Whole-genome sequencing in early-onset movement disorders: diagnosis, discovery, and deepening understanding of rare conditions

Dora Batia Dyne Steel

University College London

January 2023

A thesis submitted for the degree of Doctor of Philosophy

Principal Supervisor: Prof Manju A. Kurian
Subsidiary Supervisors: Dr Kimberley M. Reid
Prof Paul Gissen
Tertiary Supervisor: Dr Katy Barwick

In memory of my grandfather Dolf Polak, who died shortly before the completion of this thesis and who still took an interest in medical research at the age of 98.

And dedicated to my beloved daughter Miriam, who has been a grave impediment at every stage of this project.

Declaration

I, Dora Batia Dyne Steel, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Early-onset movement disorders include dystonia, chorea, myoclonus, parkinsonism, tremor, spasticity, ataxia or a combination of these. While some occur due to identifiable pre- or perinatal insults, many are suspected to be due to genetic faults. Diagnosing these remains challenging, partly because many genes responsible for early-onset movement disorders have yet to be identified, and partly because the phenotypic spectrum of ultra-rare genetic disorders is often incompletely described, complicating the interpretation of results.

In this thesis I describe using whole-genome sequencing (WGS) to investigate 166 families with suspected genetic movement disorders. Utilizing a combination of detailed clinical phenotyping and either very broad gene panels or panel-free/gene-agnostic analysis, I aimed to identify variants in genes previously known to have an associated disease phenotype, or for which the phenotype was inadequately reported. Where a likely new disease-gene relationship was identified, I investigated further by laboratory experimentation including studies of splicing, gene expression and protein localisation and by case-finding in collaboration with other research groups.

My analysis resulted in findings suspected to be relevant in 45.8% (76/166) of participants, including a probable or definite diagnosis in 31.3% (55/166). I identified three genes not previously reported in disease: *VPS16*, *VPS41* and *DRD1*. I also contributed to the index case series of three more previously undescribed genetic conditions. Many participants had presentations which expanded the known phenotypic spectrum of their disorder, including those with variants in *SLC30A9, RHOBTB2* and *JPH3*.

Overall, I identified one participant with an identifiable previously undescribed genetic disorder for every 19 analyses conducted. This high rate of significant findings confirms the value of WGS in populations with rare childhood movement disorders as a tool for both diagnosis and gene discovery. I discuss the challenges and opportunities of the use of WGS in this field.

Impact statement

Early-onset movement disorders can be disabling, stigmatising, painful and in extreme cases life-threatening. A high proportion remain undiagnosed – and undiagnosable – as the causative gene has not yet been found. Furthermore, interpreting the results of genetic tests is complicated by the fact that the range of ways very rare genetic disorders can present is often not fully understood.

In this project I identified three novel genetic causes of movement disorders, and contributed to the identification of a further three. This new information means that others affected by these conditions can now be diagnosed through clinical testing, giving them an explanation for their symptoms and the option of genetic counselling. It can also help guide treatment choices and allow the formation of peer support and advocacy groups. Gene identification is also a necessary first step in the development of precision treatments such as genetic therapies.

Besides identifying previously unknown causes of movement disorders, I also looked at ultra-rare genetic movement disorders where the range of presentations had not been fully described, including conditions due to variants in *SLC30A9, RHOBTB2, JPH3* and others. This will help genetic laboratory services to plan and interpret tests more accurately (for example, by deciding which genes to include in panels) and clinicians to understand the symptoms and needs of people affected by these conditions.

This project demonstrates the power of WGS as a tool to identify genetic causes of ultra-rare disorders and to improve our understanding of them. More specifically, it highlights the value of combining gene-agnostic genomic analysis with detailed phenotypic information about participants. As WGS increasingly becomes a part of clinical practice, understanding how to make the best use of it, both for research and diagnostics, will be essential.

Acknowledgements

Thank you very much to my supervisors: both Prof Manju Kurian, my principal supervisor, whose guidance and help was key to this project, and my subsidiary supervisors Dr Kimberley Reid and Prof Paul Gissen, and tertiary supervisor Dr Katy Barwick, for their invaluable advice and assistance.

I would also like to thank all members of the Kurian Neurogenetics Research group for their vital support, encouragement and teaching. Especial thanks go to Drs Robert Spaull and Audrey Soo for their assistance in recruiting and phenotyping participants and to soon-to-be Dr Dimitri Budinger for giving me my first lessons in quantitative real-time PCR. My supervisors Drs Reid and Barwick have spent many, many hours teaching and advising me and both have the patience of saints.

Much of this thesis describes collaborative work and I am indebted to all my many co-authors, future co-authors and collaborators.

I am very grateful for the support of my family, especially my mother, Louisa Polak, who provided essential extra childcare during the final months of the project, and my father, Sam Steel, who diligently proofread despite having no interest whatsoever in the subject matter.

Finally, and most importantly, I would like to thank our study participants, including parents, without whose contribution research such as this would be literally impossible.

My fellowship was funded by a grant to Prof Kurian from the National Institute of Health Research and work presented in this thesis also benefitted from funding from the Sir Jules Thorn Trust and Rosetrees Trust.

Division of Labour Statement

Many parts of this thesis describe collaborative work. I have indicated this in the main text throughout and have referenced our published and in-press papers where appropriate. Where I have included lightly-edited text from referenced papers (published or in preparation) I have only used text which I drafted myself. Where figures are reproduced, they are figures I prepared myself except where otherwise specified.

The following is a statement of specific contributions by others to the work described. Throughout, Sanger confirmation of some variants was kindly shared by Dr Katy Barwick. Clinical phenotyping depended on the contributions of a large number of referring clinicians who supplied information (with special thanks to Dr Suvasini Sharma, Dr Laura Cif, Dr Belen Perez Dueñas and her team, Dr Vivek Jain, Dr Yilmaz Yildiz, Dr Monica Troncoso, Dr Hilla Ben-Pazi, Dr Mohamed Abunada, Dr Lucinda Carr, Dr Michael Zech, Dr Niccolò Mencacci and Dr Jean-Pierre Lin) and also on the Neurogenetic Movement Disorders Clinic team (especially Prof Kurian, Dr Robert, Spaull, Dr Audrey Soo, Dr Aikaterini Vezyroglou, Dr Amit Batla, Dr Kailash Bhatia, Ms Laura Nallen and the administrative team including Ms Charlotte Green and Ms Anna Swieczkowska). For neuroradiology analysis I am indebted to the Great Ormond Street Hospital paediatric neuroradiology team, especially but not only Dr Sniya Sudhakar, Dr Kshitij Mankad and Dr Felice D'Arco.

Chapter 3: For the *INTS11, H2AC6* and *TNPO2* case series, as made clear in the text, projects were led by other teams and my role was genomic analysis in specific cases and contribution of genetic and phenotypic data. As the manuscripts for *INTS11* and *H2AC6* remain unsubmitted and without a finalised authors' list I am unable to credit the teams in full, but the *H2AC6* work is led by Dr Heather Mefford while that on *INTS11* is led by Dr Erica Macke and Dr Marcello Niceta.

Dr Barwick made the final diagnosis for PE13: following my initial analysis, re-called data from this triome needed complete re-analysis due to the data issue briefly referred to in section 2.6.1.

Chapter 4: In this chapter I refer to several published papers (on *SLC30A9, YY1, GNAO1, STXBP1, H3.3, STXBP1* and *RHOBTB2*). Specific author contributions are discussed further in the Published Paper Declaration forms. Experimental work is described for *SLC30A9*: fibroblast culture, RNA extraction and cDNA synthesis was undertaken by Dr Kimberley Reid. Molecular modelling was undertaken by Dr Maya Topf and her team.

Chapter 5: This chapter refers extensively to our published paper on pathogenic variants in *VPS16* and *VPS41*. Contributions are explained within the text: statistical analysis of a cohort of patients with dystonia was undertaken by members of Prof Juliane Winkelmann's team, especially Dr Michael Zech and Dr Chen Zhiao. Fibroblast preparation and culture was performed by Dr Derek Burke's laboratory at GOSH and electron microscopy by Mr Glenn Anderson. Sequencing of the splicing products for DY173P was undertaken by Dr Katy Barwick. Clinical images are stills from a video provided by Dr Laura Cif's team.

Chapter 6: This chapter refers to a manuscript in preparation, with proposed authors and their contributions listed on the Published Papers form. Experimental work as follows was undertaken by Dr Kimberley Reid (as made clear within the text): cell culture, site-directed mutagenesis, transfection, Western blotting, biotinylation studies, the cAMP assay for receptor activity (including testing of dopaminergic agonists) and ligand binding assays. Figures 6.8-6.10, which refer directly to this work, were prepared by Dr Reid.

The following are specific acknowledgements using the wording requested by the organisations in question:

This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org.

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.

Figures in this thesis were created with Biorender.com.

Research Paper Declaration Forms are provided overleaf.

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
a) Where was the work published? (e.g. journal name)	Neurology	
b) Who published the work? (e.g. Elsevier/Oxford University Press):	Wolters Kluwer	
c) When was the work published?	01/03/2021	
d) Was the work subject to academic peer review?	Yes	
e) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
c) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
d) Stage of publication	Please select.	

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**

Sara Zagaglia, Dora Steel, S Krithika Laura Hernandez-Hernandez, Helena Martins Custodio, Kathleen M Gorman, Aikaterini Vezyroglou, Rikke S Møller, Mary D King, Trine Bjørg Hammer, Robert Spaull, Walid Fazeli, Tobias Bartolomaeus, Diane Doummar, Boris Keren, Cyril Mignot, Nathalie Bednarek, J Helen Cross, Andrew A Mallick, Alba Sanchis-Juan, Anna Basu, F Lucy Raymond, Bryan J Lynch, Anirban Majumdar, Hannah Stamberger, Sarah Weckhuysen, Sanjay M Sisodiya, Manju A Kurian. I am joint first author on this manuscript with Sara Zagaglia. I undertook genomic analysis and clinical phenotyping for one participant, helped coordinate data collection, and drafted the manuscript jointly with Sara Zagaglia (who drafted the section on epilepsy. I also prepared figures. Any figures or near-verbatim text I have used in the thesis I prepared myself..

4. In which chapter(s) of your thesis can this material be found?

4

5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Dora Steel	**Date:**	19/12/2022
Supervisor/ Senior Author (where appropriate)**:**	Manju Kurian	**Date:**	21/12/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

6. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
f) Where was the work published? (e.g. journal name)	Movement Disorders Clinical Practice	
g) Who published the work? (e.g. Elsevier/Oxford University Press):	Wiley	
h) When was the work published?	01/07/2022	
i) Was the work subject to academic peer review?	Yes	
j) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
7. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
e) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
f) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
g) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
h) Stage of publication	Please select.	

8. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**
Robert Spaull, Dora Steel, Katy Barwick, Prab Prabhakar, Emma Wakeling, Manju A Kurian. My contribution was the genomic analysis of the proband.
9. In which chapter(s) of your thesis can this material be found?
4
10. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Dora Steel	**Date:**	19/12/2022
Supervisor/ Senior Author (where appropriate)**:**	Manju Kurian	**Date:**	21/12/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

11. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
k) Where was the work published? (e.g. journal name)	Movement Disorders	
l) Who published the work? (e.g. Elsevier/Oxford University Press):	Wiley	
m) When was the work published?	01/07/2022	
n) Was the work subject to academic peer review?	Yes	
o) Have you retained the copyright for the work?	Yes	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
12. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
i) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
j) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
k) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
l) Stage of publication	Please select.	

13. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**

Thomas Wirth, Giacomo Garone, Manju A Kurian, Amélie Piton, Francisca Millan, Aida Telegrafi, Nathalie Drouot, Gabrielle Rudolf, Jamel Chelly, Warren Marks, Lydie Burglen, Diane Demailly, Phillipe Coubes, Mayte Castro-Jimenez, Sylvie Joriot, Jamal Ghoumid, Jérémie Belin, Jean-Marc Faucheux, Lubov Blumkin, Mariam Hull, Mered Parnes, Claudia Ravelli Gaëtan Poulen, Nadège Calmels, Andrea H Nemeth, Martin Smith, Angela Barnicoat, Claire Ewenczyk, Aurélie Méneret, Emmanuel Roze, Boris Keren, Cyril Mignot, Christophe Beroud, Fernando Acosta Jr, Catherine Nowak, William G Wilson, Dora Steel, Alessandro Capuano, Marie Vidailhet, Jean-Pierre Lin, Christine Tranchant, Laura Cif, Diane Doummar, Mathieu Anheim. My contribution was genomic analysis of two participating families and clinical phenotyping of a third.

14. In which chapter(s) of your thesis can this material be found?

4

15. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Dora Steel	**Date:**	19/12/2022
Supervisor/ Senior Author (where appropriate)**:**	Manju Kurian	**Date:**	21/12/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

<table>
<tr><td colspan="3">16. For a research manuscript that has already been published (if not yet published, please skip to section 2):</td></tr>
<tr><td>p) Where was the work published? (e.g. journal name)</td><td colspan="2">Annals of Neurology</td></tr>
<tr><td>q) Who published the work? (e.g. Elsevier/Oxford University Press):</td><td colspan="2">Wiley</td></tr>
<tr><td>r) When was the work published?</td><td colspan="2">01/11/2020</td></tr>
<tr><td>s) Was the work subject to academic peer review?</td><td colspan="2">Yes</td></tr>
<tr><td>t) Have you retained the copyright for the work?</td><td colspan="2">Yes</td></tr>
<tr><td colspan="3">[If no, please seek permission from the relevant publisher and check the box next to the below statement]:
☐ I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</td></tr>
<tr><td colspan="3">17. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):</td></tr>
<tr><td>m) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):</td><td>Please select.</td><td>If yes, which server? Click or tap here to enter text.</td></tr>
<tr><td>n) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)</td><td colspan="2">Click or tap here to enter text.</td></tr>
<tr><td>o) List the manuscript's authors in the intended authorship order:</td><td colspan="2">Click or tap here to enter text.</td></tr>
<tr><td>p) Stage of publication</td><td colspan="2">Please select.</td></tr>
</table>

18. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**

Steel D, Zech M, Zhao C, Barwick KES, Burke D, Demailly D, Kumar KR, Zorzi G, Nardocci N, Kaiyrzhanov R, Wagner M, Iuso A, Berutti R, Škorvánek M, Necpál J, Davis R, Wiethoff S, Mankad K, Sudhakar S, Ferrini A, Sharma S, Kamsteeg EJ, Tijssen MA, Verschuuren C, van Egmond ME, Flowers JM, McEntagart M, Tucci A, Coubes P, Bustos BI, Gonzalez-Latapi P, Tisch S, Darveniza P, Gorman KM, Peall KJ, Bötzel K, Koch JC, Kmieć T, Plecko B, Boesch S, Haslinger B, Jech R, Garavaglia B, Wood N, Houlden H, Gissen P, Lubbe SJ, Sue CM, Cif L, Mencacci NE, Anderson G, Kurian MA, Winkelmann J; Genomics England Research Consortium. I am the first author on this work: my contribution was drafting and editing the manuscript; genomic analysis of the original family for *VPS16* and the *VPS41* family; design of splicing primers for *VPS41* and coordinating collection of clinical, radiological and microscopic data.

19. In which chapter(s) of your thesis can this material be found?

5

20. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Dora Steel	**Date:**	19/12/2022
Supervisor/ Senior Author (where appropriate)**:**	Manju Kurian	**Date:**	21/12/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

21. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
u) Where was the work published? (e.g. journal name)	Click or tap here to enter text.	
v) Who published the work? (e.g. Elsevier/Oxford University Press):	Click or tap here to enter text.	
w) When was the work published?	Click or tap to enter a date.	
x) Was the work subject to academic peer review?	Please select.	
y) Have you retained the copyright for the work?	Please select.	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
22. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
q) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):	No	**If yes, which server?** Click or tap here to enter text.
r) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)	Brain	
s) List the manuscript's authors in the intended authorship order:	*Kimberley M Reid, *Dora Steel, Sanjana Nair, Sanjay Bhate, Lorenzo Biassoni, Sniya Sudhakar, Michelle Heys, Elizabeth Burke, Erik-Jan Kamsteeg, Genomics England Research Consortium, Biju Hameed,	

	Michael Zech, Niccolo Mencacci, Katy Barwick, Maya Topf, Manju A. Kurian
t) Stage of publication	Not yet submitted

23. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**

Dr Reid and I will be listed as equal contributors to this manuscript: I undertook clinical phenotyping and genomic analysis. We both did confocal microscopy (and slide preparation), RNA extraction and qrtPCR. All other laboratory techniques were undertaken by Dr Reid. We prepared the manuscript and figures jointly: the figures in the thesis (except where otherwise specified) are those prepared by me and any text used verbatim was written by me. Sanjana Nair and Maya Topf undertook molecular modelling. Lorenzo Biassoni and Sniya Sudhakar provided radiological analysis. Sanjay Bhate, Michelle Heys and Biju Hameed were involved in care and assessment of the proband. Katy Barwick undertook Sanger sequencing using primers I designed. Prof Kurian directed the project and edited the manuscript. Other authors interrogated databases looking for additional affected individuals.

24. In which chapter(s) of your thesis can this material be found?

6

25. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate:	Dora Steel	**Date:**	19/12/2022
Supervisor/ Senior Author (where appropriate)**:**	Manju Kurian	**Date:**	21/12/2022

Table of Contents

Abbreviations

5HIAA: 5-hydroxyindoleacetic acid

5-HT: 5-hydroxytryptamine

AADC: aromatic L-amino acid decarboxylase

AAV: adeno-associated virus

ACGS: Association for Clinical Genomic Science

ACMG: American College of Medical Genetics and Genomics

AD: autosomal dominant

AHC: alternating hemiplegia of childhood

AR: autosomal recessive

ASO: antisense oligonucleotide

BAM: Binary Alignment/Map

BLPS: Birk-Landau-Perez syndrome

CADD: Combined Annotation Dependent Depletion

cAMP: cyclic adenosine monophosphate

cDNA: complementary DNA

CGH: comparative genomic hybridisation

CNS: central nervous system

CNV: copy number variant

CP: cerebral palsy

CSF: cerebrospinal fluid

DEE: developmental and epileptic encephalopathy

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

EDTA: ethylenediaminetetraacetic acid

EEG: electroencephalography/electroencephalographic

ERT: enzyme replacement therapy

FND: functional neurological disorder

GOSH: Great Ormond Street Hospital for Children

HGNC: HUGO Gene Nomenclature Committee

HIE: hypoxic-ischaemic encephalopathy

HOPS: Homotypic fusion and protein sorting

HOPSANDs: HOPS-associated neurological disorders

HUGO: Human Genome Organisation

HVA: homovanillic acid

iPSC: induced pluripotent stem cells

KO: knockout

miRNA: microRNA

MLPA: multiplex ligation-dependent probe amplification

MRI: magnetic resonance imaging

mRNA: messenger RNA

MSN: medium spiny neuron

NBIA: neurodegeneration with brain iron accumulation

NDIM: neurodevelopmental disorder with involuntary movements

NGS: next-generation sequencing

NHS: National Health Service

NMDC: Neurogenetic Movement Disorders Clinic

oe: observed/expected score

OMIM: Online Mendelian Inheritance in Man

PAM: positive allosteric modulator

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PD: Parkinson’s disease

PKA: protein kinase A

pLI: probability of loss intolerance

PVL: periventricular leukomalacia

QRT PCR: quantitative real-time PCR

RNA: ribonucleic acid

saRNA: small activating RNA

SCAR: spinocerebellar ataxia, autosomal recessive

SDM: site-directed mutagenesis

SIMD: Study of Inherited Metabolic Disorders

siRNA: small interfering RNA

SNARE: snap receptor

SNP: single nucleotide polymorphism

SNV: single nucleotide variant

TBE: tris-borate EDTA

Tm: melting temperature

VCF: Variant Call Format

VER/ERG: visual evoked potentials and electroretinography

VMAT2: vesicular monoamine transporter 2

VUS: variant of uncertain significance

WES: whole-exome sequencing

WGS: whole-genome sequencing

WT: wild-type

Chapter 1: Introduction

1.1 Movement disorders and motor control

Since so much human activity – from spinal reflexes and gut motility to speech and locomotion – involves movement in one form another, the term “movement disorder” could potentially cover a vast amount of human pathology. In practice, however, it refers specifically to disorders of the voluntary control of movement, and therefore of the central nervous system (CNS): primary disorders of muscle and peripheral nerves are excluded.

To understand movement disorders, it is necessary to have some grasp of the physiological basis of human motor control. This is a vast topic and to cover it in full would be beyond the scope of this (or any) thesis, but I will start with a brief summary of the key systems and mechanisms in order to provide a context for discussion of pathology later.

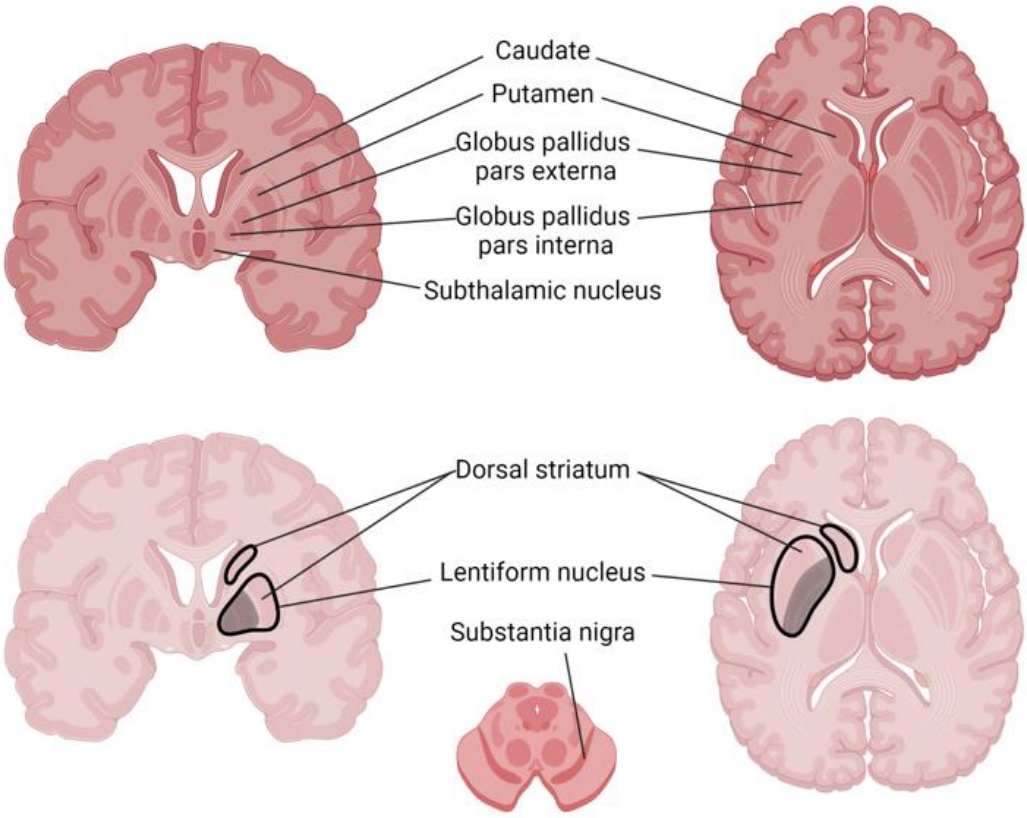
Motor control in humans relies – to oversimplify vastly – on three key elements: the pyramidal system, the cerebellum and (most importantly for my purposes) the basal ganglia. I will deal with the first two very briefly and devote a little more attention to the basal ganglia, which are the most relevant to the conditions I will go on to discuss.

The pyramidal system consists of the corticospinal and corticobulbar tracts. These are very long bundles of neurons, largely but not exclusively arising from the motor cortex, premotor cortex and supplementary motor areas. They pass through the internal capsule and go on (if corticobulbar) to synapse with the cranial nerve nuclei or (if corticospinal) to decussate in the pyramids of the medulla and ultimately synapse with the lower motor neurons or interneurons in the anterior column of the spinal cord. Disorders involving the CNS components of the pyramidal system result in the familiar “upper motor neuron” syndrome of spasticity, weakness, hyperreflexia and upgoing plantars.

The cerebellum is a hindbrain structure which contributes to balance, coordination, rhythmic movement and motor learning. Projections from the deep cerebellar nuclei reach a variety of targets including the cerebral cortex (via the thalamus), the red nucleus and the reticular formation. The classic presentation of cerebellar dysfunction includes dysmetria, dysarthria, dysdiadochokinesis, nystagmus and ataxia.

The basal ganglia are a collection of deep grey matter structures within the cerebral hemispheres and midbrain. Their basic anatomy is shown in **Figure 1.1**. (The ventral pallidum, including the nucleus accumbens, is omitted for simplicity; the substantia nigra – not always considered strictly one of the basal ganglia – is included due to its central role in motor pathways and pathology.) The role of the basal ganglia is complex and besides regulation of movement it includes contributions to learning,(1) motivation,(2) cognition,(3) and impulse control(4) – but movement will be our primary concern here.

