ridder, J., 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* 19, 90. <https://doi.org/10.1186/s13059-018-1462-9>
- Rapti, M., Zouaghi, Y., Meylan, J., Ranza, E., Antonarakis, S.E., Santoni, F.A., 2022. CoverageMaster: comprehensive CNV detection and visualization from NGS short reads for genetic medicine applications. *Brief Bioinform* 23, bbac049. <https://doi.org/10.1093/bib/bbac049>
- Rare disease tiering - Genomics England Research Environment User Guide [WWW Document], n.d. URL <https://re-docs.genomicsengland.co.uk/tiering/> (accessed 12.17.22).
- Renaux, A., Papadimitriou, S., Versbraegen, N., Nachtegaal, C., Boutry, S., Nowé, A., Smits, G., Lenaerts, T., 2019. ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Res* 47, W93–W98. <https://doi.org/10.1093/nar/gkz437>
- Rentzsch, P., Schubach, M., Shendure, J., Kircher, M., 2021. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* 13, 31. <https://doi.org/10.1186/s13073-021-00835-9>
- Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., Kircher, M., 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886–D894. <https://doi.org/10.1093/nar/gky1016>
- Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39, e118. <https://doi.org/10.1093/nar/gkr407>
- Rexach, J., Lee, H., Martinez-Agosto, J.A., Németh, A.H., Fogel, B.L., 2019. Clinical application of next-generation sequencing to the practice of neurology. *Lancet Neurol* 18, 492–503. [https://doi.org/10.1016/S1474-4422\(19\)30033-X](https://doi.org/10.1016/S1474-4422(19)30033-X)
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17, 405–424. <https://doi.org/10.1038/gim.2015.30>
- Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., Pineda-Alvarez, D., Aradhya, S., Martin, C.L., 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genetics in Medicine* 22, 245–257. <https://doi.org/10.1038/s41436-019-0686-8>
- Rosenberg, A.B., Patwardhan, R.P., Shendure, J., Seelig, G., 2015. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* 163, 698–711. <https://doi.org/10.1016/j.cell.2015.09.054>

Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., Frankish, A., Lovell, F.L., Howe, K.L., Ashurst, J.L., Fulton, R.S., Sudbrak, R., Wen, G., Jones, M.C., Hurles, M.E., Andrews, T.D., Scott, C.E., Searle, S., Ramser, J., Whittaker, A., Deadman, R., Carter, N.P., Hunt, S.E., Chen, R., Cree, A., Gunaratne, P., Havlak, P., Hodgson, A., Metzker, M.L., Richards, S., Scott, G., Steffen, D., Sodergren, E., Wheeler, D.A., Worley, K.C., Ainscough, R., Ambrose, K.D., Ansari-Lari, M.A., Aradhya, S., Ashwell, R.I.S., Babbage, A.K., Bagguley, C.L., Ballabio, A., Banerjee, R., Barker, G.E., Barlow, K.F., Barrett, I.P., Bates, K.N., Beare, D.M., Beasley, H., Beasley, O., Beck, A., Bethel, G., Blechschmidt, K., Brady, N., Bray-Allen, S., Bridgeman, A.M., Brown, A.J., Brown, M.J., Bonnini, D., Bruford, E.A., Buhay, C., Burch, P., Burford, D., Burgess, J., Burrill, W., Burton, J., Bye, J.M., Carder, C., Carrel, L., Chako, J., Chapman, J.C., Chavez, D., Chen, E., Chen, G., Chen, Y., Chen, Z., Chinault, C., Ciccodicola, A., Clark, S.Y., Clarke, G., Clee, C.M., Clegg, S., Clerc-Blankenburg, K., Clifford, K., Copley, V., Cole, C.G., Conquer, J.S., Corby, N., Connor, R.E., David, R., Davies, J., Davis, C., Davis, J., Delgado, O., DeShazo, D., Dhami, P., Ding, Y., Dinh, H., Dodsworth, S., Draper, H., Dugan-Rocha, S., Dunham, A., Dunn, M., Durbin, K.J., Dutta, I., Eades, T., Ellwood, M., Emery-Cohen, A., Errington, H., Evans, K.L., Faulkner, L., Francis, F., Frankland, J., Fraser, A.E., Galgoczy, P., Gilbert, J., Gill, R., Glöckner, G., Gregory, S.G., Gribble, S., Griffiths, C., Grocock, R., Gu, Y., Gwilliam, R., Hamilton, C., Hart, E.A., Hawes, A., Heath, P.D., Heitmann, K., Hennig, S., Hernandez, J., Hinzmann, B., Ho, S., Hoffs, M., Howden, P.J., Huckle, E.J., Hume, J., Hunt, P.J., Hunt, A.R., Isherwood, J., Jacob, L., Johnson, D., Jones, S., de Jong, P.J., Joseph, S.S., Keenan, S., Kelly, S., Kershaw, J.K., Khan, Z., Kioschis, P., Klages, S., Knights, A.J., Kosiura, A., Kovar-Smith, C., Laird, G.K., Langford, C., Lawlor, S., Leversha, M., Lewis, L., Liu, W., Lloyd, C., Lloyd, D.M., Louseged, H., Loveland, J.E., Lovell, J.D., Lozado, R., Lu, J., Lyne, R., Ma, J., Maheshwari, M., Matthews, L.H., McDowall, J., McLaren, S., McMurray, A., Meidl, P., Meitinger, T., Milne, S., Miner, G., Mistry, S.L., Morgan, M., Morris, S., Müller, I., Mullikin, J.C., Nguyen, N., Nordsiek, G., Nyakatura, G., O'Dell, C.N., Okwuonu, G., Palmer, S., Pandian, R., Parker, D., Parrish, J., Pasternak, S., Patel, D., Pearce, A.V., Pearson, D.M., Pelan, S.E., Perez, L., Porter, K.M., Ramsey, Y., Reichwald, K., Rhodes, S., Ridler, K.A., Schlessinger, D., Schueler, M.G., Sehra, H.K., Shaw-Smith, C., Shen, H., Sheridan, E.M., Shownkeen, R., Skuce, C.D., Smith, M.L., Sotharan, E.C., Steingruber, H.E., Steward, C.A., Storey, R., Swann, R.M., Swarbreck, D., Tabor, P.E., Taudien, S., Taylor, T., Teague, B., Thomas, K., Thorpe, A., Timms, K., Tracey, A., Trevanion, S., Tromans, A.C., d'Urso, M., Verduzco, D., Villasana, D., Waldron, L., Wall, M., Wang, Q., Warren, J., Warry, G.L., Wei, X., West, A., Whitehead, S.L., Whiteley, M.N., Wilkinson, J.E., Willey, D.L., Williams, G., Williams, L., Williamson, A., Williamson, H., Wilming, L., Woodmansey, R.L., Wray, P.W., Yen, J., Zhang, J., Zhou, J., Zoghbi, H., Zorilla, S., Buck, D., Reinhardt, R., Poustka, A., Rosenthal, A., Lehrach, H., Meindl, A., Minx, P.J., Hillier, L.W., Willard, H.F., Wilson, R.K., Waterston, R.H., Rice, C.M., Vaudin, M., Coulson, A., Nelson, D.L., Weinstock, G., Sulston, J.E., Durbin, R., Hubbard, T., Gibbs, R.A., Beck, S., Rogers, J., Bentley, D.R., 2005. The DNA sequence of the human X chromosome. *Nature* 434, 325–337. <https://doi.org/10.1038/nature03440>

Rowlands, C.F., Baralle, D., Ellingford, J.M., 2019. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells* 8, E1513. <https://doi.org/10.3390/cells8121513>

- Sackton, T.B., Hartl, D.L., 2016. Genotypic Context and Epistasis in Individuals and Populations. *Cell* 166, 279–287. <https://doi.org/10.1016/j.cell.2016.06.047>
- Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., Wall, D.P., MacArthur, D.G., Gabriel, S.B., DePristo, M., Purcell, S.M., Palotie, A., Boerwinkle, E., Buxbaum, J.D., Cook, E.H., Gibbs, R.A., Schellenberg, G.D., Sutcliffe, J.S., Devlin, B., Roeder, K., Neale, B.M., Daly, M.J., 2014. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 46, 944–950. <https://doi.org/10.1038/ng.3050>
- Sanchis-Juan, A., Stephens, J., French, C.E., Gleadall, N., Mégy, K., Penkett, C., Shamardina, O., Stirrups, K., Delon, I., Dewhurst, E., Dolling, H., Erwood, M., Grozeva, D., Stefanucci, L., Arno, G., Webster, A.R., Cole, T., Austin, T., Branco, R.G., Ouwehand, W.H., Raymond, F.L., Carss, K.J., 2018. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med* 10, 95. <https://doi.org/10.1186/s13073-018-0606-6>
- Savant, A., Lyman, B., Bojanowski, C., Upadia, J., 1993. Cystic Fibrosis, in: Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A. (Eds.), *GeneReviews®*. University of Washington, Seattle, Seattle (WA).
- Sawyer, S.L., Hartley, T., Dymont, D.A., Beaulieu, C.L., Schwartzenruber, J., Smith, A., Bedford, H.M., Bernard, G., Bernier, F.P., Brais, B., Bulman, D.E., Warman Chardon, J., Chitayat, D., Deladoëy, J., Fernandez, B.A., Frosk, P., Geraghty, M.T., Gerull, B., Gibson, W., Gow, R.M., Graham, G.E., Green, J.S., Heon, E., Horvath, G., Innes, A.M., Jabado, N., Kim, R.H., Koeneke, R.K., Khan, A., Lehmann, O.J., Mendoza-Londono, R., Michaud, J.L., Nikkel, S.M., Penney, L.S., Polychronakos, C., Richer, J., Rouleau, G.A., Samuels, M.E., Siu, V.M., Suchowersky, O., Tarnopolsky, M.A., Yoon, G., Zahir, F.R., Majewski, J., Boycott, K.M., 2016. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* 89, 275–284. <https://doi.org/10.1111/cge.12654>
- Schwarz, J.M., Rödelberger, C., Schuelke, M., Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. <https://doi.org/10.1038/nmeth0810-575>
- Seki, T., Hayashi, N., Nishimoto, T., 1996. RCC1 in the Ran pathway. *J Biochem* 120, 207–214. <https://doi.org/10.1093/oxfordjournals.jbchem.a021400>
- Setty, S.T., Scott-Boyer, M.-P., Cuppens, T., Droit, A., 2022. New Developments and Possibilities in Reanalysis and Reinterpretation of Whole Exome Sequencing Datasets for Unsolved Rare Diseases Using Machine Learning Approaches. *Int J Mol Sci* 23, 6792. <https://doi.org/10.3390/ijms23126792>
- Shamseldin, H.E., Makhseed, N., Ibrahim, N., Al-Sheddi, T., Alobeid, E., Abdulwahab, F., Alkuraya, F.S., 2019. NUP214 deficiency causes severe encephalopathy and microcephaly in humans. *Hum Genet* 138, 221–229. <https://doi.org/10.1007/s00439-019-01979-w>
- Shihab, H.A., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M., Gaunt, T.R., 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum. Genomics* 8, 11. <https://doi.org/10.1186/1479-7364-8-11>
- Shihab, H.A., Rogers, M.F., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M., Gaunt, T.R., Campbell, C., 2015. An integrative approach to predicting the functional effects of



- non-coding and coding sequence variation. *Bioinformatics* 31, 1536–1543.  
<https://doi.org/10.1093/bioinformatics/btv009>
- Shinkai, S., Onami, S., Nakato, R., 2020. Toward understanding the dynamic state of 3D genome. *Comput Struct Biotechnol J* 18, 2259–2269.  
<https://doi.org/10.1016/j.csbj.2020.08.014>
- Shinohara, M., Saitoh, M., Takanashi, J., Yamanouchi, H., Kubota, M., Goto, T., Kikuchi, M., Shiihara, T., Yamanaka, G., Mizuguchi, M., 2011. Carnitine palmitoyl transferase II polymorphism is associated with multiple syndromes of acute encephalopathy with various infectious diseases. *Brain Dev* 33, 512–517.  
<https://doi.org/10.1016/j.braindev.2010.09.002>
- Singh, R.S., 2021. Decoding “Unnecessary Complexity”: A Law of Complexity and a Concept of Hidden Variation Behind ‘Missing Heritability’ in Precision Medicine. *J Mol Evol* 89, 513–526. <https://doi.org/10.1007/s00239-021-10023-3>
- Smith, M.J., Beetz, C., Williams, S.G., Bhaskar, S.S., O’Sullivan, J., Anderson, B., Daly, S.B., Urquhart, J.E., Bholah, Z., Oudit, D., Cheesman, E., Kelsey, A., McCabe, M.G., Newman, W.G., Evans, D.G.R., 2014. Germline mutations in SUFU cause Gorlin syndrome-associated childhood medulloblastoma and redefine the risk associated with PTCH1 mutations. *J Clin Oncol* 32, 4155–4161.  
<https://doi.org/10.1200/JCO.2014.58.2569>
- Snape, K., Wedderburn, S., Barwell, J., 2019. The new genomic medicine service and implications for patients. *Clin Med (Lond)* 19, 273–277.  
<https://doi.org/10.7861/clinmedicine.19-4-273>
- Sobczyńska-Tomaszewska, A., Ołtarzewski, M., Czerska, K., Wertheim-Tysarowska, K., Sands, D., Walkowiak, J., Bal, J., Mazurczak, T., NBS CF working group, 2013. Newborn screening for cystic fibrosis: Polish 4 years’ experience with CFTR sequencing strategy. *Eur J Hum Genet* 21, 391–396. <https://doi.org/10.1038/ejhg.2012.180>
- Spector, J.D., Wiita, A.P., 2019. ClinTAD: a tool for copy number variant interpretation in the context of topologically associated domains. *J Hum Genet* 64, 437–443.  
<https://doi.org/10.1038/s10038-019-0573-9>
- Stankiewicz, P., Lupski, J.R., 2010. Structural variation in the human genome and its role in disease. *Annu Rev Med* 61, 437–455. <https://doi.org/10.1146/annurev-med-100708-204735>
- Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., Fowler, D.M., 2017. Variant Interpretation: Functional Assays to the Rescue. *Am J Hum Genet* 101, 315–325. <https://doi.org/10.1016/j.ajhg.2017.07.014>
- Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shaw, K., Cooper, D.N., 2012. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinformatics* Chapter 1, Unit1.13.  
<https://doi.org/10.1002/0471250953.bi0113s39>
- Stoeger, T., Gerlach, M., Morimoto, R.I., Nunes Amaral, L.A., 2018. Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol* 16, e2006643.  
<https://doi.org/10.1371/journal.pbio.2006643>
- Sullivan, P.J., Gayevskiy, V., Davis, R.L., Wong, M., Mayoh, C., Mallawaarachchi, A., Hort, Y., McCabe, M.J., Beecroft, S., Jackson, M.R., Arts, P., Dubowsky, A., Laing, N., Dinger, M.E., Scott, H.S., Oates, E., Pinese, M., Cowley, M.J., 2023. Introne accurately predicts the impact of coding and non-coding variants on gene splicing, with clinical applications. *Genome Biol* 24, 118. <https://doi.org/10.1186/s13059-023-02936-7>

- Tanaka, A.J., Cho, M.T., Willaert, R., Retterer, K., Zarate, Y.A., Bosanko, K., Stefans, V., Oishi, K., Williamson, A., Wilson, G.N., Basinger, A., Barbaro-Dieber, T., Ortega, L., Sorrentino, S., Gabriel, M.K., Anderson, I.J., Sacoto, M.J.G., Schnur, R.E., Chung, W.K., 2017. De novo variants in EBF3 are associated with hypotonia, developmental delay, intellectual disability, and autism. *Cold Spring Harb Mol Case Stud* 3. <https://doi.org/10.1101/mcs.a002097>
- Tavtigian, S.V., Deffenbaugh, A.M., Yin, L., Judkins, T., Scholl, T., Samollow, P.B., de Silva, D., Zharkikh, A., Thomas, A., 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 43, 295–305. <https://doi.org/10.1136/jmg.2005.033878>
- Tavtigian, S.V., Greenblatt, M.S., Harrison, S.M., Nussbaum, R.L., Prabhu, S.A., Boucher, K.M., Biesecker, L.G., ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI), 2018. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 20, 1054–1060. <https://doi.org/10.1038/gim.2017.210>
- Turnbull, C., Scott, R.H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F.B., Halai, D., Baple, E., Craig, C., Hamblin, A., Henderson, S., Patch, C., O'Neill, A., Devereau, A., Smith, K., Martin, A.R., Sosinsky, A., McDonagh, E.M., Sultana, R., Mueller, M., Smedley, D., Toms, A., Dinh, L., Fowler, T., Bale, M., Hubbard, T., Rendon, A., Hill, S., Caulfield, M.J., 100 000 Genomes Project, 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* 361, k1687. <https://doi.org/10.1136/bmj.k1687>
- van Weerd, J.H., Mohan, R.A., van Duijvenboden, K., Hooijkaas, I.B., Wakker, V., Boukens, B.J., Barnett, P., Christoffels, V.M., 2020. Trait-associated non-coding variant regions affect TBX3 regulation and cardiac conduction. *eLife* 9, e56697. <https://doi.org/10.7554/eLife.56697>
- Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., Ng, P.C., 2016. SIFT missense predictions for genomes. *Nat Protoc* 11, 1–9. <https://doi.org/10.1038/nprot.2015.123>
- Vellarikkal, S.K., Jayarajan, R., Verma, A., Ravi, R., Senthilvel, V., Kumar, A., Saini, L., Gulati, S., Lal, M., Mathur, A., Chhetri, M.K., Faruq, M., Scaria, V., Sivasubbu, S., 2018. A founder mutation MLC1 c.736delA associated with megalencephalic leukoencephalopathy with subcortical cysts-1 in north Indian kindred. *Clin. Genet.* 94, 271–273. <https://doi.org/10.1111/cge.13251>
- Vellarikkal, S.K., Patowary, A., Singh, M., Kumari, R., Faruq, M., Master, D.C., Sivasubbu, S., Scaria, V., 2014. Exome sequencing reveals a novel mutation, p.L325H, in the KRT5 gene associated with autosomal dominant Epidermolysis Bullosa Simplex Koebner type in a large family from western India. *Hum Genome Var* 1, 14007. <https://doi.org/10.1038/hgv.2014.7>
- Veltman, J.A., Brunner, H.G., 2012. De novo mutations in human genetic disease. *Nat. Rev. Genet.* 13, 565–575. <https://doi.org/10.1038/nrg3241>
- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 101, 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Wangler, M.F., Yamamoto, S., Chao, H.-T., Posey, J.E., Westerfield, M., Postlethwait, J., Members of the Undiagnosed Diseases Network (UDN), Hieter, P., Boycott, K.M., Campeau, P.M., Bellen, H.J., 2017. Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics* 207, 9–27. <https://doi.org/10.1534/genetics.117.203067>

- Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., Au, K.F., 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* 6, 100. <https://doi.org/10.12688/f1000research.10571.2>
- Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. <https://doi.org/10.1038/nature05911>
- Woimant, F., Poujois, A., Bloch, A., Jordi, T., Laplanche, J.-L., Morel, H., Collet, C., 2020. A novel deep intronic variant in ATP7B in five unrelated families affected by Wilson disease. *Mol Genet Genomic Med* 8, e1428. <https://doi.org/10.1002/mgg3.1428>
- Wojcik, M.H., Reimers, R., Poorvu, T., Agrawal, P.B., 2020. Genetic diagnosis in the fetus. *J Perinatol* 40, 997–1006. <https://doi.org/10.1038/s41372-020-0627-z>
- Wood, K.A., Ellingford, J.M., Thomas, H.B., Genomics UK Research Consortium, Douzgou, S., Beaman, G.M., Hobson, E., Prescott, K., O’Keefe, R.T., Newman, W.G., 2022. Expanding the genotypic spectrum of TXNL4A variants in Burn-McKeown syndrome. *Clin Genet* 101, 255–259. <https://doi.org/10.1111/cge.14082>
- Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T., Bevan, A.P., Bragin, E., Chatzimichali, E.A., Gribble, S., Jones, P., Krishnappa, N., Mason, L.E., Miller, R., Morley, K.I., Parthiban, V., Prigmore, E., Rajan, D., Sifrim, A., Swaminathan, G.J., Tivey, A.R., Middleton, A., Parker, M., Carter, N.P., Barrett, J.C., Hurles, M.E., FitzPatrick, D.R., Firth, H.V., DDD study, 2015. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 385, 1305–1314. [https://doi.org/10.1016/S0140-6736\(14\)61705-0](https://doi.org/10.1016/S0140-6736(14)61705-0)
- Wright, C.F., Campbell, P., Eberhardt, R.Y., Aitken, S., Perrett, D., Brent, S., Danecek, P., Gardner, E.J., Chundru, V.K., Lindsay, S.J., Andrews, K., Hampstead, J., Kaplanis, J., Samocha, K.E., Middleton, A., Foreman, J., Hobson, R.J., Parker, M.J., Martin, H.C., FitzPatrick, D.R., Hurles, M.E., Firth, H.V., DDD Study, 2023. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *N Engl J Med* 388, 1559–1571. <https://doi.org/10.1056/NEJMoa2209046>
- Xu, J., Xue, C., Wang, X., Zhang, L., Mei, C., Mao, Z., 2022. Chromatin Methylation Abnormalities in Autosomal Dominant Polycystic Kidney Disease. *Front Med (Lausanne)* 9, 921631. <https://doi.org/10.3389/fmed.2022.921631>
- Yamamura, T., Horinouchi, T., Aoto, Y., Lennon, R., Nozu, K., 2022. The Contribution of COL4A5 Splicing Variants to the Pathogenesis of X-Linked Alport Syndrome. *Front Med (Lausanne)* 9, 841391. <https://doi.org/10.3389/fmed.2022.841391>
- Yang, H., Wang, K., 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* 10, 1556–1566. <https://doi.org/10.1038/nprot.2015.105>
- Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., Veeraraghavan, N., Hawes, A., Chiang, T., Leduc, M., Beuten, J., Zhang, J., He, W., Scull, J., Willis, A., Landsverk, M., Craigen, W.J., Bekheirnia, M.R., Stray-Pedersen, A., Liu, P., Wen, S., Alcaraz, W., Cui, H., Walkiewicz, M., Reid, J., Bainbridge, M., Patel, A., Boerwinkle, E., Beaudet, A.L., Lupski, J.R., Plon, S.E., Gibbs, R.A., Eng, C.M., 2014. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312, 1870–1879. <https://doi.org/10.1001/jama.2014.14601>