Figure 1.1: *Basic anatomy of the human basal ganglia, shown in the cerebrum in coronal (left) and axial (right) views and the midbrain in axial view (bottom). The term "lentiform" or "lenticular nucleus" is used collectively for the globus pallidus and the putamen, while the striatum consists of the dorsal striatum (putamen and caudate) and ventral striatum (nucleus accumbens and olfactory tubercle, not shown).*

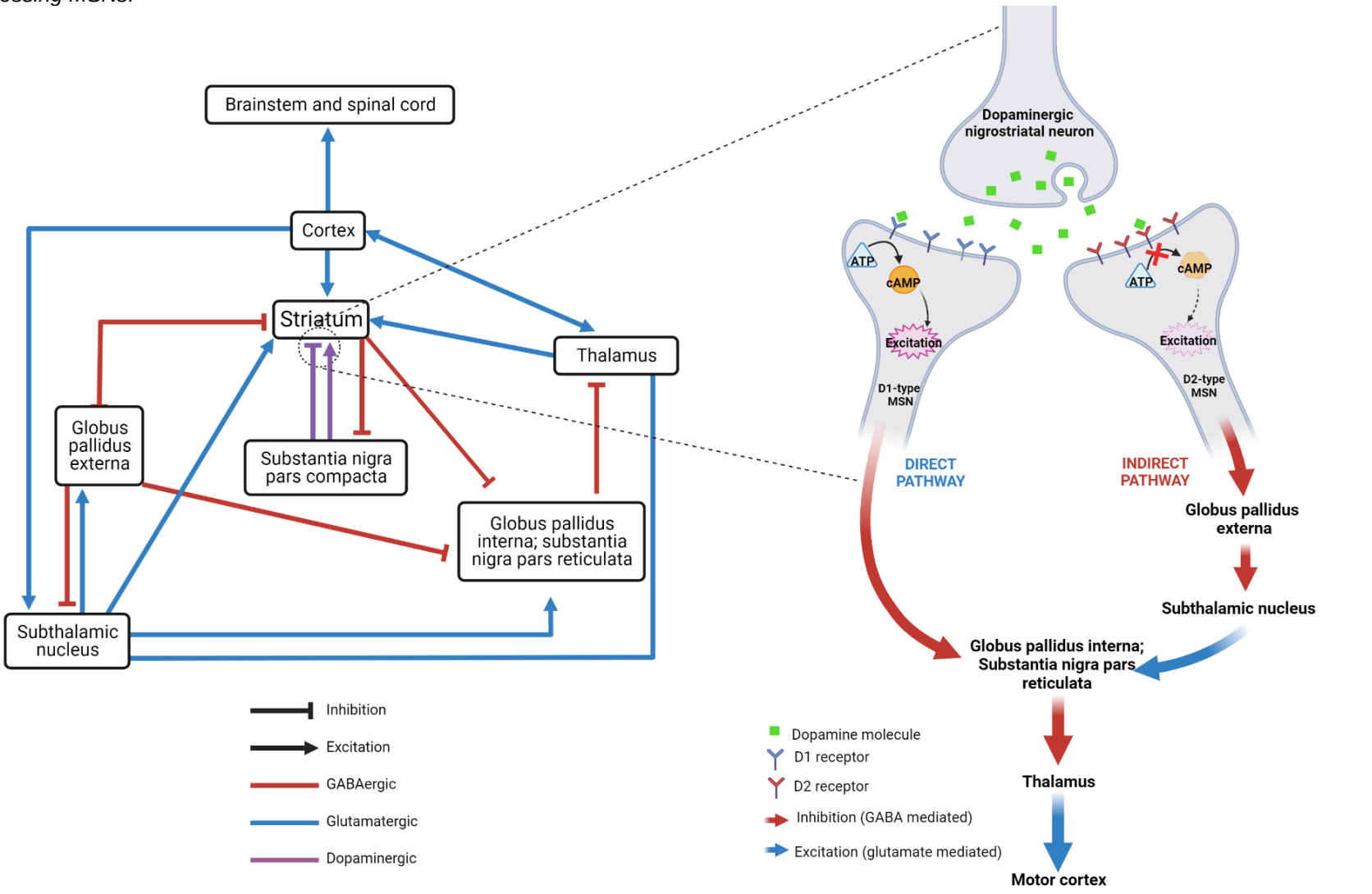


The classic model by which the basal ganglia modulate voluntary movement is that of the direct and indirect pathways **(Figure 1.2)**.(5, 6) The striatum receives primarily excitatory, glutamatergic input from the motor cortex and ultimately feeds back to it via the thalamus. This feedback can be either positive, as in the direct pathway, where the striatum acts to reduce inhibitory input to the thalamus, or negative, as in the indirect pathway where inhibition of the thalamus is increased. Thus the direct pathway stimulates increased movement while the indirect pathway reduces it. In health, the two pathways exist in a state of equilibrium, modulated by the demands of the environment and the individual's intentional behaviours. The balance between the two is modulated by the substantia nigra pars compacta, which provides dopaminergic input to the medium spiny neurons (MSNs) which make up the striatum. Dopamine has an excitatory effect on the direct pathway, via D_1 receptors, but an inhibitory effect on the indirect pathway, via D_2 receptors.

Initially it was believed there were two entirely distinct populations of MSNs, one expressing exclusively D_1 receptors and the other D_2s, but it is now known that a proportion of cells express both types.(7) Other respects in which the original direct/indirect pathway model is to some extent an oversimplification – notably the existence of multiple interlocking modulatory feedback loops(8) – have also been noted over the years, but in most key respects the model has survived and continues to inform our understanding of the how the basal ganglia regulate movement. Newer evidence continues to support it: for example, in 2010 a team using optogenetic techniques to selectively activate either D_1 or D_2 receptors in mice were able to demonstrate that the former increased locomotor activity while the latter induced hypokinetic symptoms.(9)

Symptoms arising from dysfunction of the basal ganglia, whether hyper- or hypokinetic, are sometimes described as "extrapyramidal" to distinguish them from the pyramidal syndrome of pure upper motor neuron dysfunction. Although still quite widely used, this term is now less favoured as it can lead to various kinds of confusion.

Figure 1.2: *Schematic showing the direct and indirect pathways of the Cortico-basal ganglia-thalamic network and the differing effects of dopamine on D_1- and D_2-expressing MSNs.*



Firstly, it fails to recognise that much of the effect of the basal ganglia is realised through their modulatory effects on the thalamus and cortex – that is, ultimately, through the pyramidal tracts. Secondly, there is no consensus on whether cerebellar symptoms are “extrapyramidal” or not. Thirdly, the “extrapyramidal tracts” of the spinal cord (the vestibulospinal, reticulospinal, tectospinal and rubrospinal tracts – this time justifiably called extrapyramidal as they do not traverse the medullary pyramids) have rather less to do with the basal ganglia than the pyramidal tracts do.

1.2 Movement disorders in clinical practice

1.2.1 What are movement disorders?

Movement disorders may be characterised by either a paucity of voluntary movement (hypokinesia) or excessive and/or involuntary movement (hyperkinesia). They comprise a broad range of diseases with a still broader range of causes. Although they are collectively a common reason for presentation to paediatric neurology services, many questions around management, causation, diagnosis and even nomenclature remain unresolved.

The archetypal hypokinetic disorder is parkinsonism. The principal categories of hyperkinetic movement disorder include dystonia, dyskinesia, myoclonus and tremor. Tics and stereotypies are often discussed separately from other types of movement disorder, possibly because of their partial susceptibility to conscious control, and will not form part of the scope of this thesis. The hypo-/hyperkinetic distinction is not always as clear-cut as it might at first appear. Spasticity is sometimes categorised as primarily a hypokinetic disorder, as it impairs voluntary movement,(10) but it commonly co-exists with dystonia; in childhood, parkinsonism is virtually always accompanied by dystonia. Cerebellar disorders do not obviously fall into either category. Many childhood-onset movement disorders are inherently complex, with multiple different movement disorder semiologies and often different presentations at different developmental stages (for example, the hypotonic infant evolving to the spastic toddler, or the baby with tyrosine hydroxylase deficiency who presents with a hyperkinetic disorder which gradually changes to brady- and then akinesia).(11)

1.2.2 Why do they matter?

Childhood-onset movement disorders are common: cerebral palsy (CP) alone affects over 1 in 500 live births.(12) Although severity varies widely, they are a significant cause of disability(13) and the more severe forms carry a real risk of mortality.(14) Apart from the obvious interference with normal motor function including mobility, self-care and communication, movement disorders may cause significant pain(15) and, being highly visible, are often socially stigmatising.(16) With the exception of CP, however, most *individual* causes of childhood onset movement disorders are rare (population prevalence <1/2000) or ultra-rare (<1/50,000-100,000).(17)

CP was classically defined (by Bax *et al*) as "a group of disorders of the development of movement and posture….attributed to non-progressive disturbances that occurred in the developing fetal or infant brain."(18) Although this remains a useful and widely-recognised definition, it leaves unanswered – perhaps deliberately – the question of where the demarcation between CP and non-CP movement disorders should be drawn, particularly where genetic factors are concerned. For the time being I will use the term 'CP' to refer to the sequelae of non-genetic brain insults including hypoxic-ischaemic brain injury, periventricular leukomalacia, intracerebral haemorrhage, perinatal infection or kernicterus, but not to identified genetic conditions. I will return to this discussion of terminology in my final chapter.

An increasing segment of the spectrum formerly defined as CP is being carved out into separate diagnostic subcategories as our knowledge of aetiology improves. Each individual disorder is rare, but collectively known rare diseases are growing ever commoner: the organisation Rare Disease UK estimates that 1 in 17 people are affected by a rare disorder.(19) The utility, for patients, families and clinicians, of making such distinctions rather than sticking to older and broader diagnostic categories, will be discussed below.

1.2.3 Paediatric versus adult practice

There are several important differences between movement disorders presenting in childhood and those presenting in adulthood. The single commonest condition seen in the adult movement disorder clinic is undoubtedly Parkinson's disease which, even

including all related disorders, is extremely rare in childhood. In children's services, dystonia and spasticity predominate.

Hyperkinetic disorders presenting in adulthood include a relatively high rate of focal and/or task-specific dystonia: these presentations are unusual in childhood and tend to peak in the third to fifth decade of life.(20) Although "pure" or isolated movement disorders can occur in children, combined phenotypes – in which, for example dystonia co-exists with chorea or spasticity – are common. The presence of comorbidities such as neurodevelopmental impairments is also common and indeed any sufficiently severe early-onset movement disorder will have neurodevelopmental impacts of its own.

The aetiological spectrum is correspondingly different. For adult-onset focal and/or task-specific dystonia, a multifactorial origin is presumed in which genetics may play only a minor role.(21) In childhood-onset movement disorders, especially the more severe or complex forms, a single aetiology is considered more likely and (excluding CP) monogenic disorders form a far greater part of the case-mix.

In the past, many severe childhood-onset movement disorders were rarely seen in adult medical practice due to their associated short life-expectancy – and many remain severely life-shortening today: median survival in the most severe form of CP increased markedly between 1983 and 2010, but only from 10.9 to 17.1 years.(22) Nevertheless, a greater number of severely affected children do now reach adulthood, and there is increasing awareness of the need to provide services which meet the specific challenges of adolescent medicine and transition to adult healthcare.(23)

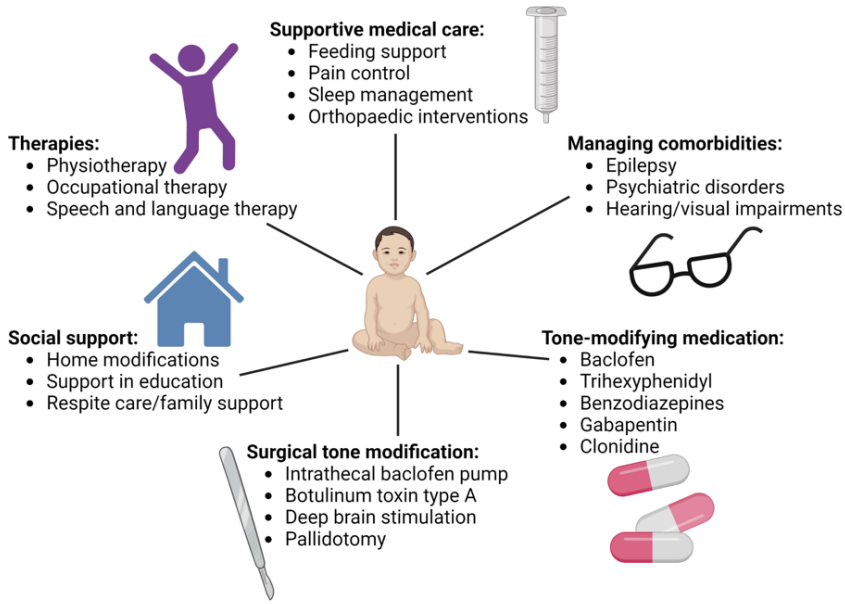
1.2.4 Current management options

Very few childhood-onset movement disorders are curable, although some (like Sydenham's chorea) may resolve spontaneously over time and in others (like Wilson's disease or DOPA-responsive dystonia) excellent symptomatic control can be achieved with appropriate therapies. For the majority, though, permanent persistence of significant symptoms is the norm. In such cases, management will generally require the support of a multidisciplinary team and will be aimed at maximising quality of life,

independence and participation. This may range from – in milder cases where the patient does not wish to risk side-effects – complete non-intervention, to major neurosurgical procedures and indeed to planning and implementing palliative care.

Figure 1.3 summarises the categories of intervention which might need to be considered for any child with a movement disorder. The lists are indicative rather than exhaustive and of course not every individual will require every type of intervention. Management along these lines, regardless of aetiology, can often provide significant benefits with or without a precise molecular diagnosis.

Figure 1.3: *Types of intervention which may be required in the care of a child with a hyperkinetic disorder*



There are, of course, a few instances where a specific medical intervention for an identified disorder can have a transformative effect, such as levodopa for DOPA-responsive dystonia(24) or caffeine for *ADCY5*-related disorders.(25) Moreover, in a small but increasing number of conditions, understanding of pathophysiology has advanced to a point where disease-modifying therapy can be attempted, as with intraventricular enzyme-replacement for *CLN2*-related late infantile Batten's disease.(26) For the large majority of hyperkinetic disorders, however, precision therapies are not an option and treatment is limited to symptomatic management. For most affected individuals, current medical treatment, though valuable, is severely limited in what it can achieve.

1.3 Causes and classifications of movement disorders

1.3.1 Aetiological classification and its limitations

The neurological system by which humans achieve motor control is among the most complex phenomena known to natural science. Its proper functioning depends on a series of interlocking systems at every scale of biology, ranging from correct functioning of ion channels at neuronal synapses, to correct growth and pruning of dendritic spines, to correct guidance of developing axons, to correctly-wired electrophysiological networks in the basal ganglia to, at the highest level, correct conscious and semi-conscious psychological mechanisms.(27) It is therefore unsurprising that the errors which can cause disorders of motor control are correspondingly diverse.

Naturally, multiple classification systems have evolved to try and deal with this complexity. Many of them have their uses but none are perfect and some bear a distressing resemblance to Borges' famous classification of animals in a fictional Chinese encyclopaedia(28)[*] – they offer a range of categories which are neither mutually exclusive nor collectively exhaustive.

Even apparently basic distinctions, such as that between "congenital" and "acquired", are not straightforward. Antenatal brain injuries are "congenital" in that they are present at birth but are clearly "acquired" by the individual, whereas genetic disorders with non-neonatal onset of symptoms are the opposite. Thus I have preferred to use "acquired" vs "genetic" – while acknowledging that in some disorders elements of both may contribute.

Table 1.1 gives one possible classification, not intended to be fully exhaustive but covering the major categories likely to be considered in differential diagnosis. Clearly,

[*] The categories are: belonging to the Emperor; embalmed; trained; sucking-pigs; mermaids; fictional; stray dogs; included within this classification; those that tremble as if mad; innumerable; drawn with a fine camel-hair brush; etcetera; having just broken a vase; those that look from a distance like flies.

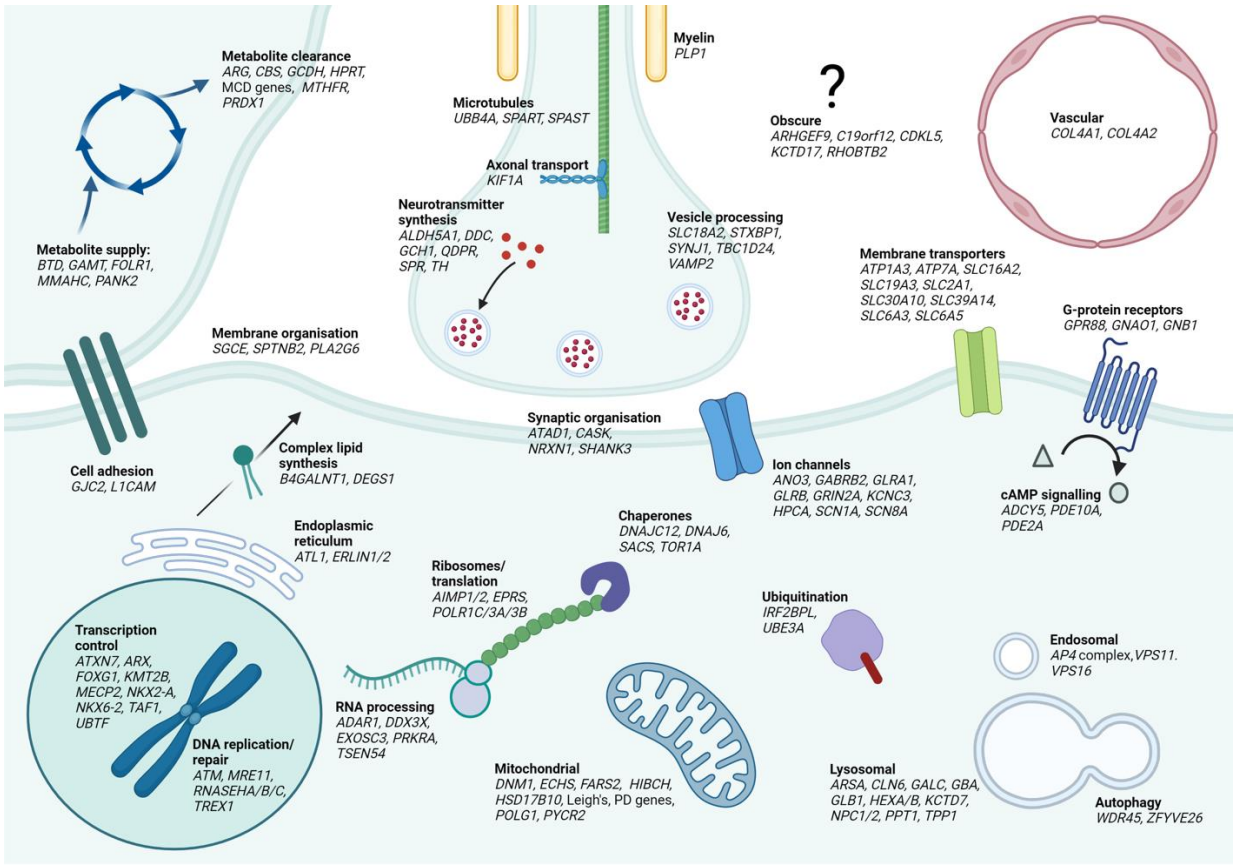
however, it remains open to some of the same criticisms as the Chinese encyclopaedia, especially among the genetic disorders. For example, the distinction between “metabolic” and “primary genetic” (historically called “idiopathic”) disorders rests largely on whether it is possible for clinicians to measure a chemical marker of the disorder in available patient samples. Hence Wilson’s disease, a monogenic disorder due to mutations in *ATP7B* of which movement disorders are often the major manifestation,(29) is considered metabolic, while, say, dopa-responsive dystonia 5, due to a failure to synthesise tetrahydrobiopterin,(30) is considered a primary dystonia. Any genetic movement disorder has, we presume, some underlying mechanism and therefore some associated abnormality at the chemical level, even if it is not easy for us to detect, and therefore the distinction between metabolic and genetic disorders does not seem to rest on solid foundations.

Table 1.1: *Categorisation of causes of childhood-onset movement disorders*

<table>
<tr><th>Acquired</th><th>Unknown</th><th>Genetic</th></tr>
<tr><td>Brain injuries (including CP):
Traumatic
Hypoxic/ischaemic
Vascular
Toxic (e.g. kernicterus)</td><td rowspan="5">A large category: many conditions currently within it are thought likely to be either genetic or multifactorial with a genetic element</td><td rowspan="5">Amenable to multiple overlapping classification systems but includes (with examples):
Abnormalities of neurotransmitter synthesis or processing (DDC, TH)
Abnormalities of cellular signalling (PDE10A, ADCY5)
Metabolic disorders resulting in neurotoxicity/basal ganglia damage (ATP7B, SLC30A10)
Failures of autophagy and other cellular maintenance processes (PANK2, EPG5, VPS16)
Ion channel disorders, including many epilepsy-dyskinesia syndromes (SCN2A, KCNT1)
Disorders leading to structural abnormalities (TUBA1A)
Disorders with complex or poorly-understood genes and pathophysiology (TOR1A, KMT2B, SGCE, NKX2-1)</td></tr>
<tr><td>Iatrogenic:
Antipsychotics
Anti-emetics
L-dopa</td></tr>
<tr><td>Infective/post-infective:
Sydenham’s chorea
PANDAS</td></tr>
<tr><td>Autoimmune:
Anti-NMDA receptor encephalitis
Acute disseminated encephalomyelitis
Systemic lupus erythematosus</td></tr>
<tr><td>Psychogenic</td></tr>
</table>

Genes associated with monogenic movement disorders act in a very wide range of subcellular locations and processes **(Figure 1.4).** From the perspective of disease gene identification, it may make sense to consider all of them together – including both those genes which have clear metabolic markers and well-understood pathophysiological pathways, and those which have neither. From the perspective of the cell biologist, however, it is vital to maintain a clear view of where and how each gene exerts its effects. The choice of classification must therefore be context dependent.

Figure 1.4: *Subcellular roles of genes implicated in childhood-onset movement disorders (selected examples)*



1.3.2 A note on nomenclature

The terminology of movement disorders is complex as regards both the description of semiology and the naming of conditions. Hypokinetic disorders can generally be

described using relatively unambiguous terms such as “bradykinesia” or “akinesia” (although, especially in paediatric practice, the latter can sometimes be confused with severe weakness) and even the eponym “parkinsonism” is well understood as indicating a combination of hypokinesia and rigidity, with or without tremor.

Describing hyperkinetic disorders, however, is more controversial. For example, chorea (low-amplitude, rapid movements resembling dancing), athetosis (slower writhing movements) and ballismus (forceful high-amplitude limb movements) may all be considered subtypes of dyskinesia, or they may be classified separately. The term “dyskinesia” (outside specific syndromic usages such as “tardive dyskinesia” and “levodopa-induced dyskinesia”) is objected to by some authorities as unhelpfully imprecise.(31) Some subtypes of tremor fall within the spectrum of dystonia (dystonic tremor), while others – notably intention tremor in the context of cerebellar dysfunction and parkinsonian tremor – are not generally considered hyperkinetic disorders at all. Myoclonus and dystonia each present particular challenges of classification which are worth discussing separately.

Myoclonus – brief, non-rhythmic jerking movements – has an exceptionally wide range of causes including physiological (such as hypnic jerks), epileptic (including cortical and subcortical-cortical myoclonus), peripheral (due to peripheral nerve dysfunction), spinal or propriospinal (secondary to a range of intraspinal pathology), and “subcortical” (denoting myoclonus arising anywhere in the brain except the cortex, and therefore lacking an electroencephalographic (EEG) correlate).(32) Only the last of these is generally considered to represent a true “movement disorder”. Distinguishing true myoclonus from other jerky movements can be hard or even impossible by clinical assessment alone, and electrophysiology including electromyography and back-averaging may be helpful.(33)

Dystonia has been defined as “sustained or intermittent muscle contractions causing abnormal, often repetitive, movements, postures, or both”, and the authors of that definition, Albanese *et al*,(34) have made an attempt to reform the nomenclature used for its classification. They reject the terms “primary” and “secondary” dystonia (in my

opinion correctly) as confusing, but their own proposed system illustrates the difficulty of designing a clear and comprehensive system of terminology.

They provide a system for classification of dystonia by age of onset, body-distribution, disease course and temporal variability. In my view these elements of the system are helpful in providing clarity and consensus, whereas their classification by aetiology – which includes some surprising features, such as providing “autosomal recessive” and “familial” as alternative, rather than complementary, categories – is less so.

They also make the following distinctions:
1) “Isolated” versus “combined” dystonia – where “combined” means the presence of an additional movement disorder;
2) Presence versus absence of other neurological and/or systemic features.

Although these features are clearly important to note, the terminology is not intuitive: an individual could have “isolated” dystonia if they had no other movement disorder, even if they also had epilepsy, neurodevelopmental impairments and multiple organ involvement. Especially in paediatric practice, where movement disorders are so often one aspect of a neurodevelopmental or neurodegenerative disease, the failure of this system to distinguish between conditions where dystonia is the major presenting complaint and those where it is a subsidiary feature is a limitation. Admittedly, this may not be an easy distinction to draw: virtually *all* forms of genetic dystonia (with the possible exception of DYT-6 due to *THAP1* variants), have associated neurological, neuropsychiatric or systemic features of some kind.

The locus symbols system used in the classification of some monogenic disorders deserves a brief mention here. It was originally introduced by the Human Genome Organisation Nomenclature Committee (HUGO/HGNC) to keep track of disorders for which a genetic locus had been identified but a gene had not, and involves the use of a short prefix indicating the phenotype and a number.(35) The HGNC no longer endorses the system but it remains in use by Online Mendelian Inheritance in Man (OMIM) for a number of conditions, among them several movement disorders including spinocerebellar ataxia (SCA), dystonia (DYT), hereditary spastic paraparesis

(HSP) and hereditary Parkinson's disease (PARK), although not, for some reason, non-paroxysmal dyskinesia or chorea.(36) This system, which is less relevant now that linkage mapping seldom precedes gene identification, has various limitations which I will use the DYT category **(Table 1.2)** to illustrate.

Over the years the classification has unavoidably incorporated some anomalies such as duplications (variants in *GCH1* cause both DYT-5 and DYT-14) and "phantom" conditions such as DYT-22. Moreover, several conditions assigned DYT- numbers, such as the paroxysmal kinesigenic and non-kinesigenic dyskinesias, would not naturally be described as dystonias at all. Conversely, a large number of disorders involving dystonia are not given a DYT- number, either escaping the system together or belonging to another category. These limitations are well summarised by Mencacci *et al* in their review.(37) Attempts have been made to improve the system, for example by adding nuance to the choice of phenotypic prefix,(38) but to my mind it adds a redundant layer of complexity to the nomenclature.

Table 1.2: *The DYT- classification of monogenic dystonias. Blue: gene has another associated disease phenotype not classified in the DYT- system. Green: gene has other associated disease phenotypes including >1 DYT number. Orange: gene has not been identified. ?: association reported in one family only.*

DYT #	Gene	Inheritance	Short phenotype name
1	*TOR1A*	AD	Torsion dystonia 1
2	*HPCA*	AR	Torsion dystonia 2
3	*TAF1*	XLR	X-linked dystonia-parkinsonism
4	*TUBB4A*	AD	Torsion dystonia 4
5	*GCH1*	AD/AR	DOPA-responsive dystonia
6	*THAP1*	AD	Torsion dystonia 6
7	.	AD	Torsion dystonia 7
8	*PNKD*	AD	Paroxysmal non-kinesigenic dyskinesia 1
9	*SLC2A1*	AD	GLUT1 deficiency syndrome
10	*PRRT2*	AD	Episodic kinesigenic dyskinesia 1
11	*SGCE*	AD	Myoclonus-dystonia syndrome
12	*ATP1A3*	AD	Rapid-onset dystonia-parkinsonism
13	.	AD	Torsion dystonia 13?
14	*GCH1*	AD/AR	DOPA-responsive dystonia
15	.	AD	Myoclonic dystonia 15?