- Yasmeen, S., Lund, K., De Paepe, A., De Bie, S., Heiberg, A., Silva, J., Martins, M., Skjørringe, T., Møller, L.B., 2014. Occipital horn syndrome and classical Menkes Syndrome caused by deep intronic mutations, leading to the activation of ATP7A pseudo-exon. *Eur J Hum Genet* 22, 517–521. <https://doi.org/10.1038/ejhg.2013.191>
- Yauy, K., Lecoquierre, F., Baert-Desurmont, S., Trost, D., Boughalem, A., Luscan, A., Costa, J.-M., Geromel, V., Raymond, L., Richard, P., Coutant, S., Broutin, M., Lanos, R., Fort, Q., Cackowski, S., Testard, Q., Diallo, A., Soirat, N., Holder, J.-M., Duforet-Frebours, N., Bouge, A.-L., Beaumeunier, S., Bertrand, D., Audoux, J., Genevieve, D., Mesnard, L., Nicolas, G., Thevenon, J., Philippe, N., 2022. Genome Alert!: A standardized procedure for genomic variant reinterpretation and automated gene-phenotype reassessment in clinical routine. *Genet Med* 24, 1316–1327. <https://doi.org/10.1016/j.gim.2022.02.008>
- Yenamandra, V.K., Shamsudheen, K.V., Madhumita, R.C., Rijith, J., Ankit, V., Scaria, V., Sridhar, S., Kabra, M., Sharma, V.K., Sethuraman, G., 2017. Autosomal recessive epidermolysis bullosa simplex: report of three cases from India. *Clin. Exp. Dermatol.* 42, 800–803. <https://doi.org/10.1111/ced.13182>
- Yokoyama, N., Hayashi, N., Seki, T., Panté, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., Aebi, U., 1995. A giant nucleopore protein that binds Ran/TC4. *Nature* 376, 184–188. <https://doi.org/10.1038/376184a0>
- Zaghloul, N.A., Katsanis, N., 2010. Functional modules, mutational load and human genetic disease. *Trends Genet* 26, 168–176. <https://doi.org/10.1016/j.tig.2010.01.006>
- Zhang, G., Pradhan, S., 2014. Mammalian epigenetic mechanisms. *IUBMB Life* 66, 240–256. <https://doi.org/10.1002/iub.1264>
- Zhang, F., Lupski, J.R., 2015. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 24, R102–110. <https://doi.org/10.1093/hmg/ddv259>
- Zhang, J., Li, J., Saucier, J.B., Feng, Y., Jiang, Y., Sinson, J., McCombs, A.K., Schmitt, E.S., Peacock, S., Chen, S., Dai, H., Ge, X., Wang, G., Shaw, C.A., Mei, H., Breman, A., Xia, F., Yang, Y., Purgason, A., Pourpak, A., Chen, Z., Wang, X., Wang, Y., Kulkarni, S., Choy, K.W., Wapner, R.J., Van den Veyver, I.B., Beaudet, A., Parmar, S., Wong, L.-J., Eng, C.M., 2019. Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nat Med* 25, 439–447. <https://doi.org/10.1038/s41591-018-0334-x>
- Zhu, Y., Hu, L., Cao, D., Ou, X., Jiang, M., 2020. Chromosomal microarray analysis of infertile men with azoospermia factor microdeletions. *Gene* 735, 144389. <https://doi.org/10.1016/j.gene.2020.144389>
- Zurek, B., Ellwanger, K., Vissers, L.E.L.M., Schüle, R., Synofzik, M., Töpf, A., de Voer, R.M., Laurie, S., Matalonga, L., Gilissen, C., Ossowski, S., 't Hoen, P.A.C., Vitobello, A., Schulze-Hentrich, J.M., Riess, O., Brunner, H.G., Brookes, A.J., Rath, A., Bonne, G., Gumus, G., Verloes, A., Hoogerbrugge, N., Evangelista, T., Harmuth, T., Swertz, M., Spalding, D., Hoischen, A., Beltran, S., Graessner, H., Solve-RD consortium, 2021. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. *Eur J Hum Genet* 29, 1325–1331. <https://doi.org/10.1038/s41431-021-00859-0>

## Appendices

## Appendix I: Supplementary data for Chapter 3

### Early B-cell Factor 3-Related Genetic Disease Can Mimic Urofacial Syndrome.

#### Supporting Information.

System	Phenotype
Craniofacial	Synophrys, bushy strait eyebrows, long eyelashes, high arched palate, deep philtrum, downturned corner of mouth, broad chin with midline cleft, hirsutism.
Eyes	Alternating squint from early childhood, myopia.
Urogenital	Ballotable kidneys during infancy, neurogenic bladder, left grade 4 vesicoureteral reflux (VUR), right grade 5 VUR, bilateral tortuous hydroureters, bilateral grossly hydronephrotic kidneys, recurrent UTIs.
Skeletal	Tapering digits with thick palmar and solar skin, 2-3 toes syndactyly, unfused L5.
Gastrointestinal	Bowel obstruction from young age, proscribed lactulose until age 12, constipation is since resolved.
Developmental	Hypotonia and weak cry Age 3: Unable to walk independently, speech limited to two words Age 6: Independent but unsteady walking, impaired motor function. Able to perform adequate self-care and read and write simple sentences.

Supplementary Table I-1: Patient developmental abnormalities.

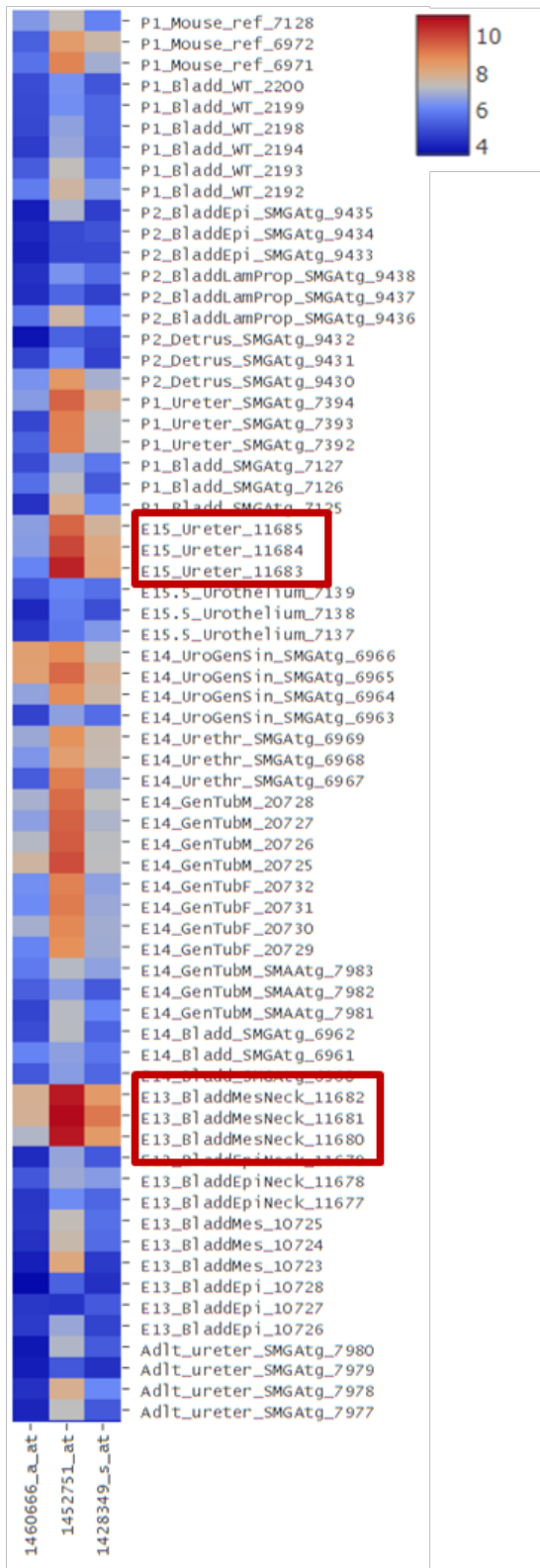
Tool	Prediction
SIFT	Damaging
PolyPhen2	Probably damaging
Mutation Taster	Disease Causing
Mutation Assessor	Predicted functional
FATHMM	Tolerated
FATHMM-MKL	Damaging
CADD Phred score	35 (Pathogenic cut-off >20)
REVEL	0.724 (Pathogenic cut-off >0.6)

Supplementary table I-2: In silico predictions for EBF3 variant c.626G>A.

Patients (n)	cDNA and <i>protein</i> coding variants	EBF3 Exon	RSID	gnomAD allele count/number	gnomAD allele Frequency
N=5	c.229A>C <i>p.Arg77Arg</i>	2	rs75074888	16766/265108	0.0632
N=1	c.639T>G <i>p.Val213Val</i>	8	rs141973685	149/234652	0.000635
N=1	c.134+40_134+45dup <i>N/A</i>	Intron 1	rs767692640	215/55380	0.003882

Supplementary table I-3: Variants detected in EBF3 targeted exome sequencing of VUR cohort (n=80). RSID (reference SNP cluster ID).

# Lower Urinary Tract (MOE430)



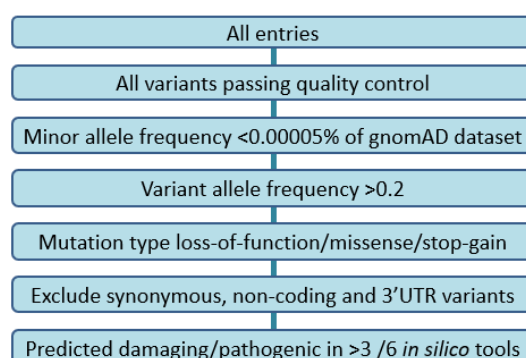
Supplementary figure I-1: Screenshot of microarray data for EBF3 in the developing murine lower urinary tract. Sourced from GUDMAP<sup>S1,S2</sup>. Red cells indicate higher expression and blue cells indicate lower expression. Red boxes indicate high expression of EBF3 in ureter at embryonic stage E15, and in the bladder mesenchyme neck at E13.

## Methods:

### Supplementary Methods

#### Whole exome sequencing (WES):

DNA samples were prepared and sent to Beijing Genomics Institution (BGI) for WGS. Sequencing was performed using a BGI exome kit, version 4 (59M) 6G BGI-Seq500. Exome read data were aligned to the genomic reference sequence (hg19) by the in-house bioinformatics team, Genetic Medicine, St Mary's Hospital. Analysis and annotation of genome data was performed in-house using VarSeq™ v2.2 (Golden Helix, Inc., Bozeman, MT). During analysis, variants were filtered for Individual 1 and 2 as stated in Figure 2.



Supplementary figure I-2: Filtering strategy used for whole exome sequencing data analysis. Resulting variants were considered based on their association with clinical data.

#### Variant analysis:

Variants of interest were identified using entries from OMIM (<https://www.omim.org/>), GeneCards (<https://www.genecards.org/>) and PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>).

#### *In silico* prediction tools:

Six *in silico* prediction tools were used to analyse mutation pathogenicity: Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Mutation taster (<http://www.mutationtaster.org/>), MutationAssessor (<http://mutationassessor.org/r3/>), SIFT (Sorting Intolerant From Tolerant) (<http://sift.bii.a-star.edu.sg/>), FATHMM (Functional Analysis Through Hidden Markov Models) (<http://fathmm.biocompute.org.uk/inherited.html>), and FATHMM-MKL (Math Kernel Library) (<http://fathmm.biocompute.org.uk/fathmmMKL.htm>).

#### Sanger sequencing:

PCR reactions were set up according to the volumes of patient, parent, or control DNA stated in Table 2. Primers were all designed and optimised in house. Reactions were carried out using the Veriti 96-Well Thermal Cycler (applied Biosystems) according to standard reaction conditions. PCR product was purified by addition of Agencourt AMPure XP beads and submitted to the NHS diagnostic laboratory at St Mary's Hospital. Purified product was prepared for Sanger sequencing (Table 3), according to standard conditions. Sequencing products were purified by addition of Agencourt CleanSEQ beads, and submitted to the NHS diagnostic laboratory, St Mary's Hospital for purification and sequencing using the automated Applied Biosystems 3730 XL DNA Analyzer.



sequencer. Analysis was performed using the Staden Software package. Sequencing files for patient, parent and control DNA were added to databases created using PreGap4 software, with control DNA processed as the reference sequence. Visualisation of aligned traces allowed for comparative analysis between patient and control or patient and parent sequences.

Reagent	Volume (μl)
GoTaq Green Mix	12.5
Forward and Reverse primer mix	2.5
Sterile Water	7.5
DNA	2.5
Total	25

Supplementary table I-4: PCR reaction components.

Reagent	Volume (μl)
PCR product	1.2
Forward or Reverse primer (2μM)	1.6
Sterile Water	1.8
Sequencing buffer	0.2
Sequencing kit v3.1	5.2
Total	10

Supplementary table I-5: Sequencing reaction components.

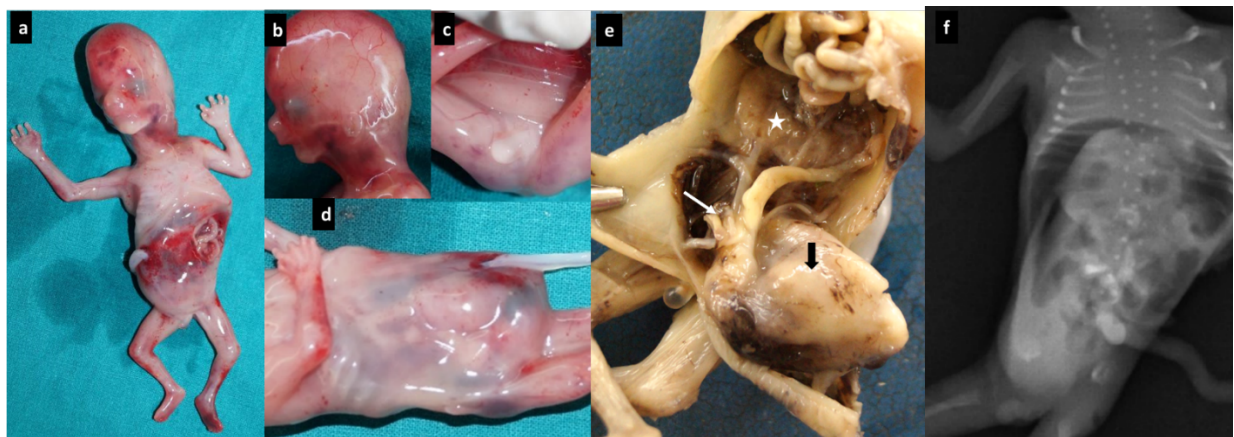
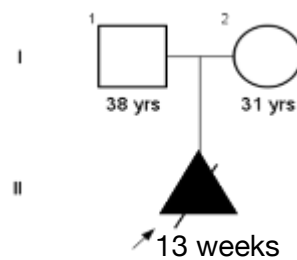
## Appendix II: Supplementary data for chapter 4

### A clinically diverse spectrum of perinatal lethal phenotypes with urorectal septum malformation sequence

Supporting information

#### Fetus 1 (Complete urorectal septum malformation sequence)

Pedigree:

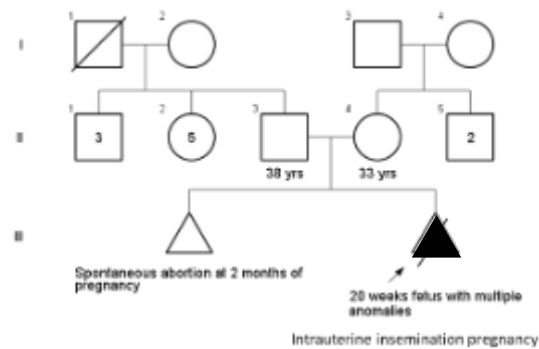


**Supplementary figure II-1:** Clinical images of fetus 1. Images show short nose with depressed nasal root, anteverted nares, long philtrum and retrognathia (a, b), protuberant abdomen with thin and transparent wall (a, d), ambiguous external genitalia with imperforate anus (c), multi-cystic and fused kidneys (white asterix) with narrow ureters, fallopian tube (white arrow) and hind gut draining into blind cloaca (black arrow, e) and absent ossification of sacrum (f).

**Clinical findings:** The fetus was examined at 13 weeks of gestation in view of megacystis on antenatal ultrasonography. Fetus weighed 31 g (normal), measured 12 cm in length (normal) with head circumference of 8.5 cm (normal). Fetus had short nose with depressed nasal root, anteverted nares, long philtrum and retrognathia. Protuberant abdomen was observed with thin and transparent abdominal wall. Ambiguous external genitalia with a small phallus-like structure with no perineal openings were noted. Multi-cystic and fused kidneys were observed with ureters running anterior to the fused part of kidney. The ureters, fallopian tubes and hind gut were connected to the enlarged and persistent blind cloaca (Supplementary figure 1). Radiographs of the fetus showed absence of ossification centers for sacrum.

**Fetus 5 (Radial ray defect, unilateral renal agenesis, urorectal septum malformation sequence and ventricular septal defect)**

**Pedigree:**



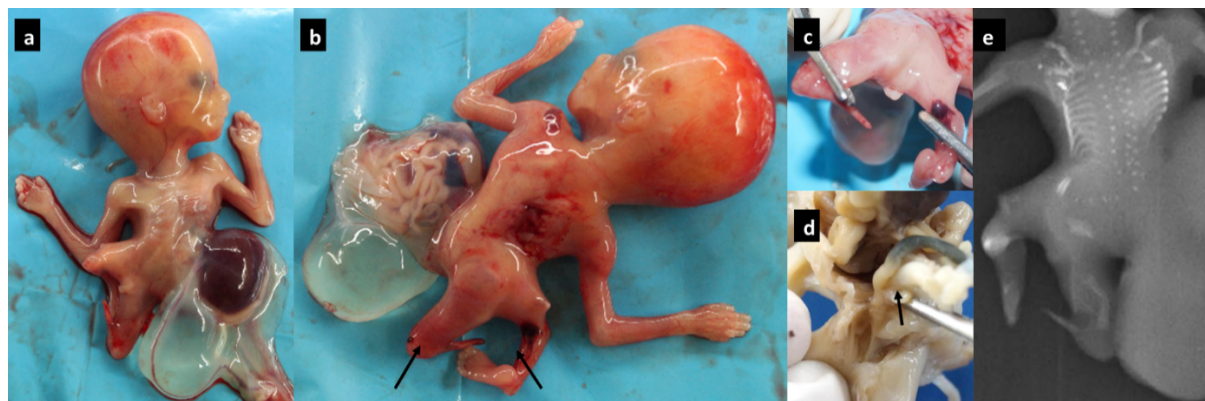
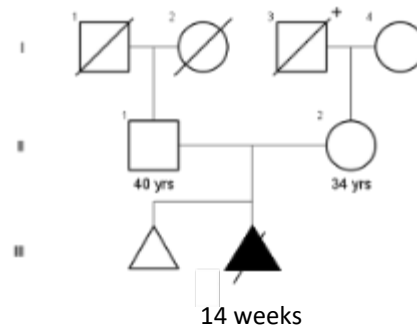
**Supplementary figure II-2:** Clinical images of fetus 5. We observed facial asymmetry with hypertelorism, short nose, anteverted nares, long philtrum, retrognathia (a), right upper limb with small radius and rudimentary thumb (b, c), left upper limb with complete absence ray 1 and 2 with absent radius (d, e), ambiguous genitalia (f), unilateral renal agenesis with blind ending cloaca (asterisk, g) and ventricular septal defect (arrow, h) in fetus 5.

**Clinical features:** Abortion in a second gravida at 20 weeks of gestation with hypoplastic nasal bone, upper limb defects, small ventricular septal defect, enlarged, echogenic and cystic right kidney in fetus on ultrasonography. Fetal anthropometry was within normal limits. Fetus had facial asymmetry, hypertelorism, short nose, anteverted nares, long philtrum, retrognathia, bilateral radial deviation of hands with bowed forearms, rudimentary right thumb and absence of two digits (ray defect - F1 and F2) on left hand. Female external genitalia was noted with no perineal openings and presence of single umbilical artery.

Membranous ventricular septal defect, unilateral (left) renal agenesis with absent left ureter, normal right kidney, ?ovaries, no uterine tissue, blind-ending persistent cloaca with right ureter and hindgut draining into it were documented on visceral examination. Infantogram showed absent radius on left side and rudimentary radius on right side, absence of metacarpals and phalanges of first and second digits on left hand and absent metacarpal and phalanges of right thumb (Supplementary figure 2).

#### **Fetus 6 (Thoraco-abdominoschisis, limb defects and urorectal septum malformation sequence)**

**Pedigree:**



**Supplementary figure II-3:** Clinical images of fetus 6. 14 weeks fetus (6) had anterior lower thoracic and abdominal wall defect with herniation of liver and intestine, kypho-scoliosis (a, b). Short right lower limb with contractures and reduction defect of left lower limb below the knee (arrows), ambiguous genitalia with absent perineal openings (b, c), absent left kidney and adrenal with blind ending persistent cloaca (arrow, d) were also noted. Radiographs show kypho-scoliosis at thoraco-lumbar region and absence of ossification of lower lumbar and sacral vertebrae and bent left femur (arrow) and underdeveloped long bones of right lower limb (e).

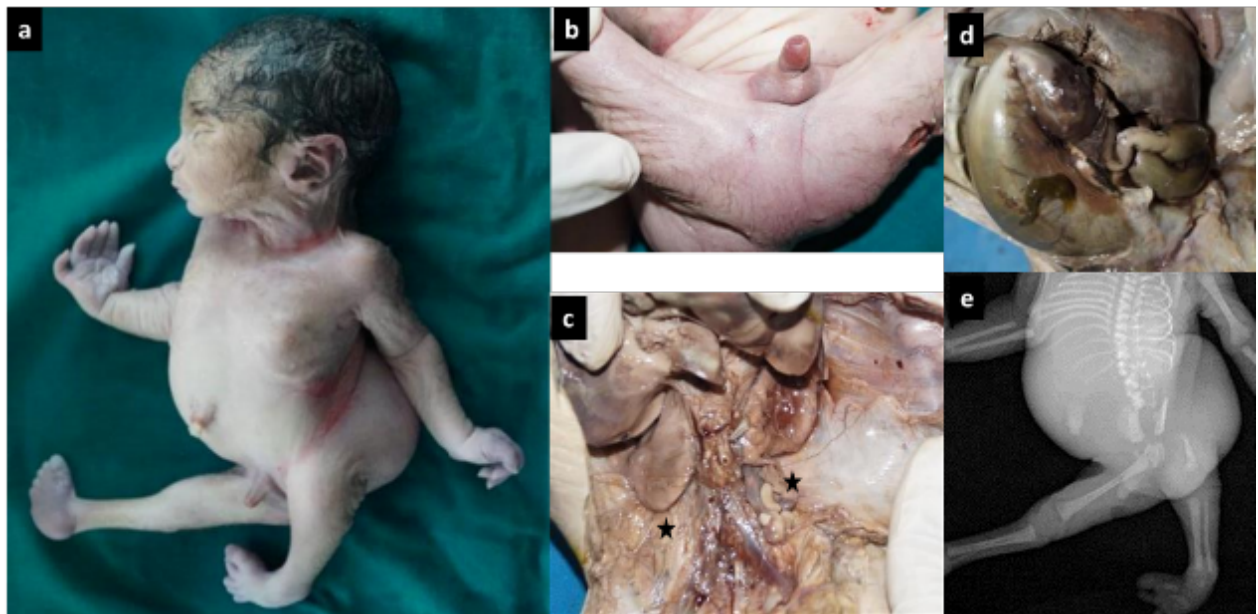
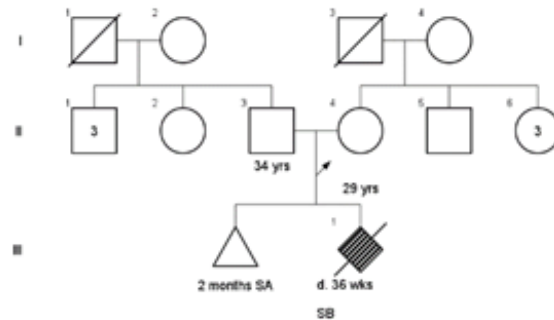
**Clinical features:** The pregnancy was interrupted at 14 weeks of gestation with anterior abdominal wall defect and herniation of liver in fetus. On evaluation, fetus measured 8.4 cm (-2.9 SD) in total length with anterior lower thoracic and abdominal wall defect, herniation of stomach, spleen, pancreas, liver and loops of small intestine covered by peritoneum, kypho-scoliosis at thoraco-lumbar region, short right lower limb with joint contractures across hips, knees and ankles with bowing of thighs and legs, transverse deficiency of left lower limb below the knee with rudimentary stem like structure (Supplementary figure 3). There was imperforate anus with ambiguous external genitalia. On internal examination, we note dextrocardia with a defect in lower portion of sternum through which the portion of left lung and apex of heart protruded into the abdominal cavity with absent left dome of diaphragm. Agenesis of left kidney and adrenal gland was noted. There was urorectal septum



malformation sequence with persistent cloaca. The colon and right ureter drained into cloaca with atresia of cloacal outlet. Indifferent gonads, kypho-scoliosis at thoraco-lumbar region, absence of ossification of lower lumbar and sacral vertebrae, bowed left femur, short and bowed right femur, short right tibia and absence of right fibula were observed.