16	*PRKRA*	AR	Dystonia 16
17	.	AR	Torsion dystonia 17?
18	*SLC2A1*	AD	GLUT1 deficiency syndrome
19	.	AD	Episodic kinesigenic dyskinesia 2
20	.	AD	Paroxysmal non-kinesigenic dyskinesia 2?
21	.	AD	Dystonia 21?
22	Doesn't exist		
23	.	AD	Dystonia 23
24	*ANO3*	AD	Dystonia 24
25	*GNAL*	AD	Dystonia 25
26	*KCTD17*	AD	Myoclonic dystonia 26
27	*COL6A3*	AR	Dystonia 27
28	*KMT2B*	AD	Childhood-onset dystonia 28
29	*MECR*	AR	Childhood-onset dystonia with optic atrophy and basal ganglia abnormalities
30	*VPS16*	AD	Dystonia 30
31	*AOPEP*	AR	Dystonia 31
32	*VPS11*	AR	Dystonia 32?
33	*EIF2AK2*	AD, AR	Dystonia 33
34	*KCNN2*	AD	Myoclonic dystonia 34?
35	*SHQ1*	AR	Childhood-onset dystonia 35?

In both clinical practice and research, it is usually preferable to adopt a pragmatic system which briefly states the phenotype and (if known) the genetic aetiology: for example, "childhood-onset generalised dystonia due to a heterozygous variant in *KMT2B*". The International League Against Epilepsy, faced with a similar challenge, has settled on a "multiaxial" system in which syndrome, aetiology and comorbidities are listed separately,(39) sacrificing the elegance of a single unified system for the sake of greater precision. Since, as with the epilepsies, no existing nomenclature of the movement disorders can offer perfection, an acceptable level of clarity must serve instead.

1.4 Diagnostic investigation of movement disorders

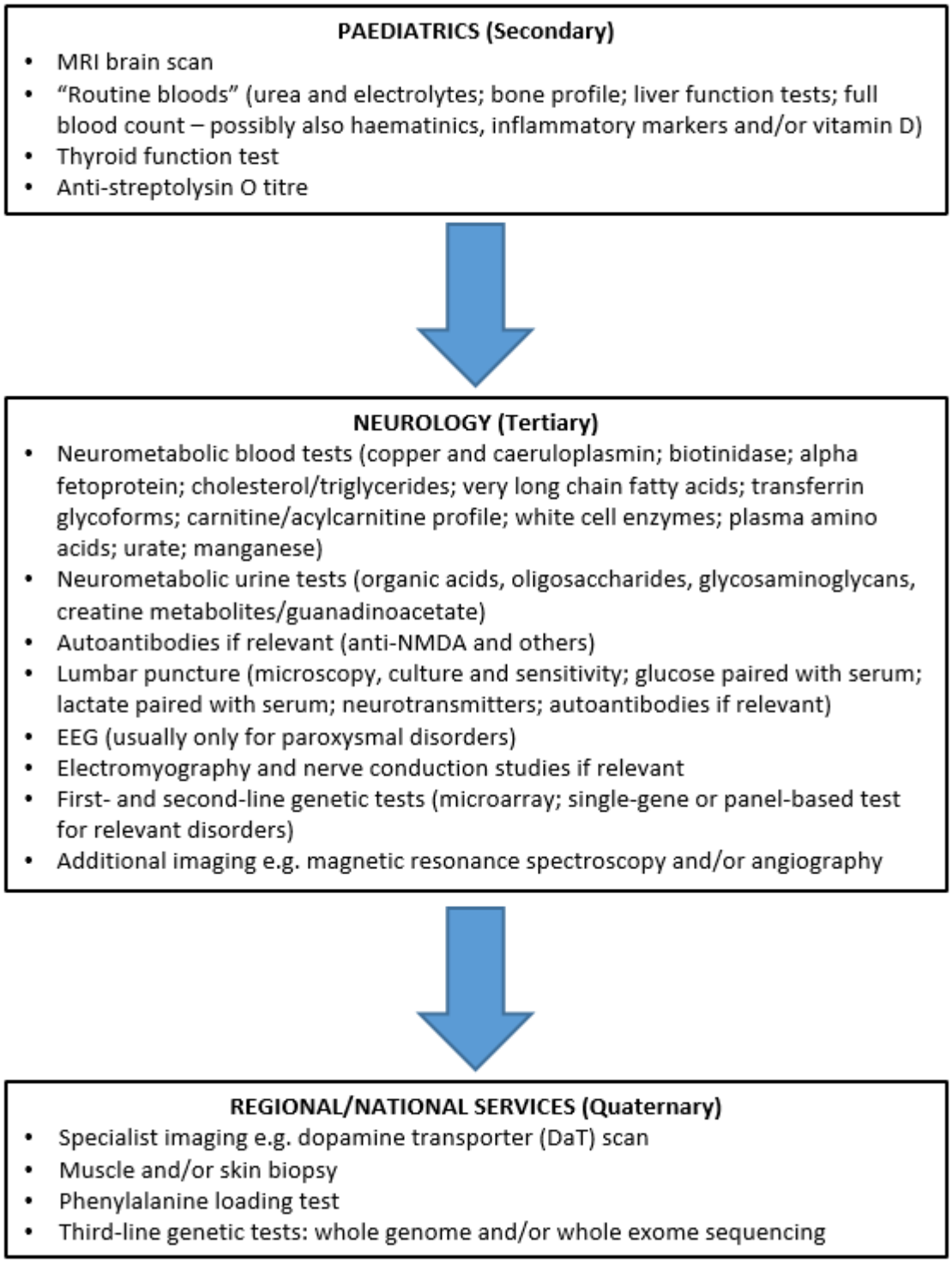
1.4.1 Typical patient pathways

A paediatric patient with a movement disorder may follow a convoluted path to diagnosis. They will initially be seen by their general practitioner, who will refer to either community or general paediatrics. At this stage a first round of investigations will be undertaken, probably including a magnetic resonance imaging (MRI) brain scan. This, together with history and examination, may result in diagnosis if there is a structural or acquired brain lesion: otherwise, a referral to paediatric neurology clinic is likely.

There, a more extensive range of tests will be performed, and a wider range of disorders – including many genetic/neurometabolic conditions – will be identified. Children or young people who still elude diagnosis at this point may be referred on to a regional or national specialist service for complex movement disorders. The investigations shown in **Figure 1.4** might be typical for a child with dystonia who remains undiagnosed at each stage.

These lists are not exhaustive or rigid, but will be recognisable as reflecting the quantity of different investigations often undertaken for a complex individual case. For young children, or those with developmental disabilities, more than one episode of general anaesthesia may be required to complete these investigations and the process of collecting blood, urine and cerebrospinal fluid may be difficult and distressing.

Figure 1.4: *Flowchart indicating investigations which might be performed at each level of for a child with an undiagnosed hyperkinetic disorder*



1.4.2 How does genetic testing fit in?

In most contexts, the first genetic test offered to an undiagnosed child with a movement disorder will be a comparative genomic hybridisation (CGH) microarray. Historically,

the next stage of genetic investigation would typically involve sequencing of one or more genes of interest. In recent years, however, as the cost of broader next-generation sequencing (NGS) based tests has converged with that of individual gene tests, this approach has become less popular. There is still a role for testing individual genes where there is a strong clinical suspicion of a certain syndrome. In particular, multiplex ligation-dependent probe amplification (MLPA) may still offer a useful service due to its high sensitivity and ability to detect copy number and structural variants as well as point mutations, which may be missed by some NGS techniques.

Until very recently, when a genetic disorder was still suspected following a negative microarray, the clinician would be likely to proceed to a panel-based approach. This involves sequencing a list of potentially relevant genes simultaneously. Numerous panels were available for the investigations of hyperkinetic disorders: in NHS practice, the most commonly-used was the Sheffield Children's Hospital's dystonia panel,(40) but many more can be accessed commercially. Without formal standardisation between panels, individual clinicians or departments had to use their judgement as to what offered their patient the best value. In the last two years, panel testing in the UK has been largely superseded by access to diagnostic whole-genome sequencing (WGS).

Broader approaches to genetic testing – WGS or whole-exome sequencing (WES) – are still largely accessible only through tertiary/quaternary settings within the NHS. WES involves sequencing protein-coding regions of the genome only, whereas WGS also sequences intronic and intergenic regions – although in fact the existence of technically difficult DNA regions including centromeres, telomeres, GC-rich regions and prolonged repeats means that only 95-98% of the genome is included in sequencing.(41)

The advantages of WGS might therefore seem to rely on our ability to interpret variants in non-coding regions, which in many cases remains limited. However, the technique also has a number of other advantages over WES. Firstly, it provides better coverage even for exons.(42) Secondly, it allows for assessment of copy number and structural variants – although interpretation of this on a large scale may be far from

straightforward, and requires a specially-adapted pipeline.(43, 44) Thirdly, it generally includes the mitochondrial genome, which WES often does not.(45) Thus although both WES and WGS are powerful techniques, WGS can be expected to capture some diagnoses which WES will miss.

Finally, the role of specialised genetic testing techniques should not be forgotten. These include mitochondrial testing (sequencing the mitochondrial genome and/or testing for rearrangements), methylation studies, and tissue-specific genetic testing where mosaicism is suspected. These techniques are only employed where a clinician has a strong enough suspicion of a particular disorder or class of disorders to request the test in the first place, but for certain conditions – for example, Angelman syndrome resulting from uniparental disomy – they are essential. There are even a few situations, such as diagnosis of suspected ring chromosome 20 syndrome, where old-fashioned karyotyping remains the investigation of choice.(46) However, the ability of WGS to capture structural variants which previously required special testing continues to improve, especially with the advent of long-read sequencing.(47)

The start of my PhD project fell within a hiatus in the availability of broad-based genetic testing in the UK. The 100,000 Genomes Project, a large-scale research project run by Genomics England offering WGS to people with undiagnosed rare disorders, closed recruitment in 2018.(48) From early 2020, WGS became available through the NHS, initially limited to specific circumstances such as critically unwell children with a suspected monogenic disorder, but increasingly to a wider patient population. Between 2018 and 2020, it was not possible to access WGS through the NHS: WES, with analysis of virtual panels of genes known to contribute to relevant disease phenotypes, was available in limited circumstances through the Exeter Clinical Laboratory International and a small number of other specialised services, including Great Ormond Street Hospital's (GOSH's) own "GOSHome". Thus research studies such as ours were, for a while, the only route to WGS for many undiagnosed families in the UK.

1.4.3 Technical aspects of NGS

In this project, WGS was entirely outsourced from the stage of DNA purification to the return of data files, but it may be useful to give a brief idea of the process. NGS refers to a variety of methods which use “massively parallel sequencing”: sequencing of very large numbers of DNA molecules at once.

The two companies we employed at different stages used different NGS platforms: DNBSeq™ and Illumina NovaSeq 6000™. Both processes start by generating a DNA library: fragmenting DNA into short lengths which are sorted by size. The fragments are then “end-repaired”, a process which allows adapter molecules to attach, usually by attaching a single adenine base to the ends of each strand for the adapter to bind to. After adding the adapters (whose nature depends on which process is to follow), the fragments are amplified using the polymerase chain reaction (PCR).

DNBSeq™ then uses a technique called DNA nanoball sequencing.(49) A second adapter called a splint oligo is added which forms the single-stranded DNA into circles. A specialised polymerase then amplifies the circular template into a single long strand consisting of the target sequence repeated multiple times: this is called rolling circle replication. Because the original strand is used as the template for each individual ball, there is less chance of replication errors being amplified. The strand automatically folds into a tight “nanoball”. The nanoballs are added to a flow cell: because they are negatively charged, they repel one another and can be reliably positioned on a positively-charged surface pattern.

The sequencing reaction then takes place: primers are added to bind the adapter sequences of the nanoballs, and fluorescently labelled, reversibly inactivated nucleotides are pumped over the cell, unbound nucleotides are washed off, and the cell is imaged under laser excitation to determine which base has attached. The fluorescent label and inactivator are then removed and the process repeated for the next base pair.

Illumina dye sequencing platforms, including NovaSeq, use a slightly different approach.(50) The adapters added to the DNA fragments contain a sequence which

binds to oligonucleotides fixed to nanowells in the flow cells. The bound fragments are amplified by a process called bridge PCR: the adapters at each end each bind a fixed oligonucleotide and bend the strand into a bridge shape. Primers can bind the adapters and replicate the strand, after which the two separate and both are left bound at one end. This generates “clusters”: thousands of vertically-bound copies of the fragment like a little forest. Sequencing then uses a process similar to that described above. In both cases, thousands of points in the genome are sequenced simultaneously.

Nanoball sequencing and Illumina dye sequencing are just two among several available techniques. Others include pyrosequencing (using an enzyme which generates light when each new base pair is bound);(51) ion semiconductor sequencing (which works by detecting ions released during DNA sequencing)(52) and nanopore sequencing (in which DNA strands are drawn through nanopores in a membrane by electrophoresis, and as each base passes through it generates tiny changes in the electrical properties of the pore).(53) The last in particular offers advantages in terms of long read sequencing which, by allowing direct reading of long repetitive sequences instead of computational reconstruction, would improve the ability of WGS to detect short tandem repeat expansions.

The usual output of NGS, regardless of technique, is a FastQ file: a file containing the full genetic sequence with each letter listed together with a quality score encoded using a single character. Using a range of bioinformatics software, these are aligned to the reference genome, generating BAM (Binary Alignment/Map) files, and undergo further processing including variant calling, to generate VCF (Variant Call Format) files, which specifically list variations from the reference sequence.

1.4.4 The utility of genetic diagnosis

There are instances where diagnosis of a rare genetic disorder may transform a patient’s therapeutic options. In the field of paediatric neurology, advances in disease-specific therapies over the last 10 years include enzyme-replacement therapy (for example, for late-infantile neuronal ceroid lipofuscinosis due to *CLN2* variants,(26) bone marrow transplantation (such as for X-linked adrenoleukodystrophy), antisense

oligonucleotide-based treatments (such as nusinersen for spinal muscular atrophy(54)) and gene therapy (as for aromatic L-amino acid decarboxylase (AADC) deficiency(55)). These life-changing treatments are available only where a genetic diagnosis has been confirmed – and, often, only if it has been confirmed in a timely manner while the patient is still at an age and a stage of disease progression where they are likely to benefit.

There are other instances where, although a direct intervention to modify the pathophysiological basis of the disease is not currently possible, a precise diagnosis can still help direct management. Examples from within the field of hyperkinetic disorders include dystonia secondary to variants in *TOR1A, KMT2B, SCGE* or, in a more palliative context, *GNAO1* and *PANK2*, which are particularly amenable to deep-brain stimulation,(56-58) or the spectrum of disorders due to GLUT1 transporter defects (*SLC2A1* variants) which are successfully managed using the ketogenic diet.(59)

Even where a diagnosis cannot guide treatment, it has other benefits. These include permitting genetic counselling and accurate prognostication. It also brings to an end a diagnostic odyssey which in some cases will have lasted for years and involved significant invasive procedures.

It has also been demonstrated that having a specific diagnosis may make it easier for patients and families to access healthcare and support services.(60) While this is regrettable, as ideally such services should be provided on the basis of need rather than diagnostic label, it can represent a significant benefit of diagnosis from the point of view of the individual family.

A further benefit which may be highly valued by families is the option to form links with condition-specific support groups. These have proliferated to a striking extent over the last few years, with support groups – often emerging on social media before becoming more formalised with website and real-world meetings – now existing for a wide range of ultra-rare genetic disorders. Unregulated social media groups do carry a risk of providing inaccurate and occasionally harmful information, but they also have the

potential to provide not just mutual support but a forum for sharing information and experience, as well as a rallying-point for a nucleus of highly motivated parents and carers to advocate for services and research. This is especially valuable for those affected by conditions so rare that the patient's own doctor is unlikely to come across more than one case in their working life.

Finally, diagnosis of rare disorders is essential for research. Our ability to study and understand a rare disorder is inevitably constrained when the worldwide cohort of diagnosed individuals is only in double figures: in this situation, each new family offers the potential to increase our knowledge of the phenotypic or genotypic spectrum. Moreover, identification of affected people opens up possibilities for basic and clinical studies, only practical where a sufficient number of families are willing to offer their information, samples and ultimately trial participation. Diagnosis is of course only the first step towards translational research, but it is an essential one.

1.5 "Gene discovery" – what, how and why

1.5.1 What is "gene discovery"?

"Gene discovery" is, of course, a useful shorthand for what should more accurately be called the identification of new gene-phenotype (or specifically gene-disease phenotype) relationships. Although some new genes probably do still await literal discovery, the vast majority of the estimated 25,000 human protein coding genes have been recognised and sequenced. For most, their products have also been experimentally characterised, although a number of unstudied open reading frames (ORFs) do still remain and the latest sequence of the human genome includes nearly 100 new protein-coding genes.(61)

In the context of genomic research on rare diseases, "gene discovery" means establishing a robust causal link between variants in a known gene and specific pathological phenotypes. In some cases, this may include identifications of second (or third) distinct disease phenotypes for genes which are already known to have one.

Historically, gene identification relied on linkage analysis(62) and/or autozygosity mapping.(63) Both these techniques are most effective where data from multiple affected members of a large family is available, so the study of rare sporadic diseases was very challenging. NGS approaches, allowing for broad-based analysis of genome-wide datasets, has made it far easier to identify *de novo* variants, as well as other ultra-rare disorders where data from multiple family members is unavailable. (It may, alas, also have increased the number of reports of gene-disease associations which ultimately prove to be spurious.(64))

Discovering a causative genetic fault is, of course, only the very first step in the process of understanding the pathophysiology of a disorder. Although in some cases there may be a clear intuitive link between an aberrant gene product and a clinical manifestation – for example, reduced activity of tyrosine hydroxylase leading to failure of dopamine and serotonin synthesis and resulting movement disorders in TH deficiency – in others this is far from the case. Why, for example, should abnormalities in *KMT2B*, a histone methyltransferase involved in regulating the expression of large numbers of other genes, lead specifically to a progressive dystonia? Nevertheless, identifying the peccant gene is an essential first step and provides a firm basis for further theoretical and experimental work.

This thesis will deal only with Mendelian genetics – that is, situations where a simple one-to-one correlation can be established between a gene and a phenotype, at least approximately speaking. This is not to devalue the role of other approaches such as genome-wide association studies, aimed at identifying polygenic contributors to disease: it is rather that the data available to me – detailed phenotypes and whole genomes of a few hundred individuals with rare disorders – lends itself to this approach.

1.5.2 Gene discovery in movement disorders

The very first human disease gene to be mapped to a specific chromosomal locus belonged to a movement disorder: mapping of *HTT,* the gene implicated in Huntington's disease, in 1983,(65) marked the start of a new era, even though the gene could not be characterised for another 10 years.(66) In those early years gene

identification was a matter of painstaking linkage analysis involving multiple affected family members: by modern standards progress was glacial, but there was a steady trickle of successes, including *TOR1A* (the gene for DYT1, mapped in 1990(67) and characterised in 1997(68)) and *ATXN1* (for SCA1, mapped in 1993(69) and characterised in 1994.(70))

As sequencing technology progressed, it became possible to characterise a gene soon after identification, and in particular the availability of NGS was a great leap forward for gene discovery – for example, 16 of the 26 known genes for DYT-numbered conditions were identified since 2010, all using NGS approaches.(58, 71-82) I started my project in the midst of a brief lull in identification of new DYT dystonias – the latest, DYT-28 due to *KMT2B* variants, having been reported in 2016(58, 74) – but since then seven more have been added. Over the same period, an even larger number of genetic conditions were reported in which a movement disorder forms one feature in a complex or syndromic presentation.(83) Today, literally hundreds of monogenic causes of movement disorders have been described, and more are published every month

Besides identifying previously blameless genes as causes of movement disorders, there has also been a growing recognition of significant phenotypic pleiotropy in genes with known disease phenotypes. For example, increasing numbers of epilepsy-dyskinesia syndromes have been described in genes which were previously known as causes of epilepsy (or developmental and epileptic encephalopathy) only.(84) In some cases, pathogenic variants in “epilepsy” genes have been reported in patients who have a movement disorder but never develop seizures at all, such as *GNAO1, RHOBTB2* and *GRIN2A*.(85-87)

1.5.3 Why haven’t we found them all already?

Genomics is still a relatively new science and has made huge advances over the last two or three decades, but some suspect that we may be reaching a period of diminishing returns. Most importantly, a high proportion of people with rare disorders remain undiagnosed after maximal investigation. In the field of movement disorders

the numbers are particularly discouraging: in one recent study, over 85% of people with dystonia of probable genetic aetiology remained undiagnosed after WGS.(88)

What, then are we missing, and why are we missing it? The answer is probably several-fold: I have summarised some of the key factors in **Table 1.3**.

Table 1.3: *Why disease-causing genes may be hard to identify, and how this could be addressed. Table adapted from Steel and Kurian, 2020.(83)*

Reason disorder is difficult to identify	**What might help?**			
	Very large/ population-wide datasets	**More sophisticated statistical/bioinformatics analysis**	**Improved understanding of biological pathways**	**Improved sequencing/variant calling technology**
Digenic inheritance	✓	✓	✓	X
Polygenic inheritance	✓	✓	✓	X
Incomplete penetrance	✓	✓	✓	X
Phenotypic pleiotropy	X	X	✓	X
Structural and copy number variants	X	✓	X	✓
Intronic variants	X	X	✓	✓
Variants in intergenic regions	X	X	✓	✓
Extreme rarity	✓	✓	✓	X
Mosaicism	X	✓	X	✓

The issue of detecting structural and copy-number variants is a particularly challenging one: although they can be detected from WGS datasets, the analysis pipeline for assessing their impact is far less straightforward than that for single-nucleotide variants (SNVs) and small insertions/deletions.(43, 44) Meanwhile, the reassurance provided by a normal microarray is only partial: CGH microarray usually offers resolutions of around 100kbp,(89) whereas a typical gene is only 10-15kbp in length.

In the field of movement disorders, the poor sensitivity of WES/WGS for triplet-repeat conditions (including a number of the spinocerebellar ataxias) is of particular concern, although novel bioinformatics approaches may be able to remedy this.(90)

Thus while existing techniques – NGS, linkage analysis and autozygosity mapping – will continue to play an important part, if we are to achieve diagnosis for a greater fraction of our patient cohort we will also need to embrace new approaches and technologies.

1.5.4 How can we know when we have found a new gene?

When a patient with a rare monogenic disorder undergoes WGS then, if the data is of good quality, their diagnosis is probably to be found somewhere within it. This does not mean, sadly, that we can always identify it. Any individual genome contains many thousands of variants, many of which could theoretically be relevant to their clinical phenotype. The trick is to determine which of those thousands is the true culprit – and seeking to answer that question will form the kernel of this thesis.

When seeking novel disease genes in a cohort of patients with rare disorders, a series of logical steps can be followed. These can be framed as a series of questions:

Question 1: Does this individual have a genetic disorder at all?

It is desirable to recruit only participants where (as far as possible) non-genetic causes of disease have been excluded. In movement disorders, however, this means excluding individuals with a history and/or scan findings suggestive of hypoxic-ischaemic encephalopathy (HIE), kernicterus, periventricular leukomalacia (PVL) or other acquired insults, or those where there is a strong suspicion of an autoimmune, post-infective or functional disorder. This approach cannot hope to be fool-proof, at least not without excluding some children in whom there is a reasonable suspicion of genetic disorder and who might therefore benefit from testing: for example, some genetic conditions present with a neonatal encephalopathy strongly resembling HIE.(91) (It is also perfectly possible for an individual to have dual diagnoses of cerebral palsy due to a perinatal insult and a genetic movement disorder: in Chapter 3 I will describe a girl with PVL in whom we found a pathogenic *SPAST* variant.)

Nevertheless, careful case-by-case assessment of the likelihood of a genetic aetiology can at least increase the pre-test probability of a genetic aetiology.

Question 2: Does this individual have a known genetic disorder?
The next step is to rule out, as far as possible, existing causes of disease. Therefore before any analysis of interesting novel variants, it is necessary to check each proband's genomic data thoroughly for potentially pathogenic variants in known relevant disease genes.

If the answers to questions 1 and 2 are "yes" (or "probably") and "no" respectively, the search for a new gene can begin in earnest.

Question 3: Does this individual's genome contain variants which are plausible candidates for pathogenicity, in the context of their phenotype?

Every individual will have a large number of variants, including some unique variants, but of course the vast majority of these will be benign. Features which would support a variant's candidacy as a cause of disease would include:

- Rarity (for heterozygous variants, preferably absent from population databases; for biallelic variants homozygotes should be absent). "Maximum credible population allele frequencies" can be calculated for dominant and recessive disorders with known population prevalence and penetrance,(92) but where a large reference dataset is available the absence of homozygotes is a reasonable proxy.

- Consistent inheritance (i.e. for an affected child of healthy parents, should be *de novo*, X-linked (if proband is male) or two variants in *trans*). Variably penetrant variants, while important, will always be harder to demonstrate and are more likely to come to light through studies using larger cohorts and computational approaches such as burden testing than through analysis of individual genomes.

- *In silico* tools predict a damaging effect on the gene product.

- Physiological function and tissue expression of the gene product is consistent with the phenotype. In some cases, this may be clear-cut – for example, the potential for the midbrain-expressed dopamine reuptake transporter gene *SLC6A3* to contribute to a phenotype of abnormal dopamine turnover is obvious – but in others, such as where ubiquitously-expressed "housekeeping" genes are involved, it is much less so.

Where a variant meets these four criteria it may be worth further attention. It is still, however, very far from being a proven cause of disease, leading on to the fourth question.

Question 4: Can the pathogenicity of the preferred variant be confirmed?

At least two authoritative bodies publish proposed criteria for assessing the pathogenicity of variants: the Association for Clinical Genomic Science (ACGS) in the UK(93) and the American College of Medical Genetics and Genomics (ACMG) in the USA.(94) Their guidelines are closely overlapping but non-identical. Both result in a classification of pathogenicity from benign (formerly Class 1) to pathogenic (formerly Class 5): illustrative examples of how this can work in practice are shown in the box below. Both systems were designed primarily for the classification of variants in known disease-causing genes: if used on novel disease genes, various inconsistencies emerge, as might be expected when using a tool outside its intended purpose. Nevertheless, they provide useful indicators of what types of evidence are considered to be supportive of pathogenicity. I will discuss these systems in more detail in section 3.3.

Example 1: A pathogenic ("class 5") variant

SP10P was a six-year-old girl of African descent with no relevant family or perinatal history. After a normal infancy, she started to crawl asymmetrically at 12 months and when she began to walk at 18 months she tiptoed. On examination she had truncal sway and occasional involuntary movements, together with several small café-au-lait spots.

Research WGS revealed that she was compound heterozygous in trans *for two variants in* ATM. *The maternal variant was previously published as pathogenic; the paternal variant (c.6081dup, p.Gln2028Thrfs*5) was novel. The paternal variant could be classified as pathogenic because a) it resulted in a frameshift; b) it was absent from normal population databases and c) it was in trans with a variant known to be pathogenic. A diagnosis of ataxia-telangiectasia syndrome was given.*

Example 2: A likely pathogenic ("class 4") variant

DY206P, a five-year-old White boy with no relevant family history, presented with global developmental delay and regression, axial hypotonia and mixed findings of spasticity and dystonia in all four limbs, and epilepsy. MRI showed progressive pontocerebellar hypoplasia.