**Fetus 7 (Myelomeningocele, urorectal septum malformation sequence, bilateral renal agenesis and vertebral segmentation defect)**

**Pedigree:**



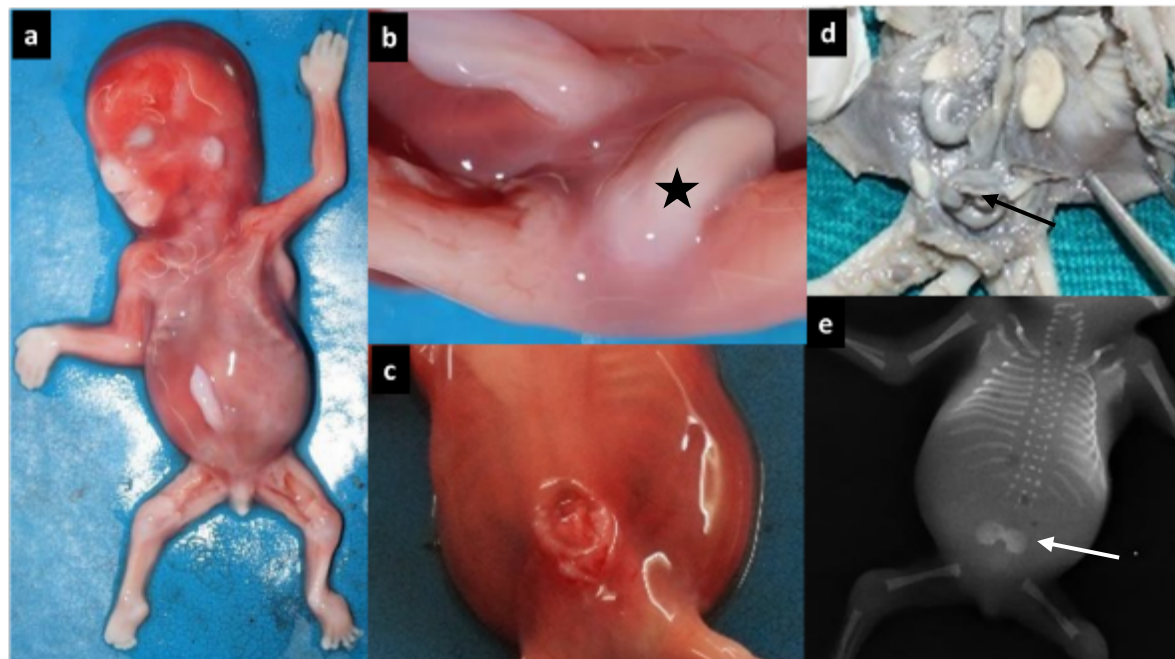
**Supplementary figure II-4:** Clinical images of fetus 7. Evaluation of fetus 7 revealed dysmorphic ears, short trunk, scoliosis, unilateral radial deviation of right hand, contractures across axillae, elbows, hips and knees joints, bilateral clubfeet and rhizomelia of left lower limb (a), ambiguous external genitalia with no perineal openings (b), bilateral renal agenesis (c, asterisks), meconium filled hindgut (d) and segmentation defects of thoracic and lumbar vertebrae, absent ossification of sacral vertebrae and short left femur (e).

**Clinical features:** A stillbirth was referred for detailed examination at 36 weeks of gestation following ultrasound findings of intrauterine growth retardation, large open neural defect at lower half of spine and anhydramnios. Anthropometry suggested [weighed 1131 g (-3SD to -4SD) measured 33 cm (-4SD) in length with head circumference of 27.5 cm (-2SD to -3SD)] fetal growth retardation. Dysmorphic

ears, short trunk, scoliosis at lumbar region, lipomeningomyelocele, unilateral radial club hand on right side, bilateral clubfeet and rhizomelia of left lower limb, bilateral contractures across axilla, elbow, hip and knee were evident in the fetus. Ambiguous external genitalia with phallus-like structure and absent perineal openings were observed.

There were pulmonary hypoplasia, bilateral renal agenesis with absent ureters, short small intestine, enlarged hind gut filled with feces draining into the blind cloaca with no intra-abdominal gonads in fetus (Supplementary figure 4). Segmentation defects of thoracic and lumbar vertebrae, absent ossification of sacral vertebrae, short left femur and herniation of dural sac at sacral region with spinal cord were reported on fetal imaging.

#### **Fetus 8 (Urorectal septal malformation sequence and neural tube defect)**

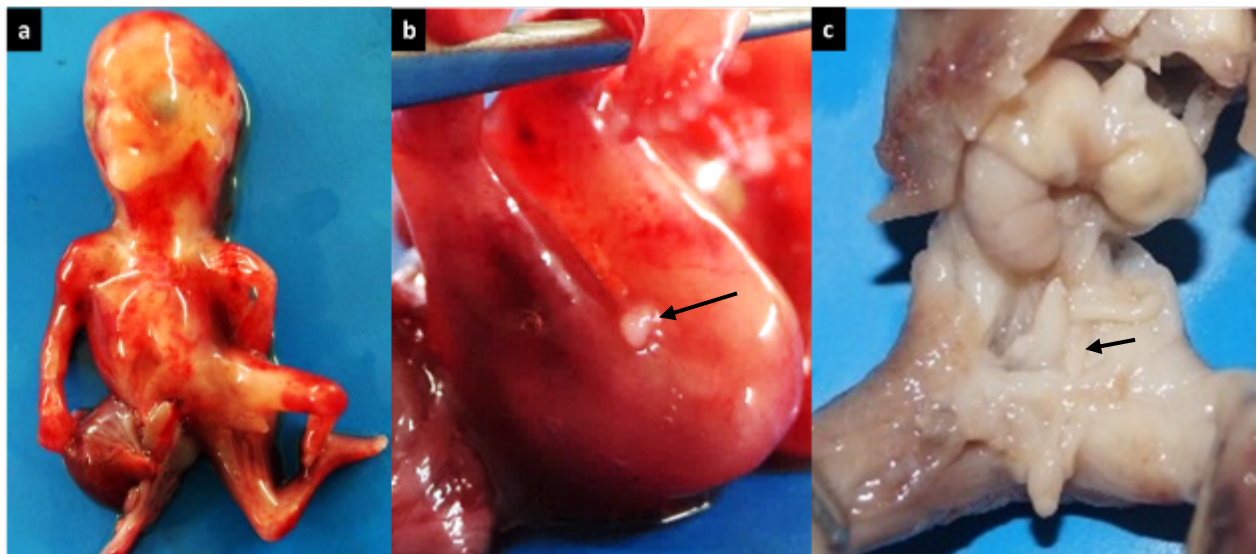
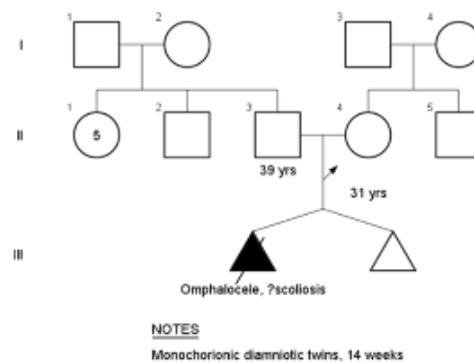


**Supplementary figure II-5:** Clinical images of fetus 8. Intrauterine demise of fetus 8 was noted with phallus-like structure (asterisk) with no perineal openings (a, b), spina bifida at lower lumbar region (c), left renal agenesis with blind ending cloaca (black arrow, d) and fused iliac bones (white arrow) and absence of ossification for lower lumbar vertebrae (e).

**Clinical features:** The fetus was examined following intrauterine demise and with an umbilical cord cyst. The fetus had open neural tube defect at lower lumbar region, ambiguous external genitalia with phallus-like structure and imperforate anus. Unilateral (left) renal agenesis was noted with right ureter and hindgut opening into blind cloaca. There were indifferent gonads. Radiographs of the fetus revealed fused iliac bones and absence of ossification of lower lumbar vertebrae (Supplementary figure 5).

**Fetus 9 (Omphalocele, partial urorectal septum malformation, imperforated anus and scoliosis)**

**Pedigree:**



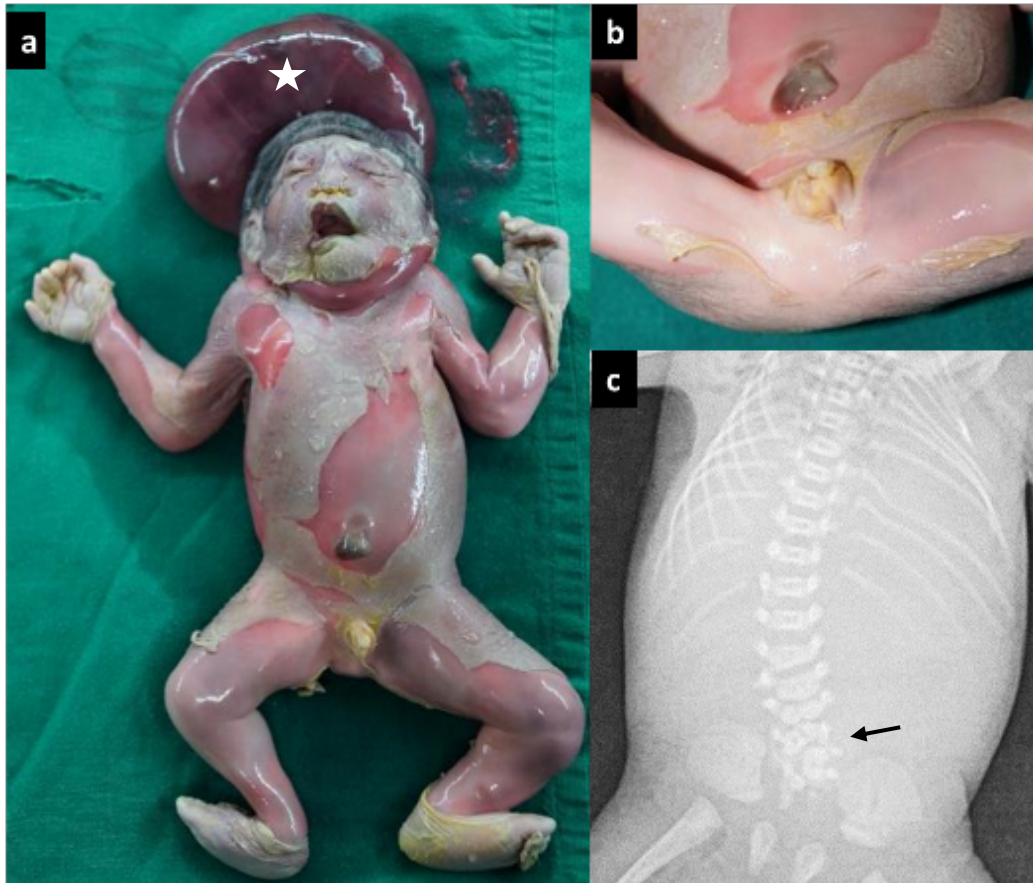
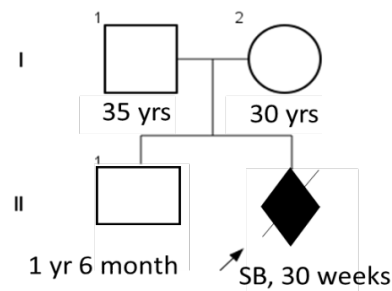
**Supplementary figure II-6:** Clinical images of fetus 9. Omphalocele, scoliosis (a), ambiguous external genitalia with no perineal openings (arrow, b) and partial separation of urogenital sinus and primitive rectum (arrow, c) were observed in fetus 9.