Research WGS found that he was compound heterozygous in trans *for two variants in* SEPSECS, *the paternal one an in-frame deletion (c.776_784del, p.Ser259_Lys261del) and the maternal one a non-canonical splicing variant (c.388+3A>G, p.?). Neither was previously reported in disease and neither by itself met the criteria for "likely pathogenic". However, the GOSH diagnostic lab was able to undertake a splicing assay on blood-derived RNA demonstrating that the non-canonical splicing variant did indeed disrupt splicing, resulting in the loss of 19 bases from the mRNA transcript and a predicted frameshift (r.270_388del, p.Phe91Cysfs*47) allowing it to be reclassified as pathogenic. This in turn permitted reclassification of the paternal variant – now in* trans *with a pathogenic variant – as "likely pathogenic" and a diagnosis of pontocerebellar hypoplasia type 2D was made.*

Example 3: A variant of uncertain significance ("class 3"), pathogenicity suspected

CD66P, a five-year-old Indian boy with no significant family or perinatal history, presented with severe global developmental delay from infancy. He had truncal hypotonia with brisk deep tendon reflexes and intermittent abnormal eye movements. By four years old, a complex hyperkinetic movement disorder with dystonia, chorea and myoclonus had

evolved. MRI brain at one year old was unremarkable and diagnostic WES in the child's home country showed no pathogenic variants.

Research WGS identified a hemizygous maternally inherited variant in CASK *as follows: c.689T>C, p.Ile230Thre. CADD score was 26.7. The child's presentation was felt to be consistent with the* CASK *spectrum and the variant had the following criteria in favour of pathogenicity: a) absence from normal population databases; b) missense variant in a gene with a low rate of benign missense variation (where missense variants are a recognised mechanism of disease); c) predicted to be pathogenic by multiple lines of computational evidence. This was not sufficient to permit re-classification of the variant and it was reported to the referrer as being of uncertain significance. Classification of X-linked recessive variants is often challenging: this variant might be reclassified if it was found to have occurred* de novo *in the mother, and/or specific biomarkers for a* CASK*-related disorder could be identified in the proband.*

Evidence which might justify publishing a variant as a putative novel cause of disease could take two main forms: firstly, experimental functional work could provide strong support for defects in the gene giving rise to the phenotype, or secondly, variants in the gene could be reported in multiple unrelated individuals with closely overlapping phenotypes. If at all possible, both criteria should be met. Indeed, a novel gene is not accepted as a proven cause of disease until a second affected family has been identified.(64) In order to feel confident that a new gene-phenotype relationship has been identified, it will clearly be preferable to draw on evidence from both the clinic and the laboratory.

1.6 Summary

Movement disorders, then, are a diverse range of conditions in which – especially in children – one or more types of movement symptom may co-exist with a range of other neurological, neurodevelopmental and systemic impairments. Although individually rare, they are a major cause of disability and individual distress with some even increasing the risk of premature death. Many are genetic in origin and of those many

remain either difficult to diagnose or, at present, literally impossible because the gene in question has yet to be identified.

WGS is on the threshold of becoming integrated into mainstream clinical practice and is also a vital tool for research, including gene discovery. However, like any test the output of which is a large and complex dataset, techniques for analysing and digesting the data are as essential as those for generating it. In the field of rare disorders, including hyperkinetic disorders, these techniques include developing our understanding of disease phenotypes, in particular understanding of phenotypic pleiotropy, to gauge the relevance of potentially pathogenic variants.

In this project, I have aimed both to contribute to our knowledge of previously undescribed genetic disease entities in the field of movement disorders, and also to build on what we have started to learn about existing but imperfectly understood rare conditions.

Chapter 2. Methods

2.1 Recruitment

Patients were recruited from two sources: the Great Ormond Street Neurogenetic Movement Disorders Clinic (NMDC), and external referrals.

2.1.1 Neurogenetic Movement Disorders Clinic

The NMDC, run by Prof Kurian, takes place approximately three times per month and provides a national and occasionally international reference service. Patients are referred by their consultant – usually a paediatric neurologist or clinical geneticist – if they are suspected to have a genetic movement disorder and initial investigations have not yielded a diagnosis. In line with the referral criteria at Great Ormond Street Hospital (GOSH), all patients are 16 years old or under at the time of their initial review, but some may be followed up into early adulthood.

A small number of patients needing highly specialised care, such as those with neurotransmitter disorders requiring medications rarely used in paediatric practice, are followed up in the clinic on a long-term basis, in collaboration with their local neurology and paediatric services. Some children and young people are also seen following a new diagnosis of an ultra-rare condition such as a disorder of neurodegeneration with brain iron accumulation (NBIA) in order to discuss management, prognosis and opportunities for involvement in research.

For most patients, however, the clinic primarily fulfils a diagnostic role and aims to complete assessments in a single visit. Consultations last about an hour and comprises a detailed history – family history, details of pregnancy, birth, development and medical background, as well as history of the presenting symptoms – followed by general and neurological examination and review of videos if available. The child's previous investigations are also reviewed, including re-reporting of any relevant external imaging by a paediatric neuroradiologist.

At the end of the consultation, a plan for further investigation is made, together with recommendations to the local team for management in some cases. If a genetic disorder is thought likely and the individual has not already had WGS, this is recommended as the next step. During most of my project, this meant offering recruitment into the Study of Inherited Metabolic Diseases (SIMD). If the family wished to proceed, DNA samples and written informed consent were collected immediately after the consultation, or at a later date if the family wanted time to consider the option.

2.1.2 Genomes and exomes from other sources

A proportion of patients seen in the NMDC had already had either WES or WGS. WES had usually been undertaken through the Deciphering Developmental Disorders Study(95) and in these instances recruitment to SIMD for WGS was still offered in view of the chance of new findings, both through improved coverage and the recognition of new disease genes since the conclusion of the study in 2015.

Some patients had had WGS through the 100,000 Genomes Project.(48) In this case I did not usually plan to undertake repeat sequencing, but I did offer patients recruitment to SIMD in some cases in the hope of being able to re-analyse their data. Accessing individual identifiable patient genomes through the Genomics England Research Environment is not permitted (although it is in fact technically possible in some circumstances). The Clinical Genetics department of GOSH supported my application for access to the portal they use for review of individual genomes (Clinical Interpretation API, or CIP-API) for clinical purposes, but this only allows access to genomes of patients recruited at GOSH itself. Given the NMDC's role as a national referral service, this only covered a small minority of our patients. At the time of writing my attempts to access genomes of patients recruited at other centres is still in progress. These genomes will be reviewed from a clinical perspective only – that is, for variants in genes with recognised disease phenotypes.

From early 2022, diagnostic WGS became available through the NHS Genomic Medicine Service. There is not, however, currently any method for individual clinicians to access unprocessed genomic information (VCF files) for patients who have had

WGS through this service, so detailed re-review of the data is not possible. It is to be hoped that this may be resolved in future.

2.1.3 External referrals

External referrals from clinicians within the UK are generally processed as clinical enquiries, with clinical advice, where required, being offered together with the opportunity of recruitment into SIMD if appropriate. In some instances this ultimately leads to the patient being seen in person in the NMDC, but in others – especially where travel is challenging – samples and consent are collected remotely.

Referrals from overseas come from a range of sources, largely through a network of clinical and academic contacts within the movement disorders community. Because in some medical systems neurologists' practice may include both adult and paediatric patients, these referrals include a small number of adults. We were also fortunate in being able to circulate a flyer describing the project to an online group of paediatric neurologists and neuroradiologists in India, which led to several additional referrals.

2.2 Ethics

SIMD is approved by the Bloomsbury Research Ethics Committee, reference 13LO168. It serves as an umbrella under which several different projects – all concerning the investigation of rare inherited disorders – take place. The consent forms and family information leaflets are standardised between these projects.

Some parts of this thesis involved retrospective collection of data from clinical records for patients who were not recruited into SIMD or another research study. As these were patients known to the NMDC, for which I am a member of the clinical team, this fell within the remit of the institutional research and development board rather than requiring formal Research Ethics Committee approval. This data collection was approved by the Great Ormond Street and Institute of Child Health Joint Research and Development Board, reference 19NM23. Where publication of any recognisable images or videos took place, separate written informed consent was obtained.

The SIMD protocol makes provision for collaboration with international colleagues. Although compliance with local legal and ethical requirements was considered to be the responsibility of the referrer, referrals were only accepted where evidence of the family's written informed consent for participation in genetic research could be provided.

2.3 Phenotyping

Utilisation of detailed phenotypic information for the analysis of genomic data was a key part of the project, as it enabled a) the use of a very large gene panel without generating unmanageable numbers of variants of uncertain significance (VUSs) and b) better prioritisation of variants of potential interest in genes without known disease phenotypes. Thus gathering such information was of the first importance.

The dataset ideally required was as follows:

History:

- **Family history:** history of neurological disorders, consanguinity, and ideally a full pedigree.
- **Antenatal and birth history:** in some instances this provided information about age of onset (as for example with a history of reduced or increased foetal movement); in others it hinted at a non-genetic origin of the movement disorder, as with perinatal events such as asphyxia or severe jaundice.
- **Medical background** including any history of serious illness and any regular medication use.
- **Developmental history** including timing of developmental milestones, any history of regression, and presence or absence of intellectual disability.
- **Onset and progression of movement disorder** including nature, time course, anatomical distribution, exacerbating or relieving factors and current status.

Examination:

- **Dysmorphic or distinctive features** if any;

- **Systemic features** such as organomegaly or abnormal cardiorespiratory findings;
- **Full neurological examination** with an emphasis on tone, posture and coordination (including ocular and bulbar movements);
- **Video recording** of paroxysmal episodes or exacerbations of fluctuating disorders where these could not be observed directly.

Investigation:

- **MRI brain scan** was the only investigation considered mandatory in all cases before considering enrolment, due to the relatively high diagnostic yield of this test.
- **Results of all other tests completed** including neurometabolic investigations (blood, urine, lumbar puncture), neurophysiology (EEG, nerve conduction studies, electromyography, VER/ERG), imaging, muscle/skin biopsy and genetic investigations.

In the case of patients seen in the NMDC, this information could be collected in person. For other patients, it was gleaned from referral letters and accompanying materials, if sufficiently detailed, but I also developed a proforma to guide referrers towards the data we needed (see **Appendix 1**).

The full dataset was not available for every proband, but those who had a comparatively full dataset were prioritised for more detailed analysis, as more likely to yield a diagnostically or scientifically useable result.

2.4 The Movement Disorders Database

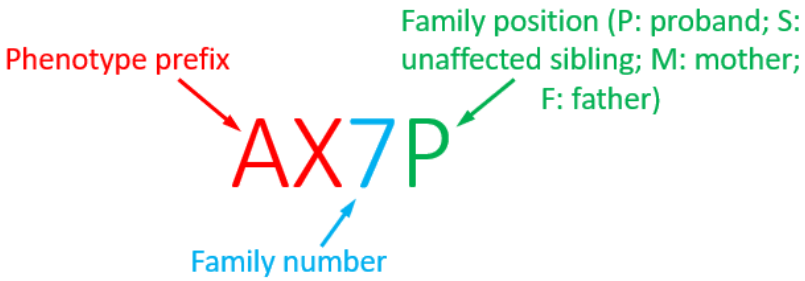
The Kurian research group already possessed a database of individuals known to us for research purposes when I arrived, and I was able to extend and refine this considerably. The database is stored on the UCL Data Safe Haven and includes patient-identifiable information alongside pseudonymisation codes and a range of genetic, phenotypic and procedural information.

To facilitate joint analyses, I classified patients into groups by their primary movement disorder phenotype as follows:

- AX: ataxia
- CD: chorea/dyskinesia
- DP: dystonia-parkinsonism
- DY: dystonia
- MY: myoclonus
- NB: neurodegeneration with brain iron
- PE: paroxysmal events
- SP: spasticity
- TR: tremor
- ZZ: unclassifiable (very complex, or patients with a movement-disorder related finding such as neurotransmitter abnormalities but no clear movement disorder).

The database currently lists details of 537 families or individuals affected by movement disorders, as shown in **Table 2.1**. Pseudonymisation codes take the form shown in **Figure 2.1.**

Figure 2.1: *Structure of pseudonymisation codes*



The ideal dataset held for each individual includes name; date of birth; referral route; consent status; disease status (affected or unaffected); location and quantity of DNA stored; previously completed genetic tests/analyses; stage of current analysis reached; and brief phenotypic details. More detailed phenotypes including free-text summaries, completed proformas and video materials are stored elsewhere within the

UCL Data Safe Haven. Where sensitive and/or patient-identifiable information is available only in hard copy (i.e. written correspondence), data are stored securely in a locked filing cabinet.

Table 2.1: *Families currently listed in the movement disorders database*

Phenotype	No. of families
Ataxia	24
Chorea/dyskinesia	72
Dystonia-parkinsonism	38
Dystonia	239
Myoclonus	18
NBIA	39
Paroxysmal	40
Spasticity	26
Tremor	8
Unclassifiable	33
Total	**537**

2.5 Sample collection and processing

2.5.1 Sample collection

DNA was collected from circulating lymphocytes derived from peripheral blood samples, collected in EDTA. For samples obtained outside GOSH, we were able to accept either EDTA blood samples or extracted DNA. Either could be transferred at room temperature provided transit time (in the case of blood) did not exceed one week. Blood samples were stored at 4°C and extracted DNA at -20°C.

Fibroblasts were obtained by skin biopsy undertaken by the proband's own clinical team. Samples were sent to us as skin in fibroblast growth medium with added penicillin and streptomycin to reduce bacterial growth, shipped at room temperature.

2.5.2 DNA extraction

DNA extraction was either performed by GOSH genetics, or in house using the Promega ReliaPrep™ Blood gDNA Miniprep System in accordance with the manufacturer's instructions. Briefly, blood samples anticoagulated with EDTA were thawed and thoroughly mixed using a rotisserie shaker. 200µl of blood was then mixed with 20µl Proteinase K and 200µl cell lysis buffer and incubated for 10 minutes at 56°C. 250µl per sample of Binding Buffer (containing a chaotropic salt) was added to the lysate and the whole was transferred to a binding column and centrifuged for 1 minute at 14,000g. The flowthrough was discarded and the binding column placed in a new collection tube and washed three times by centrifuging at 14,000g for three minutes with 500µl of Column Wash Solution, discarding the flowthrough after each wash. DNA was then eluted into a clean collection tube by centrifugation at 14,000g for 1 minute with 50µl nuclease- and nucleic acid-free water. DNA samples were pseudonymised using their database codes for storage. DNA concentration was measured using the NanoDrop™ spectrophotometer according to the manufacturer's protocol. DNA quality was confirmed by electrophoresis on a 1.5% agarose gel made using 1X tris-borate EDTA (TBE) buffer, stained with SYBR safe™ 1.5µl in 100ml, run for 30 minutes at 100V, before samples were sent for sequencing.

2.6 Whole exome/genome sequencing pipeline

2.6.1 Sequencing and variant calling

Whole exome/genome sequencing was performed externally by a commercial service. Initially using BGI Genomics (www.bgigenomics.co.uk; Hong Kong; UK offices in London), based on recommendation by GOSgene. However, problems with the accuracy of the variant calling came to light (Chapter 8), leading us to switch to Novogene (www.novogene.com; Beijing, UK offices in Cambridge) from late autumn 2019.

I will not discuss the reasons for this change in detail here, but in brief I became aware – from unexpected Sanger sequencing results – that some relevant variants were not being called (false negatives) in the VCF files from BGI, although they were present as expected in the Fastq and BAM files. Hence it was possible to reanalyse "raw" data

from BGI without full resequencing: some of this was kindly undertaken by the UCL Queen Square Institute of Neurology Genomics Unit; some by Novogene. This reanalysis confirmed our impression of significant discrepancies.

We have not identified any concerns regarding false negative findings in data from Novogene. There was a difficulty regarding large numbers of false positive calls, again involving only VCFs rather than other data files – again, I am unable to discuss this in depth here, but the problem appears to have been resolved with a modification to the variant calling process.

BGI performed sequencing with the DNBSeq^{TM} NGS technology platform, using paired-end 100bp reads to a depth of 30X. Reads were aligned to the GRCh37/hg19 reference genome. Variant calling was undertaken using SAMtools,(96) SOAPsnp(97) and GATK.(98) Data was returned to us as both BAM files and VCF files for each individual. After concerns about accuracy of variant calling emerged, we sought re-analysis of the BAM files (whose accuracy we had no reason to doubt) from the team at the Neurogenetics Laboratory of the Institute of Neurology, Queen Square, London, or from Novogene.

Novogene used the NEBNext Ultra II DNA Library Prep Kit for generation of sequencing libraries, then performed WGS using the Illumina Novoseq 6000 platform PE150, i.e. paired-end 150bp reads. Reads were aligned to the GRCh38/hg38 reference genome using Burrows-Wheeler Aligner to generate BAM files. Variant calling took place using SAMtools and GATK, plus DELLY(99) for structural variants and Control-FREEC(100) for copy-number variants. Detected variants were annotated relative to RefSeq(101) transcripts using ANNOVAR.(102)

BAM files were consulted directly using the Broad Institute's Integrative Genomics Viewer (IGV) where it was necessary to check for apparent copy number variation of loss of heterozygosity, usually in the context of seeking a second variant for an autosomal recessive condition where a single pathogenic variant had been found. For the Novogene data, structural and copy-number variants were not provided in a form which could be uploaded to our filtering software, so these were used exclusively for

manually checking second variants and other genes with a pre-existing high index of suspicion.

Variants discussed in this thesis are given using build GRCh38/hg38: those provided by BGI in GRCh37/hg19 have been converted.

2.6.2 Filtering

Initial stages of variant prioritisation were performed using QIAGEN Ingenuity Variant Analysis™, and later an updated version, QIAGEN Clinical Insight™. This software allows VCF files to be uploaded for analysis and then multiple layers of variant filtration to be applied. The exact filter settings varied depending on the aim of the analysis but broadly included those listed in **Table 2.2**. (The settings specified are typical but were adjusted when required, for example when seeking low-confidence or low-allele fraction variants in a particular gene of interest, or seeking a second variant in an autosomal recessive gene where a single pathogenic variant had already been found.)

Initially healthy unrelated individuals were used as additional controls but it became apparent that this did not add any extra information beyond using population allele frequency cut-offs, so was discontinued.

As a minimum, three analyses were conducted for each proband, looking respectively for homozygous, compound heterozygous or heterozygous variants. For probands where triome genomic data was available, the initial heterozygous variants analysis focused on *de novo* changes only. If no clearly pathogenic variant had yet been found, and inheritance with variable penetrance was considered plausible – i.e. excluding very severe, very early onset phenotypes, where inheritance from a wholly unaffected parent is much less likely – an analysis for inherited heterozygous variants was then added. (For singleton probands and those with DNA for only one parent available, of course, the analysis could not distinguish between *de novo* and inherited variants.)

First-pass analysis of heterozygous variants, analysis of inherited heterozygous variants, and any analysis of heterozygous variants in singleton probands was guided by the use of a virtual gene panel (see below). Analysis of biallelic variants was usually

panel-free from the start, and for triomes panel-free analysis of *de novo* variants was performed in all cases where no clearly pathogenic variant had yet been identified. For families with multiple affected individuals, panel-free analysis of heterozygous variants shared by affected individuals was standard except where a diagnosis was immediately obvious on the more limited analysis.

Table 2.2: *Filters for variant prioritisation with QIAGEN software*

Filter	Usual settings and notes
Confidence	Call quality at least 20 Read depth at least 10X Allele fraction at least 15% (except where specifically looking for low-level mosaicism) Outside the 5% most exonically variable 100 base windows in healthy public genomes.
Common variants	Biallelic variant analyses: population minor allele frequency (MAF) >1% in the gnomAD(103) dataset (>3% if seeking second variants in AR genes of interest) Heterozygous variant analyses: MAF >0.01%
Predicted deleterious	Exonic variants predicted to result in a frameshift, indel, missense or change to a start/stop codon Variants <20 bases into an intron Variants predicted to result in splicing changes Any variant with CADD >15
Genetic analysis	Specifies homozygous, compound heterozygous, hemizygous and/or heterozygous variants depending on analysis Can exclude variants present in homozygosity or heterozygosity in one or both parents (or other controls)
Biological context	If required, a gene panel (list of genes of interest) can be applied: not used in gene-agnostic/panel-free analysis

2.6.3 Panel development

Despite the suggestion of complete comprehensiveness contained in the name “whole-genome sequencing”, much analysis relies on panels – lists of genes of interest. There is no single standardised panel for dystonia (or any other disorder) and a range of competing clinical, research and commercial panels may be used. Naturally these change and expand over time. As a point of comparison, widely-used panels for

dystonic disorders as of March 2020 included those offered by the Sheffield Children's Hospital, with 73 genes;(40) Centogene, with 88,(104) and Invitae, with just 23.(105) The panels used by Genomics England, whose diagnostic service became available in 2022 are considerably broader: the 'R57 Childhood onset dystonia, chorea or related movement disorder' panel contains 967 genes.(106)

At the start of my project, the Kurian Group was using a number of condition-specific panels as show in **Table 2.3**:

Table 2.3: *Condition-specific panels in use by the Kurian group at the start of my project*

Panel	No. of genes
Aicardi-Goutières syndrome-like/basal ganglia calcification	9
Alternating hemiplegia-like/paroxysmal brain channelopathies	263
Ataxia	105
Cerebellar atrophy	145
Childhood neurodegenerative disorders	1055
Chromatin-related genes	3189*
Dystonia/hyperkinetic disorders	1100
Dystonia-parkinsonism	241
Early-onset epileptic encephalopathy (sic), extended	1018
Folate-related genes	7
Hereditary ataxia	172
Hereditary spastic paraparesis	76
Leukodystrophy	138
Mitochondrial	349
Myoclonus	261
NBIA	315
Neurodegenerative disorders (adult)	19
Neuropathy	217
Neurotransmitter disorders	245
Synaptic genes	749*

**Genes were selected for the synaptic and chromatin-related panels based on physiological function rather than known contribution to pathology and many had no known disease association: these panels were therefore not in regular use.*

As my interest was in identifying ultra-rare disorders, novel phenotypes and novel disease genes, I wished to take a broader approach. This appeared particularly important in view of our increasing understanding that, rather than variants in a single gene corresponding to a single well-defined genotype, there is often a substantial degree of phenotypic pleiotropy. One well-known example of this is the growing

number of genes originally identified as causes of epilepsy which are now known also to present in some instances with movement disorders, such as *SCN8A,*(107) *TBC1D24*(108) and *KCNT1*.(109) Thus excessive focus on a too-narrowly-defined phenotypic category could lead to overlooking relevant variants in genes implicated in related but non-identical disorders.

I therefore developed a “megapanel”. This involved amalgamating all the condition-specific panels we were already using, together with additional genes from the relevant Genomics England PanelApp(110) panels. All panels relevant to movement disorders of any kind, together with those related to neurotransmitter disorders, brain channelopathies, and epilepsy, were incorporated. After removal of the very considerable overlap, this yielded a list of 3477 genes.

PanelApp rates genes as “green”, “red” or “amber” depending on the level of confidence in the scientific community as to whether variants in that gene do really cause the phenotype in question. “Green” genes have a clearly established relevant disease phenotype whereas “red” genes have limited or conflicting evidence, or a purely theoretical suggestion that they might affect relevant processes. For the purposes of the megapanel, red and amber genes were included as well as green, to give the broadest possible results while still allowing some degree of prioritisation compared with panel-free analysis.

Additional genes of potential interest were added to the megapanel as they arose, either from the work of members of the group, from new publications or from other geneticists expressing interest in particular genes.

2.6.4 Variant analysis

Following the initial filtering stages carried out using Ingenuity, a “longlist” (still comprising a few hundred variants per proband) was exported for further manual filtration. This involved several stages.

Firstly, I removed certain categories of variants, namely:

- Variants with a gnomAD(111) homozygote count >1;
- Variants from the compound heterozygous analysis known to occur in *cis*;
- All variants from certain very large and/or variable genes which came up in nearly every analysis and were thought unlikely to be relevant, including (but not limited to) mucin proteins; olfactory receptors; very large muscle protein genes such as *TTN*; and members of the leukocyte immunoglobulin-like receptor subfamily;
- Variants with a CADD(112) score <10, except those predicted to affect splicing;
- Variants in non-protein-coding genes including pseudogenes and long intergenic non-protein-coding (LINC) RNA genes.

I recognised that this would entail removal of a some true-positive pathogenic variants, but it was necessary to prioritise variants with a higher probability of relevance.

The remaining variants were analysed in greater detail using Alamut® Visual 2.11 and the subsequent version Alamut® Visual Plus 1.5.1, a commercially available genome browser. This allows access to a number of *in silico* prediction tools for variant pathogenicity, including Align GVGD (v2007)(113), SIFT (v.6.2.0)(114), MutationTaster(115, 116) and PolyPhen-2(117), together with direct information about the conservation of the relevant base and nucleotide, the Grantham distance(118) for amino acid substitutions, and the location of the variant within a known protein domain.

For splicing variants the prediction tools MaxEnt, NNSPLICE and SSF were used: a mean effect on splice site of >50% between the three scores was considered to be potentially significant, although variants with lower scores which were close to splice sites were sometimes explored further if there were other features increasing their likelihood of being pathogenic (such as occurrence *de novo* or in *trans* a pathogenic variant).

For missense variants no further *in silico* analysis was routinely performed, but for indels, frameshifts, stop-gain (nonsense) or stop-loss variants, PROVEAN v.1.1.3(119) was also used.

In general, variants not previously reported as pathogenic were considered to be of interest if at least three predictors (for missense variants), two (for indels, frameshifts, stop-gain and stop-loss variants) or the average predicted effect on splicing suggested pathogenicity. These thresholds were adjusted depending on the relevance of the gene to the phenotype and/or the biological plausibility of its contribution to disease. For example, variants in genes with little or no expression in brain were discarded, while those with clear roles in, for example, synaptic transmission were given additional weight. These judgements were made with the assistance of OMIM(120) and Uniprot(121), the online gene and protein databases, which include access to a number of gene expression atlases.

For heterozygous variants in genes with no known autosomal dominant disease phenotype, constraint scores from gnomAD(103) were consulted, specifically the observed/expected (oe) score for either missense or loss-of-function variants as relevant. The probability of loss-of-function intolerance (pLI) score was also noted for loss-of function variants. In general, an oe of <0.35 and/or a pLI >0.9 was considered to increase the plausibility of pathogenicity for a heterozygous variant.

The outcome of this analysis was a shortlist, generally of between one and 15 variants per proband, of potential pathogenic interest.

2.7 Sanger confirmation

Initially, variants where either reporting to the referrer or further analysis was planned were confirmed by Sanger sequencing as follows. During lab closure for the coronavirus lockdown, confirmation was necessarily suspended and variants were reported without confirmation, with a note to that effect. Subsequently, confirmation was not restarted for reported variants of uncertain significance (American College of Medical Genetics and Genomics (ACMG) Class 3) or pathogenic/likely pathogenic variants where the read depth, allele fraction and call quality were all high and the variant was also seen to be present in the BAM file.

2.7.1 Primer design and PCR

Forward and reverse primers were designed using the Primer3Plus(122) software with the following requirements:

- Length 18-25bp
- Melting temperature (Tm) 57-63C, with a maximum difference of 3C
- GC% 40-60, with 50% as optimum
- Maximum poly-X 3 (less if possible)
- Maximum 3' self-complementarity 3 (less if possible)

All other settings were left at the programme's default. Proposed primers were checked using Alamut® Visual 2.11/Alamut® Visual Plus 1.5.1 to exclude common single nucleotide polymorphisms (SNPs) with a population MAF of >0.01%, and using the UCSC Genome Browser Human Blat Search(123) in case of non-specific binding. Primers were obtained from Sigma-Aldrich in desiccated form and diluted in nucleotide-free water to a concentration of 10pM/µl (with stocks stored at 100pM/µl).

Optimisation took place as follows. A reaction mixture was prepared with the following ingredients (multiplied by six to allow for six temperature bands for optimisation):

Table 2.4: *Reaction mixture for primer optimisation*

Reagent	**Volume (µl)**
Nucleotide-free water	2
BioMix™ Red 2x reaction mix	5
Forward primer	1
Reverse primer	1
DNA	1

The polymerase chain reaction (PCR) was then run on a Veriti™ 96-well thermal cycler at the following settings. For primer optimisation a gradient of six temperatures (usually 54-64°C at 2°C intervals) was used; for subsequent PCRs, Tm (which determines annealing temperature) was the determined optimum temperature.

Table 2.5: *Thermal cycler settings for touchdown PCR*

Number of Cycles	Temperature (°C)	Time (s)
1	95	240
2	95	30
	Tm + 4	30
	72	30*
2	95	30
	Tm + 2	30
	72	30*
35	95	30
	Tm	30
	72	30*
1	72	300

* For primers >1000bp in length, elongation times were increased by 30s per Kb.

A 100ml 1.5% agarose gel stained with 1.5µl of SYBR Safe was prepared and 3.5µl of the product at each temperature was run for 30 minutes at 100V, in an electrophoresis tank filled with 1X TBE buffer. The GeneRuler™ 100bp DNA Ladder from Thermo Fisher Scientific (or the 1kb version for larger expected products) was used as a reference for product size. The gel was imaged using Image Lab™ software a ChemiDoc™ imager.

Where clear single amplicon bands were not seen at any of the standard temperatures, hotter or cooler temperatures, or addition of GC Rich polymerase or dimethyl sulfoxide (DMSO) was tried, or as a last resort primers were redesigned.

A PCR reaction mixture was then prepared as follows (this is identical to the optimisation mixture above, except that 20µl reaction volumes were used instead of 10µl, and the DNA added was that intended for analysis or comparison, rather than a control sample):

Table 2.6: *Reaction mixture for PCR*

Reagent	Volume (µl)
Nucleotide-free water	4
BioMix™ Red	10
Forward primer	2
Reverse primer	2
DNA	2

PCR was performed with the settings above, where Tm was the primer's optimum annealing temperature. Annealing time was increased as above for anticipated product sizes >1kb. 3.5µl of the product was then run on a gel as above to confirm amplification of the expected product.

2.7.2 Purification

The PCR products were purified by combining 3.5µl of each PCR product with 3.5µl of MicroCLEAN (from Clent Life Science) and allowing the mixture to stand at room temperature for five minutes. The sample was then centrifuged at 4000rpm at 4°C for 40 minutes. Excess liquid was disposed of by centrifuging the sample upside-down on tissue paper for 1 minute at 500rpm, after which the PCR products were re-suspended in 4.5µl of nuclease-free water.

2.7.3 Sequencing

A 10µl sequencing reaction mixture was then prepared for each individual primer as follows, using the BigDye™ Terminator v1.1 Cycle Sequencing Kit from Thermo Fisher Scientific:

Table 2.7: *Reaction mixture for sequencing reaction*

Reagent	Volume (µl)
BigDye Terminator Buffer	2
BigDye Terminator	0.5
Primer	1
Nucleotide-free water	2
PCR product	4.5

From 2022, BigDye™ Terminator v1.3 was used instead: despite the manufacturer's cautions, this turned out not to require any adjustments in the protocol or other reagents.

The sequencing reaction was then run on the same Veriti™ 96-well thermal cycler with the following settings:

Table 2.8: *Thermal cycler settings for sequencing reaction*

Number of cycles	**Temperature (°C)**	**Time (s)**
1	96	180
35	96	30
	50	15
	60	240

The products were then purified and precipitated as follows. A 25:1 volume:volume mixture of 100% ethanol and sodium acetate was prepared and 50µl was added to the sequencing product in each well, before being incubated at room temperature for 20 minutes. The sample was then centrifuged at 3000rpm at 4°C for 40 minutes. Excess liquid was then removed by centrifuging upside-down on tissue paper for one minute at 300rpm. 50µl of 70% ethanol was then added to each well and the mixture centrifuged at 3000rpm at 4°C for 10 minutes and the supernatant was again removed as above. 10 µl 1X tris-EDTA buffer was then added and the sample stored at -20°C until it could be taken for sequencing.

Sequencing was performed by the GOSH diagnostic genetics laboratory using an ABI 3730xl DNA Analyzer. Sequencing files were returned electronically and analysed using Mutation Surveyor® (manufactured by SoftGenetics®) and/or the FinchTV chromatogram viewer from Geospiza Inc. to confirm the presence and segregation of the variant.

2.8 Reporting

Results were reported back to the referring clinician in writing (see Appendix 2 for a report template). Reports included a request that the clinical team seek confirmation of the variant in an accredited diagnostic laboratory. For some patients known to the NMDC this confirmation was sought through the GOSH diagnostic laboratory and results were fed back directly to the family, but a report was also copied to their referring clinician and Clinical Genetics consultation was generally advised. This was because even when it had been possible to explain the implications of the results in an unambiguous manner we felt most families would benefit from a further opportunity to discuss (for example) the implications for other relatives, and to re-review potentially unfamiliar and challenging new information. It also provides a form of quality control in that an independent Clinical Genetics team would review our results and interpretation.

Use of very large gene panels risks producing a large number of VUSs and we were keen to keep this to a minimum in order to facilitate interpretation of the reports. It would clearly be inappropriate to report variants in genes without a confirmed disease phenotype to non-research clinicians or to families, and therefore even where these genes formed part of a panel used, we did not include them in the report. We also did not report variants in genes (even in the panel used) where the phenotype was felt to be a very poor fit for patient's condition, as far as this was known to us. To avoid confusion, a full list of genes in the megapanel or any other panels applied was *not* routinely provided with the report.

2.9 Matchmaking

Where variants of interest were identified in genes without a known disease phenotype, additional cases were sought both in our own cohort and from other teams. The main platform used for the latter was GeneMatcher.(124) The Deciphering Developmental Disorders(125) online database was also checked. In some cases we also contacted other neurology or genetics teams directly with a request that they

interrogate their own databases of exomic or genomic data, and I often undertook the same task reciprocally for them.

The interpersonal, political and negotiatory aspects of managing the matchmaking process would be an interesting subject for quite a different thesis, possibly in ethnography, but I do not propose to explore them here.

2.10 Extraction of RNA, reverse transcription and splicing experiments

Living patient-derived fibroblast cultures were provided to our laboratory by Dr Derek Burke's team at GOSH. RNA was extracted from pelleted fibroblasts using the Qiagen RNeasy® Mini Kit as follows. 10µl β-mercaptoethanol was added to 1ml Buffer RLT and 350µl of the resulting mixture was added to the cell pellets. Cells were then dissociated and lysed by vortexing for one minute. 350µl 70% ethanol was added and mixed and the 700µl of sample was placed in an RNeasy® spin column over a 2ml collection tube and centrifuged for 15s at 8000g. The flow-through was discarded.

The column contents was washed by adding 700µl of Buffer RW1 and centrifuging again for 15s at 8000g, again discarding the flow-through. The process was repeated with 500µl of Buffer RPE, with time centrifuged for 2 minutes, followed by an additional 1 minute over a fresh collection tube to ensure full elimination.

The spin column was transferred to another fresh 1.5ml collection tube and 30µl of RNAse-free water was added. It was centrifuged for 1 minute at 8000g. This time the flow-through, containing the eluted RNA, was retained.

After measurement using the Nanodrop, the Invitrogen DNase 1 kit was used to digest any contaminating DNA. 1µg of RNA was added to 1µl of 10X DNase 1 Reaction Buffer, 1µl of DNase 1 (diluted to 2.5 Kunitz units/µl) and RNase-free water to a total of 10µl. The mixture was incubated for 15 minutes at room temperature and 1µl of EDTA solution was then added to inactivate the DNase I. It was then heated to 65°C

for 10 minutes using a heat-block. The RNA, due to its instability, was stored at -80°C until needed.

Reverse transcription was performed using reagents from the Invitrogen SuperScript^{TM} III Reverse Transcriptase kit with Oligo(dT) Ambion^{R} primers (Invitrogen) and dNTP mix prepared from dNTPs from BioLabs (supplied as individual dNTPs at 100mM; combined and diluted at 1:10 in nucleotide-free water). Oligo(dT) primers are oligonucleotides which include a string of deoxythymidines (dTs) capable of binding to the poly-A tail which distinguishes messenger RNA (mRNA), and thus selecting it for conversion to complementary DNA (cDNA).

A mix was then prepared containing 1µl oligo(dT) primer solution (50µM), 1µl dNTP Mix, 1000ng RNA (volume varying depending on measured concentration) and sufficient nuclease-free water to achieve a total volume of (at least) 13µl. This was then incubated at 65°C for 5 minutes, followed by 1 minute at 4°C on ice.

During this stage a second mixture was prepared containing (per sample) 4µl 5X First Strand Buffer, 1µl 0.1M dithiothreitol (DTT), 1µl SuperScript^{TM} III Reverse Transcriptase (200U/µl) and 1µl nuclease-free water. A total of 7µl was added to each sample before incubation at 50°C for one hour, followed by 70°C for a further 15 minutes. The resulting cDNA was diluted 1:25 with nucleotide-free water (i.e. to a total volume of 500µl) and stored at -20C.

To assess the impact of variants on splicing, primers were designed using the cDNA sequence such that the forward and reverse primers lay on either side of the exon of interest, and were diluted and optimised as previously described. For example, in the case of DY173, described in Chapter 5, exons were designed to lie on either side of exon 7 of *VPS41*, as shown in **Table 2.9.**

PCR was then carried out as above and the product run on a 1.5% agarose gel to allow comparison of product sizes, using a 100kb DNA ladder. To improve resolution of different product sizes, a 2-3% agarose gel was used.

Table 2.9: *Primer characteristics used for splicing experiment for DY173*

Primer	Exon	Sequence	Conditions
Forward	1	AGCAGAGGAGCAGGAAACTG	Touchdown PCR; Tm 56°C
Reverse	8	TGAACGGTCTTGGATGAACA	

Bands were excised from the gel using a scalpel and an ultraviolet transilluminator and purified using the QiaQuick® Gel Extraction Kit in accordance with the manufacturer's instructions, as follows. Gel volume was estimated at 100mg:100µl and 6 gel volumes of buffer QG were added. The sample was incubated for 10 minutes at 56°C, vortexing every two to three minutes until the gel was fully dissolved. One gel volume of isopropanol 100% was added and the sample was centrifuged in a QiaQuick® spin column at 13,000rpm for one minute. The flow-through was discarded. 500µl QG buffer was added to the column and column was centrifuged for a further minute, again discarding the flowthrough. The sample was then washed with 750µl buffer PE and centrifuged again. 3µl EB buffer was then added directly to the membrane, it was allowed to stand for up to four minutes to maximise yield, and it was centrifuged for a further minute over a clean 1.5ml microcentrifuge tube to elute the DNA.

Purified DNA yields were (as expected) low so a further PCR was undertaken using the same primers and conditions in order to amplify the sequence of interest before preparation for sequencing as above.

2.11 Quantitative real-team PCR

Quantitative real-team PCR (qRT PCR) was used to quantify gene expression by measuring amounts of mRNA to assess whether variants had an impact on transcription. Following extraction of RNA and synthesis and dilution of cDNA as above, cDNA for each sample was diluted 1:2 with nuclease-free water to provide 9µl for each of three technical replicates for each target sequence (i.e. each sequence of interest for quantification, plus one housekeeping gene for use as a reference: I used *GAPDH*).

For each target, 10µl per sample of Mesa Blue qPCR MasterMix (from Eurogentec) was mixed with 1µl of primer mix (containing the forward and reverse cDNA primers for the variant in question). 11µl of the resulting mixture was added to each well of a qRT PCR plate and 9µl of diluted cDNA was added before sealing with optical adhesive film. The plate was run in a StepOne Real-Time PCR machine set for quantification as shown in **Table 2.10.**

Table 2.10: *qRT PCR thermal cycler settings*

Number of cycles	Temperature (°C)	Time (s)
1	95	600
40	95	15
	60	60
1	95	15
	60	60
	95	15

After qRT PCR, relative expression of the target sequence in each sample was analysed by calculating fold change in expression (FC) using the $\Delta\Delta C_T$ method, where C_T is cycle threshold (number of PCR cycles required to reach threshold of detection):

$$\Delta C_T = (\text{Mean } C_T \text{ for target sequence}) - (\text{Mean } C_T \text{ for } \textit{GAPDH})$$

$$\Delta\Delta C_T = \Delta C_T \text{ in sample} - \Delta C_T \text{ in control}$$

$$FC = 2^{-\Delta\Delta C_T}$$

2.12 Immunofluorescence and confocal microscopy

HEK-293T cells for this technique were cultured on coverslips by Dr Kimberley Reid as follows: sterilised 13mm glass coverslips were coated in poly-D-lysine and placed in a 24-well culture plate. 2-3 X 10^5 cells were added to each well and allowed to settle overnight. They were then transfected as appropriate and cultured at 37°C for 24 hours. I prepared the slides from this point onwards.

Cells were rinsed with phosphate buffered saline (PBS) followed by fixation with 4% paraformaldehyde (500µl approx.) at room temperature for 10 minutes before rinsing. Coverslips were then rinsed in PBS and permeabilised with 0.1% Triton™X-100 in PBS for 10 minutes at room temperature. Cells were washed three times before blocking with blocking buffer (5% foetal bovine serum (FBS) in PBS) at room temperature one hour, with gentle agitation.

Following blocking, cells were incubated with the primary antibody (mouse anti-Myc IgG diluted 1:8000 in blocking buffer) at 4°C overnight. They were then rinsed three times with PBS and the secondary antibody (goat-anti-mouse IgG conjugated with Alexa Fluor™ 594) was added and incubated for one hour at room temperature in the dark and then rinsed again. The cells were then incubated in the dark for 5 minutes with 4',6-diaminido-2-phenylindole (DAPI, a nuclear stain) and wheat germ agglutinin (WGA, a cell membrane stain), both diluted 1:1000 in PBS. After three final rinses, the coverslips were attached to SuperFrost™ microscopy slides using ProLong™ Gold antifade mountant (approx. 10µl per coverslip). They were left to set overnight in the dark at room temperature and then refrigerated until use.

Microscopy was performed using a Zeiss LSM710 confocal microscope at a magnification of X63 with oil.

2.13 Methods relevant to this thesis undertaken by others

For the collaborative projects described in this thesis, there are some techniques which form an essential part of the research but which I did not undertake myself. I will outline these briefly in this section for completeness, indicating in each case who performed them.

2.13.1 Weighted burden analysis

This statistical technique was employed by Dr Michael Zech and his team based at the Helmholtz Zentrum Institute of Experimental Genetics in Munich, Germany, to identify enrichment of variants in specific genes in their cohort of 138 people with undiagnosed suspected genetic dystonia compared with a control population taken

from gnomAD. They used a method called TRAPD (Test Rare vAriants with Public Data)(103) and full details are given in our publication.(77)

2.13.2 Cell culture, site-directed mutagenesis and transfection

Fibroblast culture was undertaken by Dr Derek Burke and his team at the GOSH Enzyme Laboratory, and also by Dr Kimberley Reid and Dimitri Budinger within our group. Fibroblasts were cultured at 37°C in either Ham's F10 medium with 12% FBS (in Dr Burke's lab) or in Dulbecco's Modified Eagle's Medium (DMEM) with 10% FBS, in our lab. Following culture, cells were disaggregated using 0.2% trypsin and processed for either electron microscopy (see below) or mRNA testing (see above). For the experiments on *DRD1*, HEK-293T cells were cultured by Dr Reid in DMEM with 10% FBS with penicillin/streptomycin and L-glutamine supplementation, at 5% CO_2.

For transfection, pCMV6 plasmids containing a FLAG- and Myc-tagged human *DRD1* were purchased from Origene. Site-directed mutagenesis (SDM) was performed using the QuikChange Lightning SDM kit from Stratagene to introduce the desired variant. This involves using a thermal cycling process with targeted primers containing the variant to synthesise mutant DNA strands. A proprietary restriction enzyme was used to digest the remaining nonmutated DNA. XL10-Gold Ultracompetent cells (an *Escherichia coli* strain modified for competency) were transformed with the mutant plasmids by means of a brief heat pulse at 42°C. They were then cultured on agar containing kanamycin, to which the pCMV6 plasmid confers resistance, to selectively expand the clones which contained the plasmid. The plasmids were then extracted and purified using QIAGEN Maxiprep kit in accordance with the manufacturer's instructions, and plasmids Sanger sequenced to confirm the presence of the variant in question. Lipofectamine 2000 at a ratio of 2:1 with DNA was used to transfect the HEK-293T cells with the plasmids. Cells were then cultured for a further 24-48 hours before analysis.

2.13.3. Electron microscopy

Microscopy preparation and microscopy was performed by Mr Glenn Anderson and his team in the GOSH Histopathology unit. Cell clusters intended for microscopy were

fixed in 2.5% glutaraldehyde in 0.1M cacodylate buffer followed by secondary fixation in osmium tetroxide. Tissues were dehydrated in graded ethanol, transferred to an intermediate reagent, propylene oxide, and then infiltrated and embedded in Agar 100 epoxy resin. Polymerisation was undertaken at 60°C for 48 hours. 90nm ultrathin sections were cut using a Diatome diamond knife on a Leica Ultracut UCT microtome. Sections were transferred to copper grids and stained with alcoholic uranyl acetate and Reynolds lead citrate. The fibroblasts were examined using a JEOL 1400 transmission electron microscope.

2.13.4 Western blotting and biotinylation

This was undertaken by Dr Reid. HEK-293T cells transiently expressing pCMV6 empty vector, *DRD1*-wild type or *DRD1*-T37K were lysed with a protease inhibitor-containing buffer and centrifuged to remove cell debris. The lysate was denatured and resolved on a polyacrylamide gel and transferred to an inert membrane. It was probed using mouse anti-Myc as a primary antibody and goat anti-mouse secondary antibody conjugated with horseradish peroxidase, and a luminol-based chemiluminescence substrate solution was used for imaging using a ChemiDoc system. An antibody against β-actin was used as a control for normalisation of relative expression.

For biotinylation studies, transfected cells were incubated with a biotinylation reagent for 30 minutes on ice before the reaction was terminated with a quenching buffer. Cells were then lysed and centrifuged as above. Biotinylated proteins were separated using NeutrAvidin beads (containing avidin, which specifically binds biotin). The total lysates and biotinylated proteins were resolved on a polyacrylamide gel and probed with primary and secondary antibodies and imaged as above. B-actin antibodies were again used for a control for the total lysate and Na^+/K^+ alpha ATPase for normalization of cell surface expression.

2.13.5 cAMP assay

This test was also performed by Dr Reid. Since the dopaminergic D1 receptor is coupled with a $G_{\alpha s}$ protein, agonist binding raises intracellular cyclic adenosine monophosphate (cAMP) levels. HEK-293T cells transiently expressing pCMV6 empty vector, *DRD1*-wild type or *DRD1*-T37K variant were incubated with Chloro APB, a D_1

agonist, in the presence of phosphodiesterase inhibitors to increase cAMP levels further. Cells were then lysed with a compound containing protein kinase A (PKA): PKA is activated in proportion to the amount of cAMP present and breaks down ATP. A reagent called Kinase-Glo™ was added: this contains luciferin and reacts with the remaining ATP in a light-generating luciferase reaction. The plate was then read using a luminometer. High levels of luminescence correspond with high levels of remaining ATP and therefore *lower* levels of intracellular cAMP. Results have been inverted for clarity in the graphs shown.

2.13.6 Tritiated dopamine ligand binding assay

HEK-293T cells transiently expressing *DRD1*-WT or *DRD1*-T37K variant were lysed with an ice-cold lysis buffer and centrifuged to pellet the membranes. The pellets were re-suspended as cell-free membrane preparations and mixed with increasing concentrations of tritiated dopamine and a binding buffer in a 96-well plate. After incubation for 10 minutes the reaction was terminated by filtration and washes through a Filtermat using a TomTec cell harvester, so that the cell membrane was retained and non-bound reagents removed. Radioactivity of the Filtermat was measured using a solid scintillation counter and normalised to membrane preparations from untransfected cells.

2.14 Table of reagents

Table 2.11: *Reagents used in the techniques described*

Reagent	Supplier	Catalogue number
2-Mercaptoethanol	Gibco	31350
4-12% Bis-Tris NuPAGE gels	Thermo Fisher	NP0326BOX
Agarose	Sigma-Aldrich	A4718
Anti-B-actin	Sigma-Aldrich	SAB3500350
BigDye Terminator v1.1	Thermo Fisher	4337452
BigDye Terminator v1.3	Thermo Fisher	4337458
BioMix Red	Meridian Bioscience	BIO-25006
cAMP-Glo™ Max Assay	Promega	V1681
Chloro-APB	Sigma-Aldrich	C130
DAPI	Thermo Fisher	62248
Deoxynucleotide (dNTP) solution set	New England BioLabs	N0446S
DMEM	Gibco	41965039

DMSO	Sigma-Aldrich	D8418
DNA primers	Sigma-Aldrich	N/A
DNase 1	Thermo Fisher	18068015
Ethanol	Sigma-Aldrich	32221
Foetal bovine serum	Gibco	11550356
GC Rich polymerase	Roche	12140306001
GeneRuler 100bp Plus DNA Ladder	Thermo Fisher	SM0321
Goat anti-mouse antibody with HRP	Cell Signalling Technology	#91196
Goat anti-mouse with Alexa Fluor 594	Invitrogen	AB_2534073
HEK-293T cells	ATCC	CRL-3216
IBMX (phosphodiesterase inhibitor)	Sigma-Aldrich	I5879
Invitrogen SuperScript™ III Reverse Transcriptase kit	Thermo Fisher	18080093
Kanamycin	Sigma-Aldrich	K1876
LB Broth with agar (Miller)	Sigma-Aldrich	L3147
L-glutamine	Sigma-Aldrich	G7513
Lipofectamine 2000	Invitrogen	11668030
Mesa Blue qPCR MasterMix	Eurogentec	RT-SY2X-03+WOUB
MicroCLEAN	Clent Life Science	2MCL-5
Mouse anti-Myc IgG	Cell Signalling Technology	2276
Na^+/K^+ alpha ATPase antibodies	Abcam	ab76020
NeutrAvidin™ Agarose	Pierce Biotechnology	29200
Oligo(dT) Ambion® Primers	ThermoFisher	AM5730G
Paraformaldehyde	Sigma-Aldrich	158127
pCMV6 plasmids + DRD1 + FLAG and Myc tags	Origene	RC210389
Penicillin-streptomycin	Gibco	15140-122
Phosphate buffered saline	Gibco	11593377
Pierce™ Premium Grade Sulfo NHS-SS-Biotin	Thermo Fisher	PG82077
Plasmid Maxi Kit	QIAGEN	12162
Poly-D lysine	Sigma-Aldrich	P6407
Protease inhibitor cocktail	Thermo Fisher	78430
Qiagen RNeasy® Mini Kit	Qiagen	74104
QiaQuick® Gel Extraction Kit	Qiagen	28706X4
QuikChange Lightning SDM kit	Agilent	210518
ReliaPrep™ Blood gDNA Miniprep System	Promega	A5081
RIPA buffer	Thermo Fisher	89900
Ro20-1724 (phosphodiesterase inhibitor)	Sigma-Aldrich	557502
Sodium acetate	Thermo Fisher	AM9740
SuperSignal West Dura chemiluminescence substrate solution	Thermo Fisher	37071
SYBR™ safe DNA gel stain	Thermo Fisher	S33102
TBE	Sigma-Aldrich	T4415
TE buffer	Applichem	A0386

Tritiated dopamine	PerkinElmer	NET1160001MC
Triton™ X100	Sigma-Aldrich	11332481001
WGA Alexa Fluor™ 488	Invitrogen	W11261
XL10-Gold Ultracompetent cells	Agilent	200315

Chapter 3. Overview of results from the cohort

3.1 Introduction

The core of my project involved analysis of genomic data for a large number of participants with rare movement disorders. In research terms, the primary aim was to identify novel gene-disease relationships and/or to improve understanding of incompletely described ones. An important secondary aim was to achieve diagnosis for as many participants as possible. In one sense, making a rapid and unambiguous diagnosis was a 'failure' for my research aims, but it was rewarding for me and, more importantly, useful for the families concerned.

In this chapter I will discuss the rate of diagnosis in the group I have analysed, the factors associated with likelihood of making a genetic diagnosis, and some of the dilemmas of variant interpretation and reporting. I will also highlight three instances where I identified variants in genes not previously implicated in disease (*TNPO2, H2AC6* and *INTS11*) and was able to contribute to the index descriptions of these conditions.

3.2 Genomes and diagnoses

At the time of writing, I have completed analysis for 168 individuals or families. I use "completed" to mean that I have undertaken the stages of analysis described in the Methods chapter: we continue to analyse genomes of undiagnosed individuals through joint analyses with other similarly affected participants, or for variants in new genes of interest.