**Clinical features:** A twin from monochorionic diamniotic pregnancy weighed 20g (-2SD to -3SD) and measured 9 cm (-2SD to -3SD) in length at 14 weeks of gestation was evaluated in view of omphalocele and scoliosis. Postnatal examination of fetus confirmed omphalocele and scoliosis at lumbar region. In addition, ambiguous external genitalia with phallus-like structure, imperforate anus and unilateral club foot on left side were noted (Supplementary figure 6). There was partial separation of urogenital sinus and primitive rectum in fetus.



**Fetus 10 (Occipital encephalocele, renal agenesis and urorectal septum malformation sequence)**

**Pedigree:**



**Supplementary figure II-7:** Clinical images of fetus 10. Occipital encephalocele (asterisk, a), small phallus-like structure with absent perineal openings (a, b) and segmentation defects of lumbar vertebrae (arrow, c) were noted in 30 weeks' fetus (10).

**Clinical features:** Stillbirth was noted at 30 weeks of gestation. Imaging findings of occipital encephalocele, bilateral echogenic kidneys and anhydramnios were noted antenatally. The fetus weighed 718 g (-2SD to -3SD) measured 29 cm (-5SD) in total length with frontal sloping, flat nasal tip, micrognathia, dysmorphic ears, short neck and occipital encephalocele (Supplementary figure 7). Ambiguous external genitalia was noted with no perineal openings. Bilateral pulmonary hypoplasia was observed. There were bilateral renal agenesis with absent ureters, the rudimentary bladder and hindgut draining into a blind, persistent cloaca. Indistinct gonads were noted with persistent left mesonephric cord. Radiographs of the fetus showed segmentation defect in lower lumbar vertebrae and absence of ossification of sacrum.



## Appendix III: Supplementary data for Chapter 5

### Deep Intronic Variant Causes Aberrant Splicing Of *ATP7A* In A Family With A Variable Occipital Horn Syndrome Phenotype

Supplementary table III-1: Rare variants identified in exome sequencing of IV:1

Chr	Start	End	Gene	No times seen	Comments
X	334,257	334,388	<i>PPP2R3B</i>	8	
X	1,497,621	1,497,711	<i>IL3RA</i>	1	Pseudoautosomal region
X	2,418,315	2,418,627	<i>ZBED1</i>	17	
X	2,609,409	2,609,592	<i>CD99</i>	7	
X	3,772,984	3,773,085	<i>ENST00000456563</i>	8	In region
X	8,699,931	8,700,103	<i>KAL1</i>	14	
X	9,754,625	9,754,818	<i>SHROOM2</i>	5	
X	17,393,878	17,394,444	<i>NHS</i>	10	
X	19,533,285	19,533,411	<i>MAP3K15</i>	9	
X	20,206,470	20,206,590	<i>RPS6KA3</i>	5	
X	21,958,955	21,959,077	<i>SMS</i>	15	
X	21,959,261	21,959,551	<i>SMS</i>	5	
X	47,003,838	47,003,958	<i>NDUFB11</i>	12	
X	49,020,098	49,020,422	<i>MAGIX</i>	4	
X	49,772,917	49,773,037	<i>CLCN5</i>	1	Within small shared region (0.6 Mb); chloride channel mutated in Dent disease (progressive proximal renal tubulopathy with hypercalciuria, low-molecular-weight proteinuria, and nephrocalcinosis). Not an exon on Alamut
X	49,773,042	49,773,162	<i>CLCN5</i>	6	

X	55,513,621	55,514,979	<i>USP51</i>	17	In region
X	57,618,454	57,618,710	<i>ZXDB</i>	5	In region
X	70,586,296	70,586,416	<i>TAF1</i>	9	In region
X	100,268,662	100,268,753	<i>TRMT2B</i>	9	
X	101,093,015	101,093,203	<i>NXF5</i>	19	
X	101,581,338	101,581,463	<i>NXF2B</i>	11	
X	103,231,281	103,231,429	<i>H2BFXP</i>	15	
X	103,231,489	103,231,579	<i>H2BFXP</i>	14	
X	106,243,090	106,243,214	<i>MORC4</i>	15	
X	115,568,919	115,569,154	<i>SLC6A14</i>	6	
X	120,504,772	120,505,002	<i>hsa-mir-3672</i>	1	Not within a shared region
X	152,751,165	152,751,576	<i>HAUS7</i>	1	Within a small shared region (0.74 Mb). Part of the mitotic spindle assembly. Not an exon on Alamut
X	153,285,281	153,285,402	<i>ENST00000369980</i>	5	
X	153,520,375	153,520,470	<i>TEX28</i>	22	
X	153,872,142	153,872,280	<i>ENST00000453062</i>	7	

## **Appendix IV: Supplementary data and updates for Chapter 6**

### **Biallelic Variants in RCC1 result in fever-associated axonal neuropathy with encephalopathy**

The data presented in chapter 6 are the product of the initial stages of characterising the novel acute neuropathy associated with biallelic RCC1 variants described. Chapter 6 was prepared in a manuscript style to conform with other chapters in the thesis, so omitted a detailed account of the cloning and purification undertaken to express recombinant Rcc1 for in vitro stability and functional assays described in section 6.3.4 Figure 2C-E. Additionally, during the time of thesis submission, further investigations into the characterisation of Rcc1 involvement in acute neuropathy were being planned and undertaken. This appendix provides a contextual narrative to provide further supplementary data to describe the initial expression of recombinant Rcc1 (section IV-1) and update the current and future work (section IV-2).

#### **IV-1 Supplementary data for expression of recombinant Rcc1 and Ran**

Initial studies to characterise the function of Rcc1 were undertaken by expressing recombinant Rcc1 and Ran in a bacterial system in order to undertake in vitro studies. This approach required cloning of expression vectors for Ran and Rcc1, and the mutagenesis of the Rcc1 vector to produce a Rcc1 protein harbouring the p.(Gly43Ser) variant. Detailed methods are provided in sections 6.5.3-6.5.6.

Following expression of His-tagged Ran and Rcc1 proteins in Rosetta (2E3) protein expression bacteria for 36 hours at 37 °C, bacterial cells were lysed, and crude lysates were purified using purification column containing His-select agarose beads. To test purity of the His-tagged proteins following column purification, samples underwent SDS-PAGE separation using 12 % gels. Figure IV-1(A) and IV-1(B) indicate strong expression of Ran, and Rcc1 proteins. Due to the poor expression of Rcc1<sup>G43S</sup>, expression was repeated with cultures incubated at 19 °C to improve protein product solubility. Purification following adjustment of culture temperature (Figure IV-1(C)) indicated improved Rcc1<sup>G43S</sup> protein concentration.

Fractions 1-4 of each protein elution were dialysed overnight to remove imidazole present in the elution buffer. Protein concentrations were ascertained, and proteins were aliquoted and stored at -80 °C until use. Purified proteins were thawed on ice for use in thermal shift assays and GTPase assays.

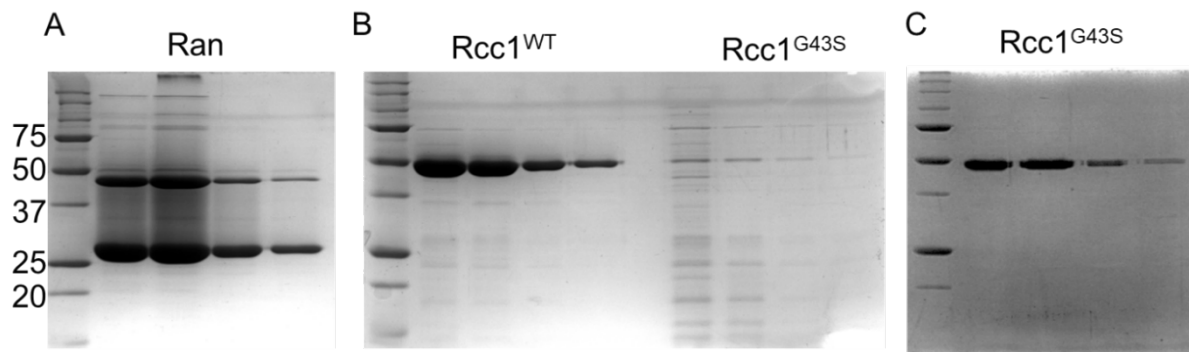


Figure IV-1: Assessment of purification of recombinant Ran and Rcc1 proteins. pET28a-Ran (A), pET28a-Rcc1-wt (B), and pET28a-Rcc1-G43S (B and C) vectors in Rosetta 2E3 bacteria. Bacterial cultures were grown for 24 hours at 37 °C (A,B) or at 19 °C (C). Molecular weights: Ran 25 kDa, Rcc1 45 kDa.

#### IV-2 Ongoing and future work

Following identification of families B and C (Section 6.3.3) from reverse phenotyping 100,000 genomes project participants, further families (D and E) affected by axonal neuropathy secondary to infection have been identified. Both families have a history of consanguinity. Two children from an Iranian family (Figure IV-2, Family D) were identified with a homozygous *RCC1*:c.280A>G;p.(Asn94Asp) variant. Sanger sequencing undertaken locally confirmed parents were heterozygous for this variant. Additionally, two children from a Turkish family (Figure IV-2, Family E) were identified following reassessment of clinical exome data locally. Both affected children from Family E were homozygous for *RCC1*:c.330G>A;p.(Met110Ile). Similarly, segregation analysis confirmed both parents as heterozygous.

Figure IC-2 indicates variants from all families cluster in exons 6 and 7 of *RCC1*, except the paternally inherited c.1195C>T variant in Family C located in exon 13. Assays are planned for

expression of recombinant forms of Rcc1 protein containing additional variants identified to extend the in vitro characterisation of Rcc1 stability and function.

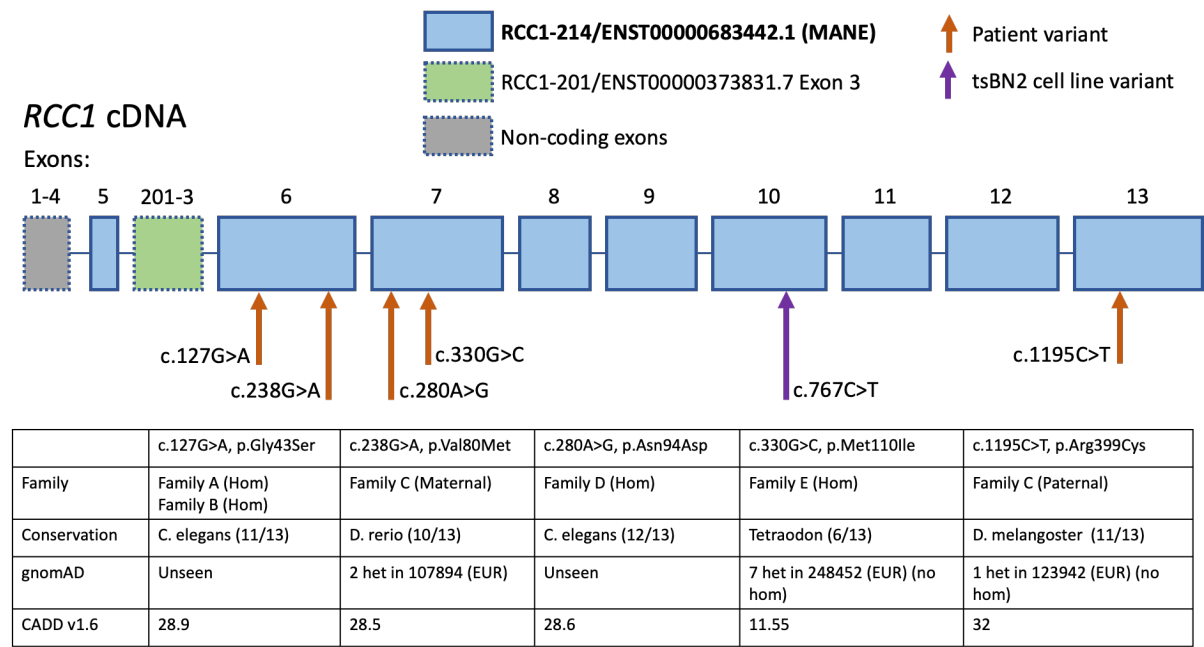


Figure IV-2: Diagram of *RCC1* cDNA variants identified in patients with fever-associated axonal neuropathy. Schematic shows *RCC1* MANE transcript (ENST00000683442.1) coding exons in blue and non-coding exons in grey. Coding exon for an alternative non-functional exon included in ENST00000373831.7 is shown in green. Orange arrows indicate variants identified in individuals with acute-onset axonal neuropathy. Purple arrow indicates homozygous missense c.767C>T;p.(Ser256Phe) present in the tsBN2 cell line. Conservation described the most distantly conserved model organism, and the number of species where the amino acid residue is conserved.