Further clinical assessment of two individuals I analysed has since led to the conclusion that they are likely to have a non-genetic diagnosis: functional neurological disorder in one instance, and cerebral palsy due to perinatal HIE in another. Of the remaining 166, analysis has resulted in probable diagnosis in 55/166 (33.1%), including diagnoses in genes which have not yet been published (excluding these, the

numbers would be 52/166, or 31.3%). They have led to a suspected diagnosis in a further 21/166 (12.7%), i.e. the total proportion to receive a finding believed likely to be clinically relevant was 76/166 (45.8%). **(Table 3.1; Figure 3.1.)** I will discuss my use of the terms “probable” and “suspected” below. All the diagnoses, as expected given the participants involved, were of rare diseases, which are often defined as those with a prevalence less than 1 in 2000.(17) Most could reasonably be described as ultra-rare, although definitions of this term remain highly variable, ranging from 1 in 50,000(126) to “worldwide prevalence of <30 patients” (equivalent to about 1 in 270 million).(127)

It will be observed that for two participants (DY88 and DY184) given suspected diagnoses, two different genes are listed: in these cases both variants were felt to be plausible candidates for pathogenicity and, unfortunately, neither further phenotypic detail nor familial testing could be obtained.

9 patients with a probable diagnosis had variants in genes newly associated with disease which were identified in part as a result of my project: namely, *VPS16* (n=4), *VPS41* (n=1), *DRD1* (n=1), *TNPO2* (n=1), *INTS11* (n=1) and *H2AC6* (n=1). Of these, the manuscripts for *INTS11, H2AC6* and *DRD1* are in preparation at the time of writing. These findings will be discussed later in this chapter, and in Chapters 5 (*VPS16* and *VPS41*) and 6 (*DRD1*).

Several pathogenic and likely pathogenic variants identified in known disease genes were in patients whose presentations were atypical and/or not part of the previously described phenotypic spectrum of the condition in question, but in whom I nevertheless felt confident in attributing pathogenicity. I will discuss these cases further in Chapter 4.

Table 3.1: *Diagnoses and suspected diagnoses resulting from my analyses. Blue: diagnosis of a newly described disorder identified (in part) as a result of the work. Green: definite or highly probable diagnosis. Yellow: suspected but unconfirmed diagnosis.*

Identifier	Analysis	Gene	Inheritance	Phenotype
DP35	Trio	*DRD1*	AR	Infantile parkinsonism-dystonia
DY179	1 parent	*H2AC6*	AD	Cerebellar atrophy and dystonia
AX10	Sibs X2	*INTS11*	AR	ND; progressive ataxia; retinal dystrophy
PE26	Trio	*TNPO2*	Mosaic	Severe ND; spasm-like episodes
DY94	1 parent	*VPS16*	AD	Progressive dystonia
DY29	Sing	*VPS16*	AD	‘’
DY212	Sing	*VPS16*	AD	‘’
DY225	Sing	*VPS16*	AD	‘’
DY173	Trio	*VPS41*	AD	Dystonia, ataxia and regression
CD56	Trio	*ADCY5*	AD	Dyskinesia
DY200	Trio	*ALS2*	AR	Infantile onset ascending hereditary spastic paralysis
SP7	Sibs X2	*ALS2*	AR	‘’
DY191	Trio +	*AP4S1*	AR	Dystonia
DY175	Trio	*ARID1B*	AD	ID, dysmorphism, unusual movements
DY127	Trio	*ASH1L*	AD	Progressive dystonia and ID
SP10	Trio	*ATM*	AR	Ataxia-telangiectasia
DP33	Trio	*ATP1A3*	AR	Severe neonatal-onset dystonia
SP21	Trio	*CCDC82*	AR	ND; spasticity
ZZ19	Sibs X2	*CNTNAP1*	AR	Lethal congenital contracture syndrome
DY189	Trio +	*DDC*	AR	AADC deficiency
DY181	Trio	*DEGS1*	AR	Hypomyelinating leukodystrophy
CD57	Trio	*DRD2*	AD	Choreiform disorder
PE34	1 parent	*EBF3*	AD	Hypotonia, ataxia and delayed development syndrome
NB36	Trio	*EPG5*	AR	Neurodegeneration; absent corpus callosum; NBIA-like features
DY120	Trio	*GNAO1*	AD	ND with involuntary movements
DY91	Sing	*GNAO1*	AD	‘’
DY72	Sing	*GNAO1*	AD	‘’
MY15	Trio	*GNB1*	AD	ID and myoclonus
DY47	Sing	*H3-3A*	AD	Progressive spastic paraplegia; mild ID
DY197	Trio	*H3-3B*	AD	Global developmental delay, epilepsy, dystonia
PE19	1 parent*	*JPH3*	AD	Paroxysmal dyskinesia and ND
PE5	Sing	*JPH3*	AR	Paroxysmal dyskinesia (AHC) and ND
DY133	1 parent	*KMT2B*	AD	Progressive dystonia
DY208	Sing	*KMT2B*	AD	‘’
DY224	Trio	*KMT2B*	AD	‘’
DY82	Sing	*KMT2B*	AD	‘’
CD67	Trio	*MRE11*	AR	Hyperkinetic disorder
DP22	Trio +	*MRM2*	AR	Dopa-responsive dystonia
NB7	Trio	*MT-ATP6*	Mito	Basal ganglia calcification and mild hyperkinetic disorder
CD16	Trio	*NKX2-1*	AD	Benign hereditary chorea
CD52	Trio	*PDE10A*	AD	Bilateral striatal necrosis
ZZ30	Trio	*PIEZO2*	AR	Distal arthrogryposis with impaired proprioception and touch

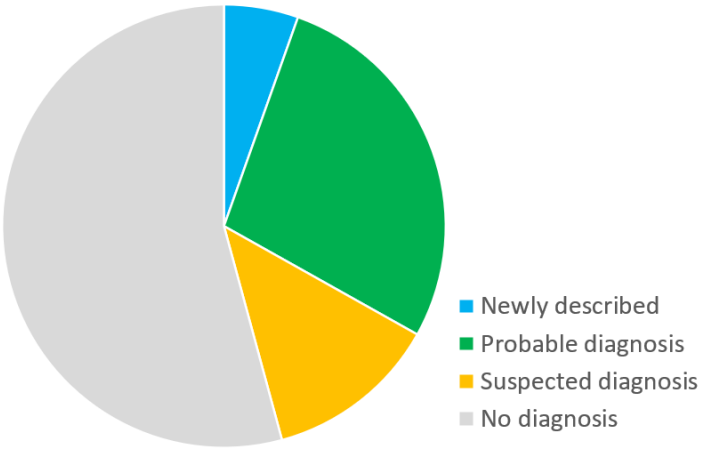
DY131	Trio	*RHOBTB2*	AD	ND, paroxysmal movement disorder
DP3	Trio	*RHOBTB2*	AD	‘’
DY206	Trio	*SEPSECS*	AR	Pontocerebellar hypoplasia
DP38	Trio	*SLC18A2*	AR	Infantile parkinsonism-dystonia
CD24	Trio	*SLC30A9*	AR	Hyperkinetic disorder; ID; oculomotor apraxia
DP27	Sing	*SLC6A3*	AR	Dopamine transporter deficiency syndrome
DY220	Trio	*SPAST*	AD	Progressive spasticity and weakness
TR5	Sing	*STXBP1*	AD	ND and tremor
CD49	Trio	*TAF1*	XLR	ID; dystonia
DY192	Sing	*THAP1*	AD	Progressive dystonia
SP19	Sing	*TUBA1A*	AD	ND, progressive spasticity, cerebellar atrophy
PE13	Trio	*UBE3A*	AD	Angelman syndrome (AHC-like presentation)
CD39	Trio	*YY1*	AD	Progressive hyperkinetic disorder and mild ID
DY88	Sing	*AFG3L2/THOC2*	AD/XLR	Dystonia, ND
SP15	Sing	*ATP1A2*	AD	ND, spasticity; early death
AX18	Sing	*CAMTA1*	AD	Mild ND and cerebellar dysfunction
CD66	Trio	*CASK*	XLR	Chorea, myoclonus, ND
DY204	Sing	*CCDC88C*	AD	Familial paroxysmal dystonia
DY136	Sing	*COL4A2*	AD	CP-like
MY16	Trio	*EEF1A2*	AD	Myoclonus-dystonia
MY9	1 parent	*GNB2*	AD	ND, myoclonus
CD62	Trio	*IRF2BPL*	AD	Dyskinesia, abnormal MRI, severe ND
DY205	Sing	*KDM3B*	AD	Dystonia
PE18	Trio	*KCNMA1*	AD	Paroxysmal dyskinesia
DY93	Sing	*LRPPRC*	AR	Progressive dystonia; hearing and visual loss
DP26	Sing	*MAGEL2*	AD	Severe ND, hypokinetic
ZZ28	Trio	*MECP2*	XLR/D	Autism, unsteady gait
SP16	Sing	*MORC2*	AD	ND, spasticity, dystonia
DY184	Sing	*NAA15/MAG*	AD/AR	Progressive dystonia
PE20	Trio	*PDE10A*	AD	Paroxysmal dystonia
AX22	Trio	*PNPT1*	AD	Ataxia, neuropathy, hearing loss
DY89	Sing	*PRKAR1B*	AD	Adult-onset dystonia and tremor
MY6	Trio	*PTCHD1*	XLR	Unsteadiness, ?autism
DY92	Sing	*SLC40A1*	AD	Adult-onset dystonia-parkinsonism

Abbreviations: Sing: singleton; AD: autosomal dominant; AR: autosomal recessive; ID: intellectual disability; NBIA: neurodegeneration with brain iron accumulation; ND: neurodevelopmental disorder: Trio: triome; Trio+: triome and additional family member(s); SCAR: autosomal recessive spinocerebellar ataxia; Sibs X 2: two siblings analysed. SPG: spastic paraplegia; XLR: X-linked recessive 1 parent: proband and one parent analysed (indicates that parent was also affected).*

The majority of diagnoses were made in one individual or family only. The exceptions to this were conditions associated with *VPS16* (n=4); *KMT2B* (n=4), *GNAO1* (n=3), *RHOBTB2* (n=2), *ALS2* (n=2) and *JPH3* (n=2, though with different modes of inheritance – see Chapter 4). For *PDE10A* and *IRF2BPL* one confident diagnosis and

one suspected diagnosis was made for each. These results cannot, in my view, be taken as any kind of “genetic landscape” or epidemiological estimate, because of the biased way in which the cohort was recruited, largely from patients who had not been able reach a diagnosis through conventional testing.

Figure 3.1: *Proportion of individuals receiving probable or suspected diagnoses. Newly described disorder 5.4%; other probable diagnostic finding 27.7%; suspected finding 12.7%; no diagnosis 54%.*



3.3 Assessment of pathogenicity and levels of certainty

Determination of pathogenicity and reporting variants of uncertain significance is a vexed area. The practice of diagnostic laboratories in the UK is based on the ACGS guidelines,(93) which are in turn based on the criteria published by the ACMG.(94) Both bodies recognise the limitations of the system and are continually working on refinements: as in so many areas, there are difficult trade-offs both between accuracy and simplicity, and between sensitivity and specificity. The current criteria for the assessment of pathogenicity (in the context of full-penetrant rare monogenic disease) are summarised in **Table 3.2**. A separate set of criteria are published for the assignment of a benign classification, but these are less relevant here.

Table 3.2 *Summary of ACMG criteria for assessment of variant pathogenicity, adapted from Richards et al, 2015.(94)*

Description	Criterion
Truncating variant, including start-loss and canonical splicing variants	PVS1
Same amino acid change (different base change) as a known pathogenic variant	PS1
De novo variant with confirmed parental identity	PS2
Functional studies support damaging effect of variant	PS3
Prevalence of variant is significantly increased compared to controls (odds ratio >5)	PS4
Located in a mutational hotspot or critical non-variable functional domain	PM1
Absent from controls (or if recessive very low frequency)	PM2
In *trans* with a pathogenic variant (if recessive)	PM3
In-frame deletion or stop-loss	PM4
Missense change affecting an amino acid where a different missense is pathogenic	PM5
Assumed *de novo*, parenthood not confirmed	PM6
Segregates with disease in multiple family members	PP1
Missense (if known disease mechanism) in a gene with low rate of benign variation	PP2
Multiple *in silico* predictors suggest a deleterious effect	PP3
Phenotype specific for a disease with a single genetic aetiology	PP4
Variant reported as pathogenic but evidence not available	PP5

To reach the classification of a variant as “pathogenic” (previously “Class 5”) or “likely pathogenic (previously “class 4), criteria are combined as follows:

Pathogenic:

- PVS +
    - 1 X PS or
    - 2 X PM or PP*
- 2 X PS*
- 1 X PS +
    - 3 X PM
    - 1 X PM + 3X PM or PP

Likely pathogenic:

- 1 X PVS or PS + 1 X PM
- 1 X PS + 2 X PP
- 3 X PM
- 2 X PM + 2 X PP
- 1 X PM + 4 X PP

*In contrast to the ACMG, the ACGS recommends that PVS + 1X PM should be considered pathogenic whereas those with 2 X PS (and no other supporting criteria) should be classified likely pathogenic.

The ACGS modifications to the ACMG criteria above are based on a Bayesian analysis of the criteria outcomes by Tavtigian *et al*.(128) The ACGS also allows the supporting criterion “phenotype specific to a condition with a single genetic aetiology” to be used as a PM or PS criterion where there is strong clinical evidence for the condition in question: it recommends that it only be considered PS if there is highly-specific supporting laboratory evidence such as absence of dystrophin on muscle biopsy for Duchenne muscular dystrophy.

These guidelines, in my view, struggle to incorporate adequate weighting for clinical evidence. For example, it is possible to classify *any de novo* truncating variant, or indeed any truncating variant which is functionally demonstrated to impair protein function regardless of inheritance, as pathogenic. Although it is specified that the criteria should only be used “if the patient’s phenotype is consistent with the disease gene association”,(93) the meaning of this in practice can be hard to pin down. For example, in a patient presenting with a neurodevelopmental disorder and paroxysmal movements, is it appropriate to classify a variant in a gene associated with an epileptic encephalopathy as “pathogenic” if it meets the other criteria, or does that overstate the likelihood of the presenting symptoms forming part of the gene-disease spectrum? Or in a patient with non-syndromic intellectual disability, for which hundreds of genes are indexed on OMIM and, it is certain, hundreds more remain undescribed, can a variant really be considered *definitively* pathogenic simply because it is truncating and can be demonstrated *in vitro* to affect the protein?

Conversely, I feel the guidelines may not allow enough weight to clinical markers which do not fully meet the “single genetic aetiology” criterion (PP4). For example, I identified compound heterozygous variants in *SEPSECS* (the causative gene for pontocerebellar hypoplasia type 2D) in a child with clear evidence of pontocerebellar hypoplasia on MRI brain scan. The diagnostic laboratory was unable to apply the aetiology criterion because pontocerebellar hypoplasia has several different genetic

causes: they were of course correct, but accuracy would surely be improved by taking into account the presence of a rare clinical manifestation associated with a limited range of genes.

A separate refinement of the ACMG criteria, called Sherloc, was published by Nykamp et al in 2017 and aims to improve various aspects of their application including the incorporation of clinical information.(129) Perhaps inevitably, a good deal of the simplicity of the ACMG system is sacrificed and decisions on variant weighting involves a number of separate decision trees. Oddly, this system does not seem to be referenced at all in the ACGS guidance.

Case History: DY220

DY220P is one of fraternal twins born at 33 weeks' gestation. The neonatal period was reasonably uneventful. Her twin developed typically but DY220P's gross motor skills were delayed from infancy. She started walking with support at two-and-a-half years old and subsequently used a Kaye walker. With it, she was initially able to walk for substantial distances but this gradually deteriorated and by the age of 12 she used a wheelchair full-time. On examination she had gradually increasing bilateral spasticity, predominantly affecting the lower limbs. She also began to develop symptoms suggestive of an autonomic and sensory neuropathy in her lower limbs (though nerve conduction studies at five years old were essentially normal), and difficulties with bladder function. MRI was consistent with periventricular leukomalacia but as this was not felt to explain her progressive symptoms, at the age of six years she also had a singleton clinical exome examining a panel of 69 genes associated with hereditary spastic paraplegia and/or dystonia.

The first page of the report was identical to a negative report, stating "Diagnosis not confirmed". At the bottom of the second page, after the panel gene lists and in small font, it was mentioned that she had a heterozygous variant of uncertain significance in SPAST *(NM_104946.3, c.1637G>T, p.Gly546Val). No additional notes or recommendations were included. The report was treated as negative by the referring clinical team.*

The breadth of the "VUS" category can, in my experience, lead to difficulties and delays in clinical interpretation if it is not supplemented with further information. This is illustrated by the case of *DY220P*: this girl had a variant in *SPAST* which was

correctly identified and correctly classified as a variant of uncertain significance in 2016, but because of the way it was presented to the referring clinical team this was treated as a non-significant finding. When – unaware of this result – I undertook triome WGS for this patient, it became apparent that the variant was *de novo* and therefore could be classified as “likely pathogenic”. After clinical review, the patient was given a diagnosis of autosomal dominant hereditary spastic paraplegia type 4: had parental testing been suggested in the clinical exome report, this diagnosis could have been made six years earlier.

The classification of homozygous variants is also problematic. Although the ACGS guidelines state that criterion PM3 (detected *in trans* with a pathogenic (ACMG) or likely pathogenic (ACGS) variant) can be applied as a supporting criterion to homozygous variants, it is not clear how this could be helpful if the “other” (identical) variant were not already classified as pathogenic. This seems likely to underestimate the likely pathogenicity of suspicious homozygous variants. The Sherloc system notes, but does not appear to resolve, this issue.

I have not, in this project, considered myself strictly bound to observe the ACGS guidelines: our reports are marked as research reports undertaken in a non-accredited laboratory and we recommend that in cases where a clinically relevant finding is made it should be confirmed both by retesting in an accredited diagnostic laboratory and by discussion with a Clinical Genetics team. We decided to report all variants classified as likely pathogenic or pathogenic, and also VUSs in genes relevant to the disease phenotype which we felt had a realistic chance of being contributory – including, in contrast to current diagnostic guidelines, heterozygous pathogenic variants in recessive disorders, especially where these were a good phenotypic fit and/or potentially treatable. All reports were discussed at the group’s Laboratory Meeting before issuing and were signed off by Professor Kurian.

For the purposes of reporting on the cohort as a whole, I have considered a patient to have a “probable” diagnosis if they have a pathogenic or likely pathogenic variant which I consider is likely to explain their presentation (as above, this is *not* necessarily the case for every patient with a variant which can technically be classified as “likely

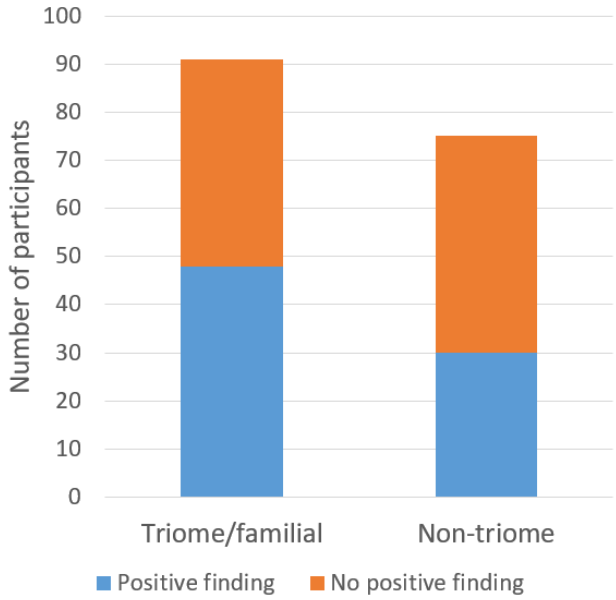
pathogenic"). I have also included a small number of participants who had a variant of uncertain significance with several criteria in favour of pathogenicity which I felt was a close fit for the phenotype described, even if it did not strictly meet the ACMG/ACGS threshold for "likely pathogenic", and especially where these have been reported in a peer-reviewed manuscript (for example, TR5, who has a novel *STXBP1* stop-loss variant for which parental inheritance could not be tested).

I have considered a patient to have a "suspected" diagnosis if they have a variant of uncertain significance with several criteria in favour of pathogenicity which I feel is a convincing but not a highly specific fit for the presenting features. Very broadly, I would estimate my "probable" diagnosis category to correspond to a probability of >90% that the identified variant is causative, and my "suspected" category to correspond to a probability of >60%, but I would not wish to claim a quantitative rigor which I cannot substantiate. A high proportion of the cases in which I did not identify a probable or suspected diagnosis had VUSs in relevant genes with some criteria in favour of pathogenicity which (after discussion with Prof Kurian) I felt had a sufficiently high chance to include in the report, but which did not meet my threshold for "suspected diagnosis".

3.4 Factors associated with diagnosis

In assessing correlation between participant characteristics and whether a diagnosis was found, for simplicity I have considered a probable or suspected diagnosis as a "positive" and any other result, with or without relevant VUSs, as "negative". The probability of a positive result will clearly be associated with two different classes of variable: firstly, those related to the adequacy of data available for interpretation (e.g. access to a complete triome; quality of phenotypic information), and secondly those related to likelihood that a monogenic disease is actually present (e.g. absence of sentinel perinatal events; presence of family history and/or consanguinity).

Figure 3.2: *Positive findings are more likely in participants analysed on a triome (or familial) basis (48//91) than in those analysed as singletons or with just one parent (30/75).*



In the first category, it seems that a positive result is associated with availability of samples for a full triome analysis (or analysis of two or more affected family members): I found a positive result in 53% of triome/familial analyses, but only 40% of singleton/uniparental analyses **(Figure 3.2)**, although this result is not statistically significant (X^2 2.68, p=0.10)**.** This difference was not explained by the greater likelihood of a genetic aetiology existing in a family with two or more affected members, as these formed only a small proportion of the analyses and the correlation persisted if they were removed. Triome analyses have two great advantages: firstly, they allow greater confidence in variant interpretation (showing, for heterozygous variants, whether they are *de novo*, and for compound heterozygous ones whether they are in *trans*); secondly, they make it practical to undertake a detailed gene-agnostic review of heterozygous variants by allowing filtering of inherited changes. Besides the difficulty of confirming a positive finding in a singleton/uniparental analysis, it is equally difficult to exclude potentially relevant VUSs, meaning that reports for these patients often include a longer list of ambiguous findings.

I am confident that participants with more detailed clinical/phenotypic information available were more likely to be diagnosed, but I cannot formally substantiate this as I do not have a rigorous means of quantifying the amount of information I had access

to. Throughout this project I found it more fruitful not to focus on forms of phenotypic documentation which are designed to be unambiguous and readily comparable (such as Human Phenotype Ontology (HPO) terms),(130) in favour of the richer clinical detail and gestalt offered by free-text summaries or, better still, direct knowledge of the patient. An example of where this served me well include CD67P: this patient was compound heterozygous for novel variants in *MRE11*, the gene associated with ataxia-telangiectasia-like disorder 1. In assessing whether the variants were likely to explain his phenotype, I re-examined his clinic letter and noticed that it was recorded *twice* that he had no telangiectasia. This supported my impression that, although it was not explicitly stated, the examining clinician felt CD67P's disorder had features resembling ataxia-telangiectasia. Another example is DY131P (discussed in section 4.2) whose symptoms included episodes of unilateral facial droop. These subtle episodes would probably not have been coded as "hemiplegia" using HPO terms, but (in the context of her other clinical and genetic findings) appear to be analogous with the hemiplegic episodes experienced by many patients with *RHOBTB2*-related disorder, which turned out to be her diagnosis.

I have not attempted to analyse the association between my broad phenotypic categories (AX, CD, DP etc) and the likelihood of a positive finding. Apart from the relatively small size of some groups, and the fact that, although I attempted to classify by main type of movement disorder many ,participants actually had two or more, the results would be confounded by the fact that a proportion of "DY" probands, referred by a specific international collaborator, came as singletons and with relatively limited phenotypic detail, reducing the rates of success in this group compared with participants from other referral sources.

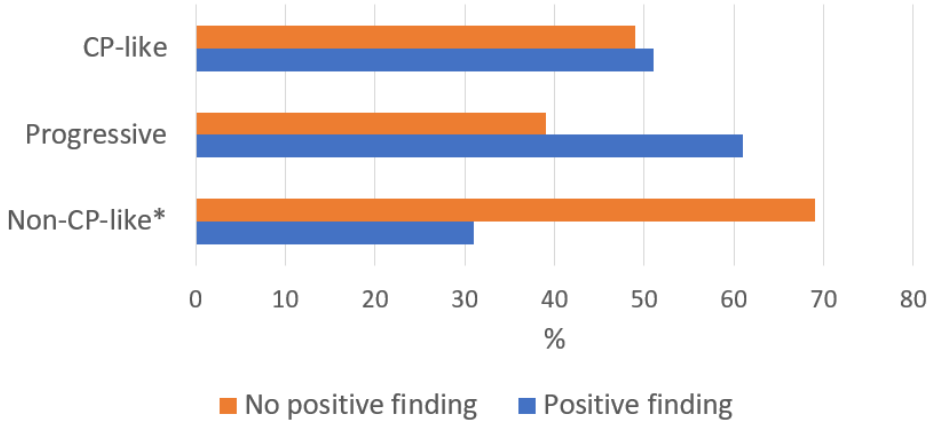
I examined whether participants who might reasonably be described as having CP were more or less likely to receive a diagnosis than others. For this purpose I classified my cohort into three groups:

1. "CP-like": those with a clear motor disorder of any kind (except purely paroxysmal) manifesting before three years of age, and no clear progression of symptoms.

2. Progressive CP mimic: as above but with clear progression over time.
3. Non-CP: motor disorder not apparent until after three years of age, and/or motor disorder is purely paroxysmal.

It should be noted, however, that the "CP-like" group is in no way a typical CP population, as many had features such as oculogyric crises, atypical or normal MRI findings, dysmorphic features or a positive family history (though none of these features would in fact disqualify their condition from classification as CP – see section 7.3). As shown in **Figure 3.3,** positive findings were most common in the progressive group, in 14/23, followed by the CP-like group (43/85), and least common in the non-CP group (19/58). Thus for participants with onset of a (non-paroxysmal) motor disorder before three years of age, there was a 53% chance of a positive finding, versus a 33% chance in the later-onset group (X^2 6.09; p=0.014; for a comparison of only CP-like participants against the non-CP group, X^2 4.46; p=0.035). Again, however, the fact that the later-onset group included more patients referred from some international collaborators working in adult services who were less likely to be able to provide triome samples and/or full phenotypic data may be a confounder here.

Figure 3.3 *Number of participants with positive or negative results classified by whether CP-like features were present. Positive findings were present in 33% of the non-CP-like group; 51% of the CP-like group; and 61% of the progressive group.*



** Onset at over three years of age and/or purely paroxysmal movement disorder.*

3.5 Comparison with other cohorts

My rate of diagnosis is broadly comparable with the rates achieved in other published cohorts, but this comparison is not necessarily meaningful: both results and cohort characteristics have varied wildly. To assess diagnostic yield of broad-spectrum genetic testing in other relevant cohorts, I performed a literature review as follows. I undertook a series of PubMed searches in October 2022 using the search term: (exome OR WES OR WGS OR “whole genome” OR “gene panel” OR “genetic panel” OR CNV OR microarray OR “copy number”) AND (“cerebral pals*” OR “movement disorder” OR dyston*).

This yielded 895 results. I used the following inclusion criteria:

- Study reports primary research (i.e. not meta-analyses, reviews etc);
- Researchers used a broad-spectrum genetic test (WES, WGS, microarray, panel testing or a combination of these);
- Study reports on a cohort rather than targeted investigation of two or fewer families;
- Participants included people with movement disorders (but were not necessarily limited to these: several studies recruited a cohort of people with various neurological conditions);
- Participants included children, or people with childhood-onset conditions (but were not necessarily limited to these);
- Cohort was not primarily defined by a non-movement disorder clinical, laboratory or radiological characteristic (e.g. studies of populations with developmental and epileptic encephalopathy, or brain calcification, were excluded even though they included some individuals with movement disorders).

This resulted in a list of 43 relevant studies. Their characteristics and diagnostic yields are summarised in **Table 3.3.**

It will be seen that diagnostic yields vary widely within all testing modalities **(Figure 3.4)**. There is an impression that WES and WGS achieve higher yields than CNV-only approaches or panel testing, as might be expected, but the diversity of the cohorts is such that it would be inappropriate to attempt statistical analysis. The highest-yield

studies have in common a highly pre-selected population: all those achieving a yield of over 60% are limited to participants with a very high index of suspicion such as people with cerebral palsy with no sentinel event or typical MRI,(131) patients already attending a genetics clinic,(132) or children with both a movement disorder and an epileptic encephalopathy.(133) The lowest yields are found in less selected populations including a high proportion of participants likely to have a non-genetic aetiology: other than those using a CNV-only approach, all the studies with <10% yield were in unselected CP cohorts.

Figure 3.4 *Diagnostic yield of studies using broad-spectrum genetic testing in movement disorder cohorts. Red: CNV only; purple: panel tests; blue: WES/clinical exome; green: WGS.*

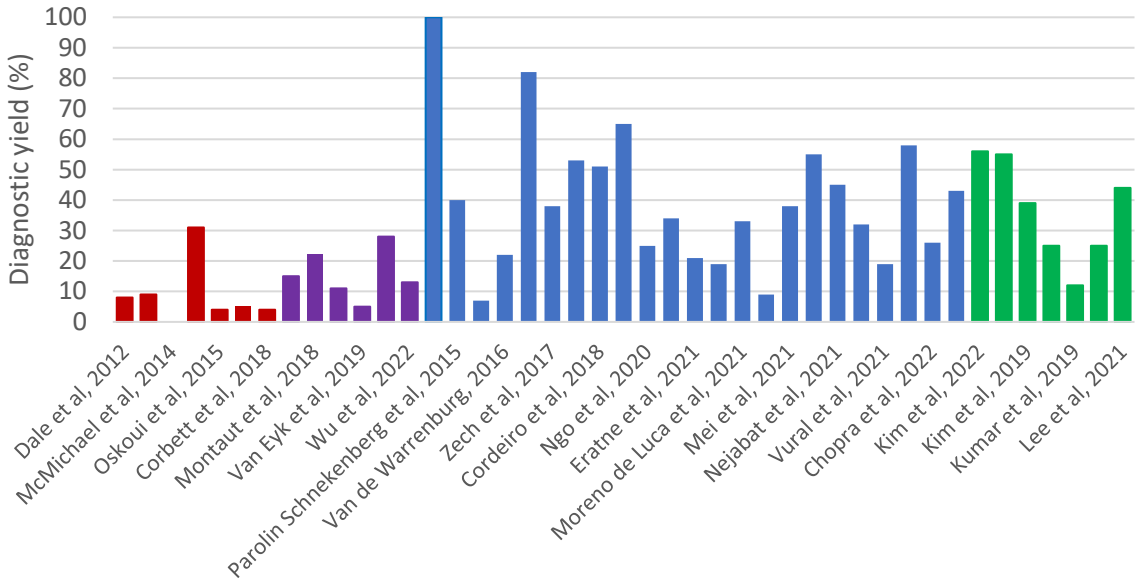


Table 3.3 *Rates of diagnosis in various movement disorder cohorts investigated using broad-spectrum genetic tests, arranged by test type and date*

Study	Method	Diagnosed/total (%)	Cohort characteristics	Notes
Dale et al, 2012	CNV only	2/25 **(8)**	Suspected genetic MD	3 additional participants had VUSs
Howell et al, 2013	CNV only	20/215 **(9)**	Various MD	10 additional participants had VUSs
McMichael et al, 2014	CNV only	0/50 **(0)**	Unselected CP	.
Segel et al, 2015	CNV only	16/52 **(31)**	"Cryptogenic" CP only	.
Oskoui et al, 2015	CNV only	5/115 **(4)**	Unselected CP	.
Zarrei et al, 2018	CNV only	5/97 **(5)**	Hemiplegic CP	Authors believe CNVs might be relevant in up to 23.7%.
Corbett et al, 2018	CNV only	7/186 **(4)**	Unselected CP	This includes the cohort from McMichael et al (2015), below.
Van Egmond et al, 2017	Panel	9/61 **(15)**	Dystonia	.
Montaut et al, 2018	Panel	83/378 **(22)**	Suspected genetic MD	.
Reale et al, 2018	Panel	25/221 **(11)**	Adult and childhood MD	.
Van Eyk et al, 2019	Panel	14/271 **(5)**	Unselected CP	.
Graziola et al, 2019	Panel	42/148 **(28)**	Childhood-onset MD	.
Wu et al, 2022	Panel + CNV	40/318 **(13)**	Various dystonia, including adults	.
Srivastava et al, 2014	WES	11/11 **(100)**	Patients with "CP-like encephalopathy" presenting to a neurogenetics clinic	.
Parolin Schnekenberg et al, 2015	WES	4/10 **(40)**	Ataxic CP	.
McMichael et al, 2015	WES	7/98 **(7)**	Unselected CP	Authors quote a higher resolution rate, but this includes unconfirmed novel genes.
Van de Warrenburg, 2016	WES	17/76 **(22)**	All HSP or cerebellar	"Clinical exome"
Kobayashi et al 2016	WES	9/11 **(82)**	All both early-onset epileptic encephalopathy and MD	.
Zech et al, 2017	WES	6/16 **(38)**	Early-onset generalised dystonia	.
Takezawa et al, 2018	WES + CNV	9/17 **(53)**	CP in people born at term without specific MRI changes	No relevant CNVs. All triomes.
Cordeiro et al, 2018	WES	26/51 **(51)**	Paediatric genetic MD clinic patients	Some had a panel first
Matthews et al, 2019	WES	32/49 **(65)**	"Atypical" CP: increased index of suspicion for genetic aetiology	.

Ngo et al, 2020	WES +CNV	65/260 **(25)**	Cerebellar ataxia, sporadic, mostly adult	Another 27% had findings described as “relevant”
Wirth et al, 2020	WES	11/32 **(34)**	Early-onset complex or familial dystonia	Participants had been undiagnosed after Montaut et al, 2018
Eratne et al, 2021	WES	34/160 **(21)**	Various child and adult neurological disorders	
Zech et al, 2021	WES	135/728 **(19)**	Dystonia	Genetic causes commoner when other neurological features present
Moreno de Luca et al, 2021	WES + CNV	499/1526 **(33)**	Unselected CP (but see notes)	Included patients referred for clinical genetic testing, in whom yield was higher than others.
May et al, 2021	WES	14/151 **(9)**	Unselected CP	Higher yield in those without risk factors (not statistically significant).
Mei et al, 2021	WES + CNV	78/217 **(38)**	Unselected CP	SNVs 5X commoner than CNVs
Rosello et al, 2021	WES	11/20 **(55)**	“Idiopathic” CP only	.
Nejabat et al, 2021	WES	30/66 **(45)**	“Atypical” CP only: increased index of suspicion for genetic aetiology	.
Kwong et al, 2021	WES	10/31 **(32)**	Idiopathic paediatric movement disorders	.
Vural et al, 2021	WES + STR	251/1296 **(19)**	Hereditary ataxia	.
Yechieli et al, 2022	WES + CNV	26/45 **(58)**	“Cryptogenic” CP only	All triomes. 18 point mutations; 8 CNVs.
Chopra et al, 2022	WES	13/50 **(26)**	Unselected CP and “CP masqueraders” (progressive)	Some genetic diagnoses made in people with perinatal risk factors.
Thomas et al, 2022	WES + CNV + mitochondrial	29/67 **(43)**	Various progressive neurological disorders neuro, all progressive	.
Kim et al, 2022	WES	42/75 **(56)**	Paediatric movement disorders	.
Martinez Rubio et al, 2022	WES	29/55 **(55)**	MD, mostly NBIA	Panel first, WES on those remaining undiagnosed
Kim et al, 2019	WGS + CNV	7/18 **(39)**	Spastic ataxia/HSP	None of the diagnoses were CNVs
Odgerel et al, 2019	WGS + CNV	2/8 **(25)**	Familial tremor only	.
Kumar et al, 2019	WGS + CNV	13/111 **(12)**	Dystonia, including adult and sporadic	3 diagnoses were CNVs
Van Eyk et al, 2021	WGS	37/150 **(25)**	Unselected CP	Relevance of some findings unclear, e.g. genes for bleeding diathesis. Mostly *not* triomes.
Lee et al, 2021	WGS	94/214 **(44)**	Childhood neurological disorders, suspected genetic	Lower rate of diagnosis in MD than in epilepsy and neurodevelopmental in this study

Abbreviations: CP: cerebral palsy; HSP: hereditary spastic paraplegia; MD: movement disorder; NBIA: neurodegeneration with brain iron; STR: short tandem repeat testing; VUS: variant of uncertain significance. References: (88, 131-172)

Another factor which will be important in determining yield is how much the participants had already been investigated at the point of recruitment: inevitably, as more investigations are performed on the same patients a degree of diminishing returns must be expected. In my project, I aimed to recruit people who had already been extensively investigated, ideally including a movement disorders-focused panel or exome sequencing. This is a higher level of investigation than almost any of the published studies, and had I been wholly successful in this aim, my yield would doubtless have been much lower than it was. In fact, as I recruited a number of participants who had had limited genetic investigations (for example, from collaborators working in health systems where access to diagnostic testing was limited), a proportion had straightforward diagnoses which did not necessarily require a detailed research analysis to identify. The fact that I preferentially recruited highly-investigated patients accounts for the prevalence of ultra-rare disorders in my cohort.

A major factor favouring a high diagnostic yield in my cohort include the recruitment of (almost) exclusively people with childhood-onset movement disorders which are clinically suspected to be genetic in origin. However, the high proportion of panel- and exome-negative patients included counteracts this to some extent. Numerical comparisons with other cohorts are therefore unlikely to be meaningful. Having said that, a rate of clinically relevant findings in excess of 45% clearly indicates that even highly-investigated cohorts can be worth detailed reanalysis.

3.6 Dilemmas in reporting

A few cases presented specific dilemmas, ethical and practical, about reporting. I will discuss one of these in order to explore the issues which may arise in projects of this type, and in broad-spectrum genetic testing more generally.

Case History: NB7

NB7P had no family or perinatal history of note. At six months old she presented to paediatric services with recurrent twitching and jerking movements which were not felt to be epileptic in origin. Intermittent involuntary movements persisted throughout childhood and she also developed a fine postural tremor. She had mild developmental delay during her early years and was subsequently diagnosed with dyslexia and dyspraxia, but could attend mainstream school. MRI brain showed bilateral basal ganglia signal changes; CT confirmed that these were calcification rather than iron. There was also a degree of cerebellar vermis atrophy.

As an adult, NB7P was able to live and work independently. She continued to have involuntary movements and sometimes blackouts. In her mid-twenties she received a diagnosis of functional neurological disorder (FND). Treatment for this condition, by a neuropsychiatric team, led to a major improvement in her symptoms although they did not resolve entirely.

I undertook reanalysis of NB7's genome some years after her initial recruitment. At that point the clinical information available to me was of her childhood presentation without details of the FND identified in adulthood. I identified a heteroplasmic mitochondrial variant in *MT-ATP6* as follows: m.9134A>G, p.Glu203Gly. Heteroplasmy in blood was about 44% and it was absent in the mother. In homoplasmy, this variant has been reported as causative for complex V deficiency manifesting with severe Leigh-like symptoms in infancy.(173)

I made contact with the participant's current clinical team and at this point discovered her good response to treatment for FND and mild current symptoms. This raised two questions: firstly, was the heteroplasmic variant indeed pathogenic, and secondly, if significant doubt remained on this point, would it be ethical to report it to the participant, given the likely resulting anxiety and distress? The adult neurological team were understandably not happy to take forward discussion and investigation of the research finding, so despite the patient's age the responsibility remained with us.

To answer the question of pathogenicity I discussed the participant in the GOSH Mitochondrial Multidisciplinary Team meeting. They agreed that heteroplasmic results

are hard to interpret with certainty but felt that, at a minimum, further investigation was required. The scan findings were consistent with, though not specific for, mitochondrial disease,(174) as indeed were the symptoms. The very early onset ruled out a fully functional explanation for the whole presentation, making it more likely that a mitochondrial disorder was at least contributory. It was also noted that even though the significance of the finding for the participant herself was unconfirmed, it would be ethically questionable to withhold information about a known pathogenic mitochondrial variant in heteroplasmy from a young woman of childbearing potential: a future child would risk homoplasmic inheritance of the variant, resulting in severe and possibly fatal disease.

After discussion with Prof Kurian and the mitochondrial team, we decided to disclose the finding to the participant, with the explanation that its significance remained unclear. She appeared quite satisfied with this and willing to undergo further investigation, for which we referred her to an adult mitochondrial team. Their assessment is awaited at the time of writing

This case illustrates two points. One is the importance of taking into account both detailed clinical data and knowledge about the specific gene and variant in reaching a conclusion about likely pathogenicity. The second is the need for careful communication with both the participant and other clinical teams in genetic research, to ensure findings are acted on constructively and, where necessary, a safe handover of care takes place.

3.7 Collaborations and case series

As part of the analysis process, if no clear diagnosis was reached after an initial review, but there were variants of interest in genes without known disease phenotypes, I listed these on GeneMatcher.(124) In some cases this led to the work described in Chapters 4 and 5, led primarily by me and the Kurian group team. In others, however, I learned that another team was already working on a gene and I could contribute to their project.

There are three genes where such projects have reached a stage where I feel confident that the proposed gene-disease relationship is likely to prove genuine (multiple unrelated cases identified with a coherent phenotype and/or supportive functional work) but so far only one of these, *TNPO2,* has been published. A fourth gene, *CCDC82*, has now been published by two other groups independently of our work.(175, 176)

3.7.1 *TNPO2:* a new neurodevelopmental disorder

Case History: PE26

PE26P is an Australian boy who was born at term after an uneventful pregnancy with no relevant family history. He presented at three months due to failure to hold his head up and lack of visual reciprocity. At eight months old he had global developmental delay including persistent head lag and experienced episodes of hyperkinetic limb movements on a daily basis. He required gastrostomy feeding due to dysphagia. On examination he had truncal hypotonia, microcephaly and intermittent stereotyped movements but no frank dystonia. MRI brain showed mild ventriculomegaly, a dysplastic appearance of the cerebellum and slightly hypoplastic caudate nuclei.

TNPO2 encodes transportin-2, a protein involved in nuclear transport. The gene is predicted to be intolerant of heterozygous change (pLI 1, missense o/e 0.28).(177) PE26's WGS did not reveal any variants likely to be pathogenic in known disease genes but did find a *de novo* mosaic variant in *TNPO2* as follows: NM_001136196.1: c.455_457del, p.Lys152del. Mosaicism in blood-derived DNA was estimated at 21% on WGS and 16% on Sanger sequencing; no other tissues were available for analysis. The variant was absent from normal population databases and *in silico* tools predicted pathogenicity.

Goodman *et al* collected a cohort of 15 unrelated individuals with heterozygous *de novo* variants in *TNPO2*, all of whom had neurodevelopmental impairment, with muscle tone abnormalities and feeding difficulties in the majority.(178) Three others had documented abnormal movements (tremor or choreiform movements). Six of the variants, including PE26P's, occurred in the Ras-related nuclear protein (RAN) binding domain, which is essential for the protein's function as a nucleocytoplasmic shuttle:

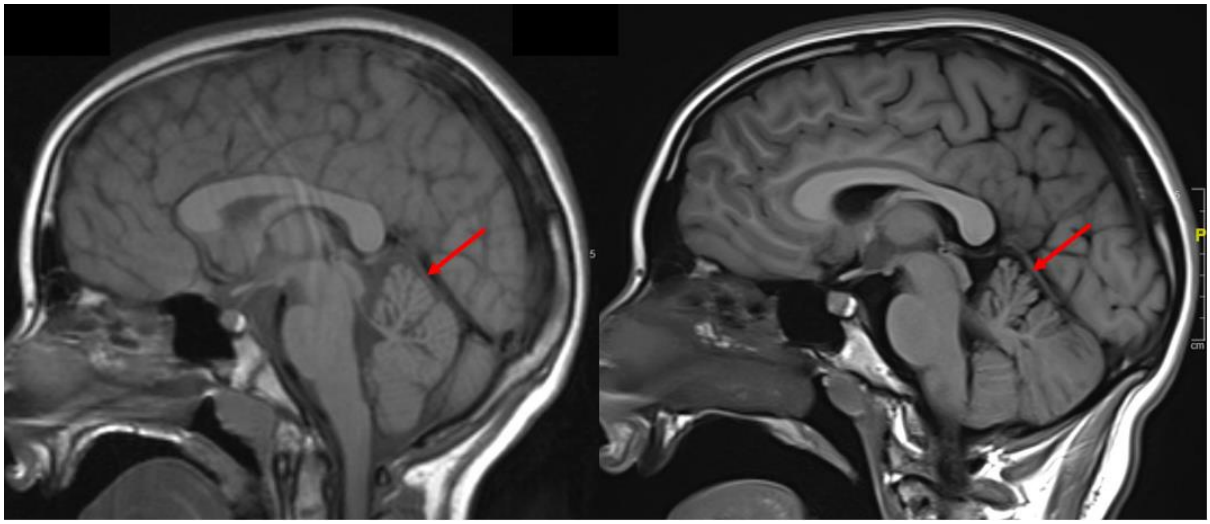
variants in this domain appeared to be associated with more severe phenotypes. The team demonstrated that both knockdown and ectopic expression of transportin-2 in *Drosophila* led to developmental defects and hypothesised that a subset of patient variants caused toxic gain-of-function effects. I am a co-author on the index report on this condition, which has now been designated “intellectual developmental disorder with hypotonia, impaired speech and dysmorphic facies” (MIM 619556).

3.7.2 *H2AC6*: a histone disorder with a variable phenotypic spectrum

H2A is one of the core histones which forms the nucleosome core particle, and *H2AC6* encodes one of the canonical H2A histones, H2A1C or HIST1HA2AC. (The confusing “HIST1” refers to the location of the gene in histone gene cluster 1.)

For participant DY179, I undertook WGS on a uniparental basis as paternal DNA was not available, and my initial report was negative. I listed some variants of interest on GeneMatcher, including one in *H2AC6:* NM_003512.4:c.245G>A, p.Arg82His (heterozygous). The variant was not present in the mother but could have been either *de novo* or paternally inherited. It was absent from normal population databases and *in silico* tools predicted pathogenicity, including a CADD score of 31. Sanger sequencing confirmed the presence of the variant.

Figure 3.5: *T1-weighted sagittal MRI images from DY179P aged 11 years (left) and 16 years (right). Despite technical differences between the two scans, progressive atrophy of the superior cerebellar vermis can be seen (red arrow).*



Case History: DY179

DY179P is a teenage girl with no relevant family history. She was born at term in unexpectedly poor condition but her blood tests were not suggestive of an acute hypoxic insult. Her infant milestones were within normal limits but when she started walking at 18 months she toe-walked and appeared to be slightly ataxic. Her family were told it was possible she had ataxic cerebral palsy. At 11 years old she developed nocturnal seizures characterised by right-sided jerking, with an EEG typical for childhood epilepsy with centrotemporal spikes.

During childhood DY179P experienced "muscle spasms" and intermittent uncomfortable posturing of her hands and feet. These gradually worsened over the years: she remained independently ambulant but found walking increasingly difficult and tiring, and struggled with tasks such as handwriting and cutting up food. On examination aged 15 she was dysarthric and mildly ataxic, with subtle orofacial and limb dystonia. There were no dysmorphic or neurocutaneous features. She had no intellectual disability.

Serial MRI brain scans at the ages of 11 years and 16 years were normal except for progressively worsening atrophy of the cerebellar vermis.

Through GeneMatcher I made contact with Dr Heather Mefford, a paediatric geneticist based in Tennessee. Her team was putting together a case series of patients with neurodevelopmental impairments and epilepsy. Initially we both felt unsure whether DY179P was suitable for inclusion, but in summer 2022 two other teams got in touch who were aware of patients with *H2AC6* variants who had progressive cerebellar atrophy and extrapyramidal signs but minimal or absent ID. These included a family with three affected members. This close phenotypic correspondence in three unrelated families made me more confident that DY179P's variant was likely to be causative.

A manuscript describing the phenotype of an autosomal dominant *H2AC6* disorder with a variable phenotypic spectrum including ataxia, dystonia, epilepsy and ID in some cases but not all, together with *in silico* modelling of variants, is currently in preparation by Dr Mefford, her clinical fellow Dr Wei-Liang Chen, and their team, based at St Jude's Children's Research Hospital in Tennessee. The disorder will join the growing catalogue of recognised movement and neurodevelopmental disorders

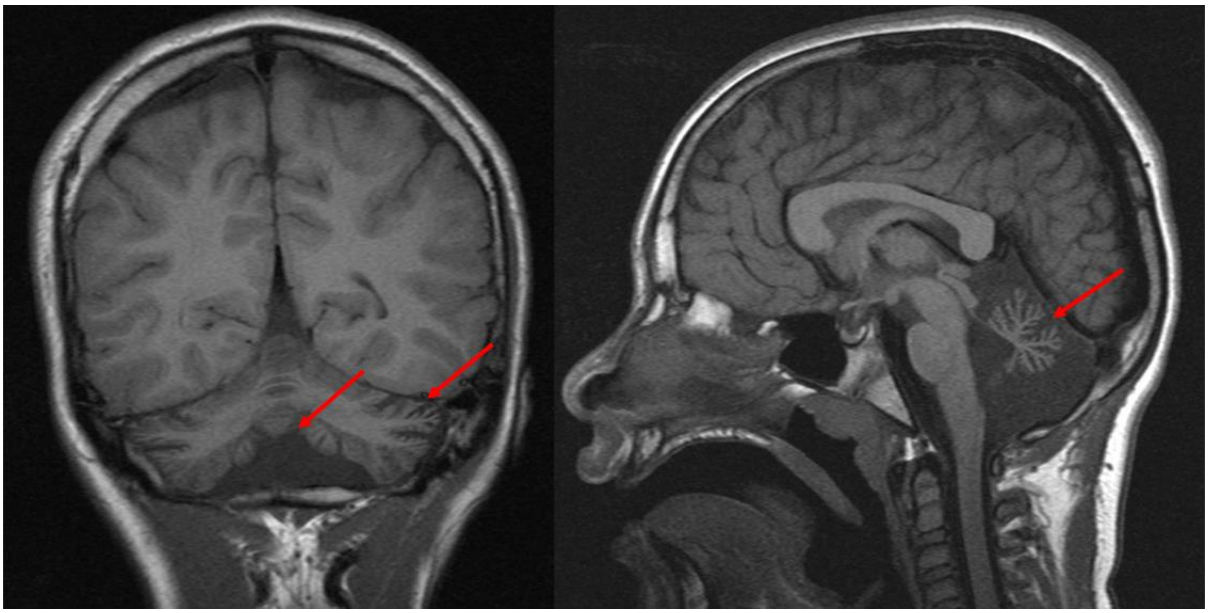
associated with variants both in histone proteins and histone modifiers such as *KMT2B*.(179, 180)

3.7.3 *INTS11:* another Integrator Complex subunit implicated in neurological disease

The integrator complex is a highly conserved protein complex consisting of 14 subunits, involved in RNA processing. Defects in genes encoding two of the subunits *INTS1* and *INTS8,* are already implicated in autosomal recessive disorders characterised by neurodevelopmental impairment, cerebellar dysfunction and ocular abnormalities.(181, 182)

The AX10 family had been recruited for WGS some years before my project started but initial analysis had been non-diagnostic. A separate analysis through the 100,000 Genomes Project had also not yielded any result (although I am told that the Clinical Genetics team was aware, and suspicious, of the variants I subsequently found, but was not in a position to report them or to investigate further).

Figure 3.6 *T1-weighted MRI images from AX10P1, aged 11 years: coronal (left) and sagittal (right). Profound hypoplasia of the cerebellum, especially the vermis, is indicated by red arrows.*



Case History: AX10

AX10P1 (a boy) and AX10P2 (a girl) are the eldest two children of healthy unrelated parents of Bangladeshi origin. Both were born by emergency caesarean section due to concerns about fetal distress, A at 33 weeks' and B at 38 weeks' gestation, but both were delivered in good condition. Global developmental delay was noted in both siblings during the first year of life. AX10P1 sat at 12 months and rolled at 18 months. AX10P2 crawled at 13 months. Neither sibling walked independently but both learned to walk with an ataxic gait using a frame. Both learned to speak in short sentences and could write some letters. During their teenage years, both gradually developed increasing difficulty with motor function and required more support to walk. AX10P1 had one seizure at the age of 14 years, but AX10P2 never had any clinical seizures.

On examination, both had an unsteady broad-based gait, jerky eye movements with horizontal nystagmus, hypotonia and brisk deep tendon reflexes. AX10P2 had an intention tremor with past-pointing and dysdiadochokinesis.

Ophthalmological assessment revealed retinal dystrophy in both siblings. Muscle biopsy in AX10P2 found moderately low complex I, II and III with normal complex IV. MRI showed progressive cerebellar and pontine atrophy and subtle signal changes in the cerebral white matter.

Once my project was established, I reanalysed data from a number of families where there was a strong suspicion of an undiagnosed genetic disorder, especially those where two or more siblings were affected. In the case of AX10, this revealed that both siblings were compound heterozygous for variants in *INTS11* as follows: NM_001256456.1:c.1237C>T, p.Pro413Ser and c.52G>A, p.Gly18Ser. One variant was absent from gnomAD and the other present at very low levels in heterozygosity only, and *in silico* predictions for both supported pathogenicity.