The *RCC1*:c.767C>T variant (Figure IV-2, purple arrow) previously identified in tsBN2 cells derived from hamster kidney is associated impaired nucleocytoplasmic transport following exposure to non-permissive temperature of 39.5 °C (Nishijima et al., 2000; Ohtsubo et al., 1987). In addition to high temperature, exposure to oxidants and ultraviolet light can induce defects in nucleocytoplasmic transport (Chatterjee and Paschal, 2015). Nucleocytoplasmic

transport defects can be rescued in tsBN2 cells by transfection with wildtype human *RCC1* (Chatterjee and Paschal, 2015; Chen et al., 2007).

This forms the basis for further characterisation of *Rcc1* variants. In addition to fibroblasts from family A, fibroblasts have been procured from families C and E. Currently, nuclear localisation of eGFP tagged with a nuclear localisation signal (eGFP-NLS) is being assessed in primary dermal fibroblasts from individuals with biallelic *RCC1* missense variants following exposure to heat shock, or H<sub>2</sub>O<sub>2</sub>. Assays are also being prepared to enable characterisation of tsBN2 cells transfected with wildtype *RCC1*, and variants identified in individuals affected by axonal neuropathy (c.127G>A, c.238G>A, c.280A>G, c.330G>C, and c.1195C>T). In this model, the hypothesis will be that variants identified in patients with axonal neuropathy will not permit rescue of the defects in nucleocytoplasmic transport.

Three additional models are also being considered to permit extension of molecular characterisation. Firstly, production of induced pluripotent stem cells (iPSCs) from primary patient fibroblasts would allow co-culture of motor neurons with Schwann cells to characterise the effect of *RCC1* variants on dysregulation of myelination identified in conditional motor neuron *Ranbp2*<sup>-/-</sup> mice (Cho et al., 2017). Secondly, vectors produced to undertake rescue experiments in tsBN2 cells can be repurposed for modelling in *Drosophila melanogaster*. Deletion of 15bp from *D. melanogaster RCC1* homolog *Bj1* results in lateral central nervous system differentiation defects (Shi and Skeath, 2004). Extending these studies with missense variants identified in patients will enable assessment of neurodevelopment and phenotyping following infection. Thirdly, production of a transgenic mouse model will permit immunophenotyping to assess immune response in mice with homozygous *Rcc1* variants. This will provide greater insight into the mechanisms precipitating the acute onset neuropathy associated with biallelic *RCC1* variants.

## References

- Chatterjee, M., Paschal, B.M., 2015. Disruption of the ran system by cysteine oxidation of the nucleotide exchange factor RCC1. *Mol Cell Biol* 35, 566–581. <https://doi.org/10.1128/MCB.01133-14>
- Chen, T., Muratore, T.L., Schaner-Tooley, C.E., Shabanowitz, J., Hunt, D.F., Macara, I.G., 2007. N-terminal alpha-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat Cell Biol* 9, 596–603. <https://doi.org/10.1038/ncb1572>
- Cho, K.-I., Yoon, D., Qiu, S., Danziger, Z., Grill, W.M., Wetsel, W.C., Ferreira, P.A., 2017. Loss of Ranbp2 in motoneurons causes disruption of nucleocytoplasmic and chemokine signaling, proteostasis of hnRNPH3 and Mmp28, and development of amyotrophic lateral sclerosis-like syndromes. *Dis Model Mech* 10, 559–579. <https://doi.org/10.1242/dmm.027730>
- Nishijima, H., Seki, T., Nishitani, H., Nishimoto, T., 2000. Premature chromatin condensation caused by loss of RCC1. *Prog Cell Cycle Res* 4, 145–156. [https://doi.org/10.1007/978-1-4615-4253-7\\_13](https://doi.org/10.1007/978-1-4615-4253-7_13)
- Ohtsubo, M., Kai, R., Furuno, N., Sekiguchi, T., Sekiguchi, M., Hayashida, H., Kuma, K., Miyata, T., Fukushige, S., Murotsu, T., 1987. Isolation and characterization of the active cDNA of the human cell cycle gene (RCC1) involved in the regulation of onset of chromosome condensation. *Genes Dev* 1, 585–593. <https://doi.org/10.1101/gad.1.6.585>
- Shi, W.-Y., Skeath, J.B., 2004. The *Drosophila* RCC1 homolog, Bj1, regulates nucleocytoplasmic transport and neural differentiation during *Drosophila* development. *Dev Biol* 270, 106–121. <https://doi.org/10.1016/j.ydbio.2004.02.011>

## Appendix V: Data from an additional investigation

### Characterising heterozygous *WARS1* missense variants in patients with variable neurodevelopmental phenotypes.

J. Robert Harkness, Leigh Demain, John McDermott, Raymond O’Keefe, Siddharth Banka

This appendix sets out additional investigations carried out relating to identification of heterozygous variants in *WARS1* among patients with variable phenotype consisting of either intellectual disability and developmental delay or peripheral neuropathy. The results presented here form part of larger body of ongoing work, including characterisation in model organisms (*Caenorhabditis elegans*, *Danio rerio*) being conducted with external collaborators. For adequate interpretation of variant effect on neurodevelopment, further data from animal models is required, but is not yet available.

My contribution to this work is the undertaking of aminoacylation and thermal shift experiments and associated data and statistical analysis. LD prepared constructs for and undertook expression of Wars1 proteins. JM and SB phenotype and identified patient variants. RO and SB guided the planning and supervision of this work.

#### Background

Tryptophanyl-tRNA synthetase 1 (Wars1, or TrpRS) encoded by the *WARS1* gene (MIM 191050) is the enzyme responsible for tryptophan-tRNA aminoacylation, allowing incorporation of tryptophan residues during protein synthesis (Shen et al., 2006). Heterozygous *WARS1* variants have previously been associated with a juvenile-onset, slowly progressive form of pure motor neuropathy termed distal hereditary motor neuropathy (dHMN) type IX (MIM 617721) (Li et al., 2019; Tsai et al., 2017; Wang et al., 2019). One patient with heterozygous missense variants was identified with a more heterogeneous phenotypic spectrum consisting of microcephaly, hypotonia, severe intellectual disability with delayed myelination, but absence of peripheral neuropathy (Okamoto et al., 2022).

Additionally, biallelic variants in *WARS1* have been identified in patients with various phenotypes. Lin et al., (2022) identified three patients with heterogeneous phenotypes. One patient had a syndromic neurodevelopmental phenotype, one patient had defects in craniofacial and skeletal development, sensorineural hearing loss, hypotonia, and walking/gait abnormalities, and a family was identified with multiple brain anomalies, microcephaly, hypotonia, adrenal insufficiency, osteopenia, and global developmental delay (Lin et al., 2022).



These cases indicate variants in *WARS1* can cause autosomal dominant or autosomal recessive disease, both of which are characterised by phenotype heterogeneity. The purpose of this study is to determine functional differences which result from heterozygous *WARS1* genetic variants which underly the phenotypic heterogeneity.

## Methods

### Protein purification

Rosetta 2 (DE3)-transformed bacteria were grown for 24 hours at 37°C in 500ml cultures containing Overnight Express media to ensure high level of protein expression. To purify the expressed proteins, cultures were centrifuged, and cells lysis initiated by storing cell pellets overnight at -80°C. Cells were thawed on ice and suspended in lysis/wash buffer (20mM Tris-Cl pH7.4, 150mM NaCl, 0.1M DTT, 0.02% Tween-20, 10mM imidazole, 15% glycerol). Cell lysis was completed by sonication 3x 10 second bursts at 30% amplitude. Lysate was cleared by centrifugation, and His-tagged Wars column-purified using His-Select beads (Sigma). Following elution of 4x 1ml fractions using elution buffer (20mM Tris-Cl pH7.4, 150mM NaCl, 0.1M DTT, 0.02% Tween-20, 250mM imidazole, 15% glycerol), purification was evaluated via SDS-PAGE. Once purification was confirmed protein fractions were dialysed in dialysis buffer (20mM Tris-Cl pH7.4, 150mM NaCl, 0.1M DTT, 15% glycerol) to remove imidazole from the elution buffer. Proteins were aliquoted and stored at -80 °C until use.

### Tryptophan-tRNA transcription

Prior to performing the Aminoacylation assay, tryptophan-tRNA was transcribed. Briefly, human tryptophan tRNA top and bottom oligomers were hybridised by heating to 95°C, and slowly cooling to 4°C. The annealed DNA template was then transcribed using Ribomax Express T7 enzyme for 2 hrs at 37°C. 10x DNase enzyme was added to the reaction for 15 min to degrade remaining DNA template. The tRNA was then purified using RNA purification kit (New England Bio), and concentration was evaluated using spectrophotometry.

### Aminoacylation assay

This assay is used to quantify the Aminoacylation activity of tRNA synthetase enzymes, including Wars, which facilitates binding of Tryptophan to tryptophan-tRNA molecules. This assay relies upon quantification of free phosphate generated during this reaction as set out by Cestari and Stewart (2013, PMID 23134734). Malachite green binds free phosphates and produces a colour change from yellow to green. The extent of colour change can be measured with a spectrophotometer.

In order to establish aminoacylation efficiency of WARS1 mutants, aminoacylation assays were conducted. A master mix of reagents was prepared containing 10x assay buffer (0.5M Tris, 0.1M MgCl<sub>2</sub>, 0.4M KCl, 0.01M DTT, 0.01M ATP), 10x Trp stock (0.04% in water), 0.1U/μl inorganic yeast pyrophosphatase, 8μM tryptophan-tRNA. 20 μg Wars protein was added to each well of a flat-bottomed 96 well plate, followed by master mix and water reaching a total reaction volume of 50 μl. The reaction was incubated at 37°C for 30 min. Following incubation, 100μl BIOMOL Green was added, and absorbance of each well was measured at 600nm at 15 min and 30 min. Reactions containing 50 μL water were used to establish a background BIOMOL absorbance, which was used to normalise reaction data. Reactions without WARS1 protein were used to determine baseline activity. Three repeats were conducted, each containing two replicate wells, and data are represented as mean ± standard deviation.

#### Thermal shift assay

In order to assess stability of expressed WARS1 mutant proteins, thermal shift assays were conducted using Protein Thermal Shift™ Dye Kit (AppliedBiosystems). 0.3 μg WARS1 protein was combined with Protein Thermal Shift™ Dye Kit Buffer and 8x Dye in 20 μL reactions in a 96 well PCR plate. Reactions were also conducted with the addition of 200 μM tryptophan and 8 μM tryptophan-tRNA. ROX reporter dye fluorescence was measured using a StepOnePlus™ Real-Time PCR System after each 1 % increase in temperature between 25 °C and 99 °C. Melting temperature (T<sub>M</sub>) was derived from the peak in fluorescence from plotting melt curves of fluorescence against temperature.

## Results

We ascertained multiple individuals with heterozygous missense *WARS1* variants, all of which occur within the Wars1 catalytic domain (Figure V-1). These individuals were ascertained through DDD, 100kGP, or GeneMatcher. These individuals have a spectrum of neurodevelopmental phenotypes. Patients identified with T160M, R162W, G163V, Q194K and D193N variants have developmental delay and intellectual disability (DD/ID). Patients identified with H257R, D313G and D314V variants also have peripheral neuropathy (PN).

TrpRS Catalytic Domain (154-362)			
			<b>Variants:</b>
<b>TrpRS Functional motifs:</b>			T160M
<b>HVGH</b> (170-173):			R162W
ATP and tRNA stabilisation during aminoacylation.			G163V
<b>AIDQ</b> (310-313):			Q194K
Conformational change upon Trp binding.			D198N
<b>KMSAS</b> (349-353):			H257R
ATP binding.			D314G/V
154	KPFYLYTGRGPSSEAMHVGH	LIPFIFTKWLQDVFNVPLVIQM	195
197	TDDEKYLWKDLTLDQAYS	AVENAKDIIACGFDINKTFIFSDL	238
239	DYMGMSGFYKNVVKIQKH	VTFNQVKGIFGFTDSDCIGKIS	279
281	FPAIQAAPSFNSFPQIFRDRTDIQCLIPC	AIDQDPYFRMTRD	322
323	VAPRIGYPKALLHSTFFPALQGAQT	KMSASDPNSSIFLT-----	362

Figure V-1. Wars1/TrpRS catalytic domain amino acid sequence (aa 154-362). Conserved functional motifs are highlighted in yellow and amino acid substitutions identified in patients are highlighted in red.

To determine the relationship between phenotype severity and *WARS1* missense variants, we conducted enzymatic and stability assays using recombinantly expressed Wars1. Ability of Wars1 with variants identified in patients to aminoacylate Trp-tRNA indicated no correlation between variant and phenotype (Figure V-2). R162W, G163V and Q194K variants identified in patients with DD/ID demonstrated large reductions in aminoacylation activity, whereas T160M and D198N variants identified in DD/ID patients retained more than 70 % of wild type activity. Similar variability was observed among variants identified in patients with peripheral neuropathy, with H257R indicating no reduction in aminoacylation activity compared to WT. Both variants at residue D314 indicate reduced aminoacylation activity, but by different extent.

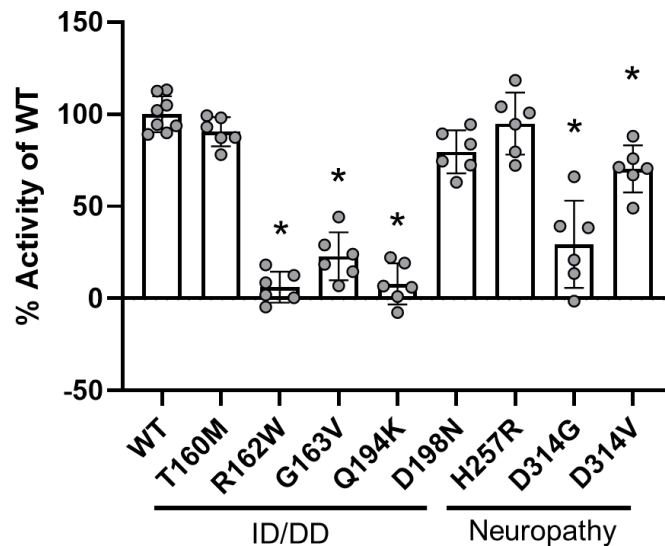


Figure V-2: Aminoacylation activity of recombinant Wars1 with variants identified in patients with ID/DD (intellectual disability and/or developmental delay) or neuropathy. Activity was quantified by change in Biomol absorbance following conversion of pyrophosphate by-product of aminoacylation to free phosphates. Data are averages of two replicate wells across three assays ( $n=6$ )  $\pm$  SD, expressed as a proportion of WT activity. \*  $p<0.01$ , Kruskal-Wallis test.

Next, the stability of Wars1 variants was established by determining the melting temperature ( $T_m$ ) in a Protein Thermal Shift™ assay. Wars1 proteins were assayed in reactions with and without Trp-tRNA and tryptophan amino acid to establish the effect of tRNA binding on protein stability. Compared to Wars1<sup>WT</sup>, the thermal shift curves of the mutant proteins varied considerably (Figure V-3A). There was little change in the curve ‘fingerprint’ from WT in Wars1 proteins with the T160M, R162W, and G163V mutations. The Q194 and H257R curves both had an additional earlier peak, indicating one part of the protein unfolds at a lower temperature compared to WT. A shift in the peak of fluorescence to the left in the Q194K and D198N assays indicates a lower melting temperature ( $T_m$ ) compared to WT (Figure V-3B), suggesting stability is affected more by these variants. Both D314G and D314V variants had curves non-typical of the WT Wars1 protein (Figure V-3A), suggesting protein folding is dramatically altered or impaired as a result of these variants. The D314G mutant did not produce a fluorescence peak, suggesting protein is already aggregated at room temperature, leading to high variability in  $T_m$  quantification (Figure V-3B). A fluorescence peak was produced at very low temperatures in reactions with the D314V mutant, leading to quantification of a low  $T_m$ . Addition of tryptophan and Trp-tRNA to the reaction had little effect on  $T_m$  of the WT, and stable mutants, however among the unstable mutants D314G and D314V,

addition of tryptophan and Trp-tRNA resulted in small rises in average  $T_m$ , suggesting the stability of these mutants is increased when assessed during a reaction with functional activity (Figure V-3C).

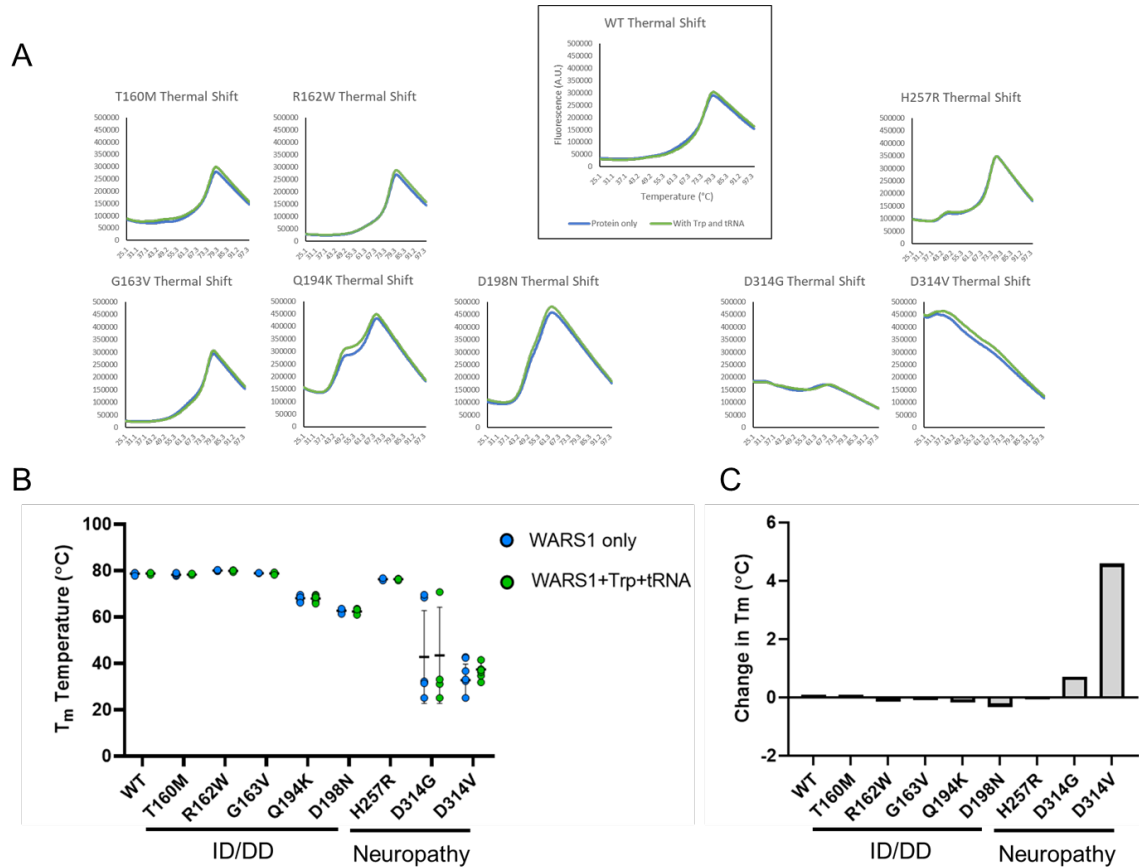


Figure V-3: Temperature stability of recombinant Wars1 with variants identified in patients with ID/DD or neuropathy. (A) Thermal shift assay fluorescence peaks for WT and mutant Wars1. Stability was assessed with and without addition of tryptophan and Trp-tRNA. Graphs display average of duplicate reactions from one reading, representative of three repeats conducted. (B) Average  $T_m \pm$  SD of Wars1 WT and mutants, quantified from the peak in Protein Thermal Shift™ assay dye fluorescence ( $n=6$ ). (C) Change in average Wars1  $T_m$  on addition of tryptophan and Trp-tRNA.

## References

- Li, J.-Q., Dong, H.-L., Chen, C.-X., Wu, Z.-Y., 2019. A novel WARS mutation causes distal hereditary motor neuropathy in a Chinese family. *Brain* 142, e49. <https://doi.org/10.1093/brain/awz218>
- Lin, S.-J., Vona, B., Porter, H.M., Izadi, M., Huang, K., Lacassie, Y., Rosenfeld, J.A., Khan, S., Petree, C., Ali, T.A., Muhammad, Nazif, Khan, S.A., Muhammad, Noor, Liu, P., Haymon, M.-L., Rüschen-dorf, F., Kong, I.-K., Schnapp, L., Shur, N., Chorich, L., Layman, L., Haaf, T., Pourkarimi, E., Kim, H.-G., Varshney, G.K., 2022. Biallelic variants in WARS1 cause a highly variable neurodevelopmental syndrome and implicate a critical exon for normal auditory function. *Hum Mutat.* <https://doi.org/10.1002/humu.24435>
- Okamoto, N., Miya, F., Tsunoda, T., Kanemura, Y., Saitoh, S., Kato, M., Yanagi, K., Kaname, T., Kosaki, K., 2022. Four pedigrees with aminoacyl-tRNA synthetase abnormalities. *Neurol Sci* 43, 2765–2774. <https://doi.org/10.1007/s10072-021-05626-z>
- Shen, N., Guo, L., Yang, B., Jin, Y., Ding, J., 2006. Structure of human tryptophanyl-tRNA synthetase in complex with tRNA<sup>Trp</sup> reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res* 34, 3246–3258. <https://doi.org/10.1093/nar/gkl441>
- Tsai, P.-C., Soong, B.-W., Mademan, I., Huang, Y.-H., Liu, C.-R., Hsiao, C.-T., Wu, H.-T., Liu, T.-T., Liu, Y.-T., Tseng, Y.-T., Lin, K.-P., Yang, U.-C., Chung, K.W., Choi, B.-O., Nicholson, G.A., Kennerson, M.L., Chan, C.-C., De Jonghe, P., Cheng, T.-H., Liao, Y.-C., Züchner, S., Baets, J., Lee, Y.-C., 2017. A recurrent WARS mutation is a novel cause of autosomal dominant distal hereditary motor neuropathy. *Brain* 140, 1252–1266. <https://doi.org/10.1093/brain/awx058>
- Wang, B., Li, X., Huang, S., Zhao, Huadong, Liu, J., Hu, Z., Lin, Z., Liu, L., Xie, Y., Jin, Q., Zhao, Huihui, Tang, B., Niu, Q., Zhang, R., 2019. A novel WARS mutation (p.Asp314Gly) identified in a Chinese distal hereditary motor neuropathy family. *Clinical Genetics* 96, 176–182. <https://doi.org/10.1111/cge.13563>