The initial analysis had been on the two siblings only (as is our usual practice in multiplex families for reasons of cost): paternal DNA was not available, but Sanger sequencing confirmed that one variant (c.52G>A) was maternally inherited while the other was not, which provides strong supporting evidence that the two variants are in *trans*.

Through GeneMatcher, I made contact with Dr Erica Macke, a geneticist based at Nationwide Children's Hospital in Ohio, who together with others has collected data on 16 individuals from 11 families with biallelic *INTS11* variants and a phenotype including neurodevelopmental impairment, progressive cerebrocerebellar atrophy and in many cases retinal and/or optic nerve abnormalities. A manuscript is in preparation.

3.7 Discussion

In this chapter I have outlined the overall results of my analysis, including the rate of diagnostic and potentially diagnostic findings and the practical and ethical dilemmas which can arise in interpreting and reporting variants. I have also briefly described three instances where I contributed to identification of a new gene-disease relationship: in the subsequent chapters I will discuss genes where, rather than simply contributing genotypic and phenotypic data, I and colleagues in the Kurian lab led the gene discovery.

A large number of published reports now confirm the usefulness of WGS and other broad-spectrum genetic tests in investigating neurological disease, including movement disorders. These studies have been too diverse in methodology for a meta-analysis to yield useful results, but the overall conclusion that diagnostic yields are significant, including in more complex and challenging populations, is clear. Moreover, in this project I came across one identifiable new genetic disorder for every 28 genomes (or triomes) analysed. I will discuss the implications of this more fully in my final chapter: it clearly supports the use of WGS as a tool not just for diagnosis but for gene discovery.

Chapter 4: Expanding and clarifying phenotypes of established rare disorders

4.1 Introduction

In a naïve understanding of Mendelian genetic disorders, defects in a specific gene result in a specific disease. Conversely, a specific phenotype, if sufficiently well-characterised, can be traced to a specific gene. Of course, this is sometimes the case, but equally often the relationship between genes and phenotypes is variable, complex, and hard to predict. (I am writing here of monogenic disorders: the contribution of genetic factors to multifactorial disease is of course even more complex, and beyond the scope of this thesis.)

There are many instances where, depending on the exact nature of the pathogenic variant, defects in a single gene can give rise to totally different phenotypes, a phenomenon known as phenotypic pleiotropy. For example, different mutations in *KIF1A* can cause a range of conditions ranging from a pure hereditary sensory and autonomic neuropathy (found with protein-truncating variants) to a profound infant-onset form of neurodegeneration (caused by certain missense variants).(183) In other situations, identical genetic variants can cause different disease phenotypes depending on other genetic and/or environmental factors in the individual in which they occur – for example, the same *SCN1A* variant can give rise to the relatively mild generalized epilepsy with febrile seizures plus in a parent, but full-blown Dravet syndrome in their child.(184) There are also well-defined clinical entities which may be due to defects in any of several genes, distinguishable only by genetic testing – Aicardi-Goutières syndrome, caused by variants in any of at least nine different genes,(185) some of which have radically different physiological functions, is a good example. This is known as genetic heterogeneity.

This matters for at least three reasons. Firstly, if a particular genotype-phenotype association has not been documented, a person presenting may remain undiagnosed: either the gene in question will not be analysed at all, or any variants found will be dismissed as irrelevant. For example, if variants in a gene are known to cause epilepsy

but not dystonia, a dystonic child may be investigated using a panel which excludes that gene. I will discuss this further in the sections on *RHOBTB2, STXBP1* and *YY1* variants.

Secondly, effective counselling of patients and families relies on accurate knowledge of the phenotypic spectrum. Where a potentially pathogenic variant has been found, especially antenatally or in a young child, the family will wish to know the likely prognosis and the range of possibilities, rather than simply the genetic diagnosis. I will elaborate on this in the sections on *SLC30A9, H3.3* and *GNAO1* variants.

Thirdly, different classes of variants, even in the same gene, may act through quite different pathophysiological pathways – as I shall discuss below in the section on *JPH3* variants. Approaches to management and therapeutic research may therefore be entirely different. Again, *SCN1A* provides an excellent example: in Dravet syndrome, related to loss-of-function variants, sodium channel blockers must be avoided as they may exacerbate seizures. In the rarer neonatal epileptic encephalopathy caused by *SCN1A gain*-of-function, sodium channel blockers may be highly effective.(186)

For all these reasons, accurate and comprehensive delineation of disease phenotypes and genotype-phenotype relationships seems to me just as important as the identification of disease phenotypes for completely “new” genes. Making, explaining and acting on molecular diagnoses requires far more than just knowing the name of the implicated gene.

In this chapter, I will highlight three genes – *RHOBTB2, JPH3* and *SLC30A9* – where, although an association with neurological disease had previously been reported, this had not fully captured the range or nature of the condition: and my work aimed to address this. For *RHOBTB2*, this meant exploring aspects of the syndrome’s presentation which had been overlooked due to the focus on epilepsy. For *JPH3*, I reported disorder which is both genetically and phenotypically distinct from the main phenotype described for the gene. For *SLC30A9*, I reported on four more families for a condition where only two had previously been reported. For three other genes,

STXBP1, GNAO1 and *H3.3*, we explored aspects of the phenotype which had not previously been fully described.

This chapter incorporates material from several manuscripts which have been published or are under review: this is indicated where appropriate. I was the first author on the papers discussing *JPH3* and *SLC30A9*; joint first author for *RHOBTB2*; second author for *STXBP1* and *YY1*, and an additional author for *GNAO1* and *H3.3.* Details of variant shortlisting for participants mentioned in this chapter can be found in **Appendix 3**.

4.2 *RHOBTB2*-related disorders: not just epileptic encephalopathy, also alternating hemiplegia of childhood

Case History: DY131

DY131P is the eldest of five children of healthy unrelated Indian parents. There is no known relevant family history. Pregnancy, birth and development during the first year of life were all unremarkable. Concerns about her intellectual development were raised when she went to nursery and she was subsequently diagnosed with moderate intellectual disability: she attended mainstream school with one-to-one support. She was also felt to be emotionally labile.

At three years old she had an episode of vomiting and encephalopathy with reduced consciousness level requiring a short period of intubation and ventilation, for which no immediate explanation could be found. EEG was encephalopathic and showed asymmetry, but was not frankly epileptiform, and was asymmetrical. After this event she continued to have occasional paroxysmal events consisting of unilateral or bilateral facial drooping, weakness, drooling, and looking vacant but not losing awareness. She also had some episodes of hand "stiffness" and stereotyped movements such as hand-rubbing.

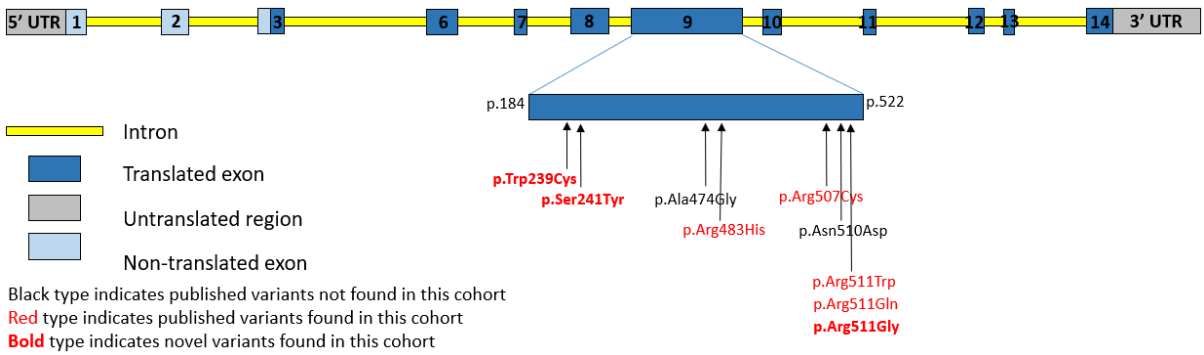
On examination at 14 years old she showed mild distal choreiform movements of the upper limbs, subtle myoclonus and dystonic posturing of the hands. Gait was normal except that she had some difficulty with tandem walking. There were no dysmorphic features.

MRI and inter-ictal EEG did not show any convincing abnormalities.

The autosomal dominant neurological disorder caused by variants in *RHOBTB2,* which encodes a Rho GTPase, was first described in 2018 and was designated "developmental and epileptic encephalopathy 64" (DEE64).(187, 188) It classically presents with global developmental delay and onset of epilepsy in the first three years of life. For many affected people, however, epilepsy is not the most prominent or disabling feature, and in some it does not occur at all. Instead, many experience a complex, polymorphic movement disorder which had not previously been described in detail. The work described in this section has been published: I am joint first author of the manuscript but it was of course a collaborative effort by all the co-authors.(87)

DY131P was in her teens when she was recruited to our study. WGS revealed no biallelic variants likely to be pathogenic. Analysis of heterozygous *de novo* variants, using the megapanel, showed a novel *RHOBTB2* variant (c.722C>A, p. Ser241Tyr) which was predicted to be pathogenic by 4/4 *in silico* tools, with no other likely candidates (see **Appendix 3**). If her presentation was considered compatible with DEE64, the variant could be classified as pathogenic based on criteria PS2 (*de novo*), PM1 (variant located in a mutational hot-spot), PM2 (absent from controls) and PP3 (supported by multiple lines of computational evidence. However, DY131P had no history of epilepsy. Fortunately, we were already aware through personal communications that some children with *RHOBTB2* variants could present with movement disorders and/or encephalopathic episodes.

Figure 4.1 *Gene diagram showing location of pathogenic variants in RHOBTB2 (NM_00160036.1). Image appears in Zagaglia et al, 2021.(87)*



We went on to describe the presentation of 11 people with pathogenic *RHOBTB2* variants in whom movement disorder was a feature – often the most prominent feature. DY131P was included in this cohort. We recruited participants with a *RHOBTB2* disorder diagnosis from the NMDC and international collaborators and collected detailed phenotypic information on them, including video where possible. All 11 participants had a heterozygous missense *RHOBTB2* variant. Testing both parents was possible for 9/11 and in all of these *de novo* occurrence was confirmed. 7/11 of the variants involved the same amino acid (Arg511) and all the variants were found within exon 9 (NM_001160036.1) (**Figure 4.1**). Three of the variants (including DY131P's) were novel.

Only 7/11 individuals had a diagnosis of epilepsy, although two others had experienced at least one seizure (electrographic without clinical correlate in one instance; a febrile convulsion in another). 5/7 had experienced status epilepticus.

All participants had some degree of motor abnormality, with a wide range of severity from complete lack of head control in one individual to only mild balance difficulties in one ambulant patient. Strikingly, 10/11 patients had paroxysmal abnormal movements: these included hemiplegic episodes in eight, focal dystonia in seven, episodic ataxia in four, episodic exacerbations of dyskinesia in four, and non-epileptic myoclonus in two. 5/11 also had a baseline choreiform hyperkinetic disorder.

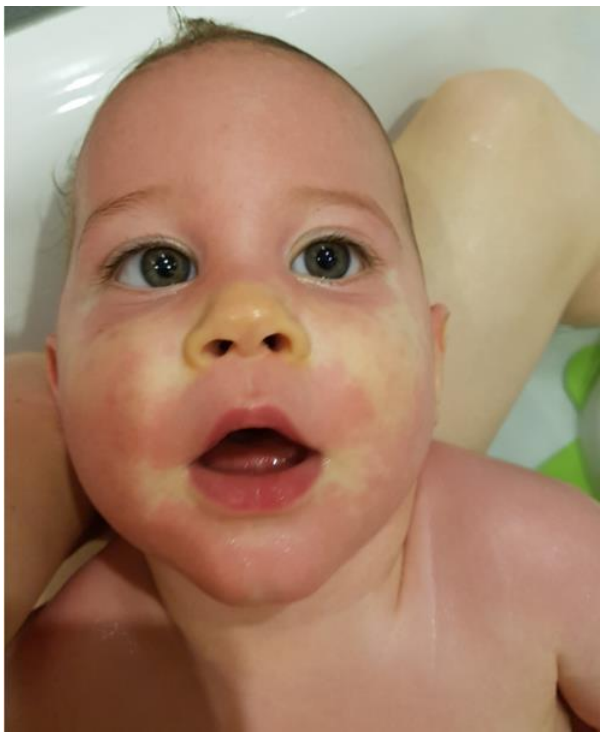
Events which had been thought to be epileptic seizures in fact proved to have no electrographic correlate in the majority of patients (6/11, with another 3/11 where most event types were never captured on EEG). These events included myoclonus and paroxysmal lateral deviation of the head and eyes: in many instances they did indeed appear highly suggestive of epileptic seizures.

Several patients exhibited most of the features of alternating hemiplegia of childhood (AHC). The key difference was that hemiplegic episodes were not reported in infancy: the youngest occurrence of hemiplegia in our cohort was in a child of 20 months, whereas the criteria for classical AHC (usually due to *ATP1A2* or *ATP1A3* variants) specify commencement of these episodes by 18 months of age(189) – although a

recently-suggested revision of the criteria would mean *RHOBTB2*-related cases were included.(190) Hemiplegic episodes in *RHOBTB2* disorders were usually described as unilateral or asymmetrical weakness accompanied by what parents described as a 'vacant appearance' (though without loss of responsiveness to external stimuli), associated with drooling and reduced activity. Duration varied from minutes to weeks. Some episodes, although by no means all, followed an identifiable trigger such as a minor head injury or a seizure. Where available, EEG during these episodes was suggestive of encephalopathy, but not epileptiform activity, in the contralateral hemisphere.

In some instances, paroxysmal events (both abnormal movements and epileptic seizures) were accompanied by thermoregulatory changes, vomiting, striking flushing and/or pallor of the skin (**Figure 4.2**) and in one instance respiratory disturbances severe enough to require bag-valve-mask ventilation. We hypothesized that these might represent autonomic disturbances, but the mechanism remains unclear.

Figure 4.2 *Patchy flushing and pallor of the skin which appeared episodically in a child with a* RHOBTB2*-related disorder. Image adapted from Zagaglia et al, 2021.(87)*



All patients had intellectual disability (ID): formal psychometric testing was not available for most patients but for 10/11, ID was clinically assessed as moderate or severe, with either absent or significantly impaired verbal communication and comprehension. One patient – our own DY131P – had much milder impairments than the rest of the cohort and was able to attend mainstream school with support: in this, as in her other symptoms, she appeared to have attenuated but recognisable versions of the symptoms experienced by the other participants, such as mild “hemiplegic” episodes consisting of unilateral facial weakness only. Only two patients experienced definite developmental regression, and this occurred in the context of either status epilepticus or an exacerbation in the frequency and severity of paroxysmal movements. In three other patients, developmental slowing was noted after the onset of epilepsy, which in all these cases was very early (3 months of age, or less): milestones (such as social smiling, visual attentiveness and partial head control) prior to this age had been apparently normal, but were delayed afterwards.

There is no good evidence for any specific medication in *RHOBTB2* disorders: the majority of patients in our cohort were using anti-epileptic medication for both epilepsy and control of paroxysmal movements, with varying success. No agent was reported to be particularly helpful by more than two families: those that were helpful to some include carbamazepine, topiramate and acetazolamide.

Thus it is clear that while epilepsy forms a major component of the symptomatic burden of *RHOBTB2* disorders it is far from the only issue, and often not the most important one from the patient’s point of view. A number of individuals, like DY131P, did not have epilepsy at all but nevertheless experienced significant disability as a result of their movement disorder and/or neurodevelopmental impairment. It appeared that both epilepsy and paroxysmal abnormal movements correlated with periods of developmental stagnation or regression.

From the clinical point of view, the importance of understanding the movement phenotype of *RHOBTB2* disorders is twofold. Firstly, it enables the diagnosis to be considered, and variants of uncertain significance to be interpreted accurately, in people presenting with movement symptoms in the absence of epilepsy, or with a

clinical diagnosis of alternating hemiplegia of childhood. (Translation of this sort information into practice is not always rapid: unfortunately, I note that *RHOBTB2* is still not on most movement disorder panels, including the NHS Genomic Medicine "Childhood-onset dystonia or chorea or related movement disorder" panel.(106)) Secondly, realising that many of the paroxysmal events experienced in this condition are not true epileptic seizures, even though they may closely resemble them, can guide appropriate treatment. *RHOBTB2* disorder goes well beyond an epileptic encephalopathy, and controlling paroxysmal movements, and averting or treating hemiplegic/encephalopathic episodes, is quite as important as seizure management.

RHOBTB2 remains a disorder urgently in need of better description, diagnosis and treatment trials: the latter both in the form of pragmatic "trials" of existing agents such as anti-epileptic medication, flunarizine and acetazolamide, and targeted approaches such as gene therapy. Working with, and using the collective expertise of, the developing patient and family advocacy groups could be a way to achieve this.

4.3 *JPH3*: one gene, two diagnoses

JPH3 encodes junctophilin-3, a component of the junctional complex linking the plasma membrane with the endoplasmic reticulum in excitable cells.(191) Expression is highest in the brain. It has a known associated disease phenotype – Huntington's-like disease 2, which as the name suggests resembles Huntington's disease: it presents in adulthood with progressive chorea, dementia and atrophy of the cortex and basal ganglia.(192) This condition is caused by the expansion of a triplet-repeat domain, leading to toxic accumulation of abnormal protein which results in cell death.(193, 194)

It has been known for around a decade that in mice loss of *JPH3* function also results in an abnormal motor phenotype, both in full knockout and, more mildly, in haploinsufficiency.(195) The gene is also predicted to be highly intolerant of heterozygous loss of function.(177) Recently, a single patient with a homozygous truncating variant in *JPH3* and a neurodevelopmental disorder involving paroxysmal dystonia was described.(196)

In my WGS cohort, I identified two patients with loss-of-function variants in *JPH3,* heterozygous in one case and homozygous in the other. Pedigrees of both families are shown in **Figure 4.3**. Both presented with a phenotype of apparently non-progressive neurodevelopmental disorder and paroxysmal dyskinesia, with the manifestations being milder in the child with a heterozygous variant. These variants did not involve the region of the gene implicated in the triplet-repeat expansion disorder (**Figure 4.4**).

Figure 4.3 *Pedigrees of families with pathogenic loss-of-function variants in JPH3. Filled shape: affected. Double line: consanguineous union. Diagonal strikethrough: deceased. Arrow indicates proband. Figure appears in Steel et al (2022).(197)*

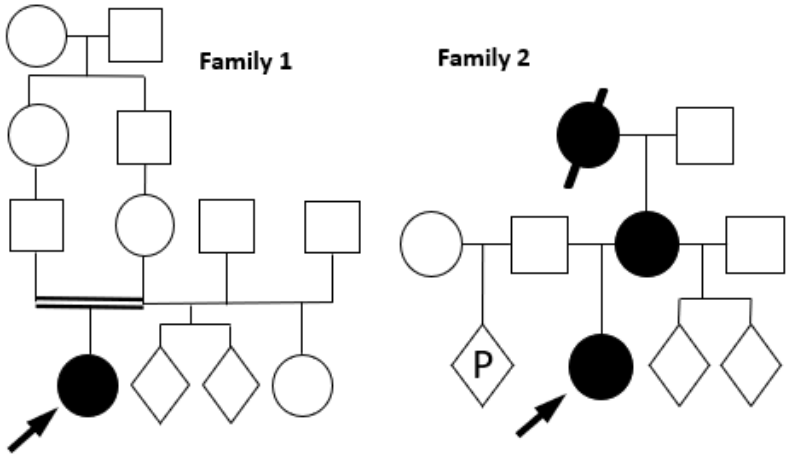
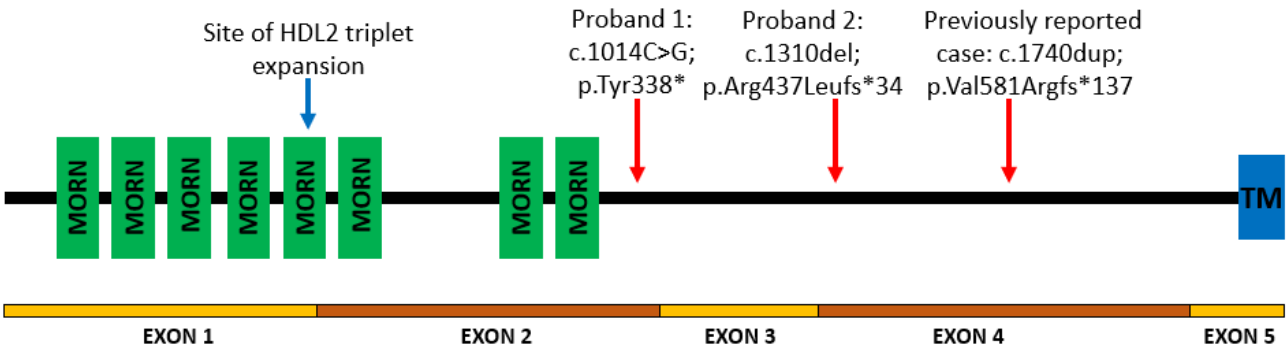


Figure 4.4 *Schematic structure of the* JPH3 *gene (below) and protein (above) showing key protein domains (MORN: Membrane Occupation and Recognition Nexus; TM: transmembrane domain); sites of pathological triplet repeat expansion (blue arrow); previously reported frameshift variant; and variants identified in my cohort (red arrow). Figure appears in Steel et al (2022).(197)*



Case History: PE5 (Family 1)

PE5P is now 15 years old and was born to first-cousin parents with no relevant family or perinatal history. Her mother was asymptomatic but the medical history of her father was unavailable. Early infant milestones were normal but she did not walk until three years old. At school she was identified as having moderate intellectual disability and behavioural difficulties. From six months old, her mother noticed episodes of abnormal body stiffening. Over time these evolved into two clear episode types. "Minor" episodes consisted of brief right or occasionally left hemidystonia which occurred several times each week. "Major" episodes happened two or three times per month and lasted up to thirty minutes: after a similar onset they progressed to involuntary flailing movements of all limbs, impaired responsiveness, dysarthria, aphasia and drooling, sometimes associated with upwards eye deviation. While "major" episodes tended to be triggered by fatigue, there was no obvious trigger for "minor" episodes. A clinical diagnosis of alternating hemiplegia of childhood (AHC) was made.

Neurological examination between episodes was unremarkable. There was no disease progression over time. There was no clinical response to L-dopa, oxcarbazepine, flunarizine, trihexyphenidyl, topiramate or gabapentin, although prolonged episodes could be terminated with buccal midazolam. Extensive investigation including brain MRI, EEG (with capture of a typical episode), and CSF neurotransmitters was normal.

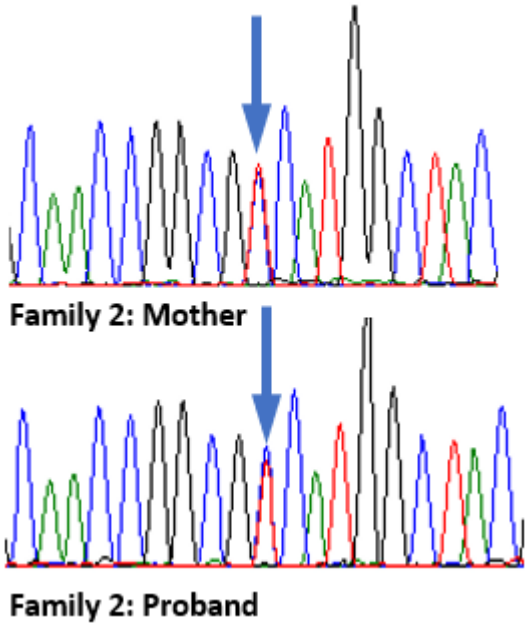
Case History: PE19 (Family 2)

PE19P, who is now 10 years old, experienced episodes of limb, facial and orolingual dyskinesia at least weekly from the age of six months. These usually started in the right upper limb before becoming generalised, and lasted for several hours. No abnormal eye movements were evident during the episodes and there was no obvious trigger. Between episodes, very mild resting dyskinesia was evident in the upper limbs. She had moderate learning difficulties but no behavioural issues. Again, extensive investigation was unremarkable, including EEG with capture of episodes. Several different medications – carbamazepine, valproate, topiramate, acetazolamide and trihexyphenidyl – proved ineffective. Her maternal grandmother had experienced lifelong episodes of orofacial dyskinesia and limb posturing and had never learned to read. Her mother reported episodes of orofacial dystonia triggered by alcohol, and had a slight resting tremor. She had completed compulsory education.

In PE5P, WGS identified a novel homozygous frameshift in *JPH3* (NM020655.3:c.1310del;p.Arg437Leufs*34), which was confirmed by Sanger sequencing. Parental DNA was not available for testing. The only other variant of interest was a homozygous *NDUFAF5* variant (NM_024120.5:c.524A>G; p.His175Arg) which was considered unlikely to be relevant both because the phenotypic fit for mitochondrial complex I deficiency was poor and in silico tools predicted that it was probably benign (see **Appendix 3**).

For PE19, the only relevant variant identified on WGS was a novel heterozygous nonsense variant in *JPH3* (NM020655.3:c.1014C>G;p.Tyr338*) shared by mother and daughter: grandmaternal DNA was unavailable (**Figure 3.5**).

Figure 4.5: *Electrophoretogram of Sanger sequencing of JPH3 gene in PE19,showing heterozygous variant shared by PE19 mother and daughter (indicated by arrow). Figure appears in Steel et al, 2022.(197)*



We reported the second AR case and the first AD family – although, as with any new observation of a genetic disorder, identification of a second family will be required before the association can be considered confirmed. The existence of overlapping AD and AR phenotypes is not unprecedented in neurological disorders (examples include mitochondrial membrane protein-associated neurodegeneration(198) and dopa-

responsive dystonia due to *GCH1* mutations(199)) and in this case the existence of a mammalian model with both AD and AR manifestations adds plausibility.

It appears to me that this represents not an expansion of the phenotypic spectrum of Huntington's-like disease 2 but a new disorder, differing both in pathophysiology and in clinical manifestations. Rather than being the consequence of progressive neuronal death and toxic misfolded protein accumulates, the condition we describe presumably results from the lack of functional juntophilin-3. Importantly, there is no reason to anticipate that this will be a progressive disorder, and the families were counselled that they should ignore descriptions of triplet-repeat associated *JPH3* disease. The full natural history and pathophysiology of the condition remain to be explored.

4.4 *SL30A9:* defining a spectrum

SLC30A9 encodes a cation transporter thought to be primarily involved in cellular zinc homeostasis, designated ZnT-9.(200) Recently a novel syndrome consisting of a movement disorder, neurodevelopmental regression, oculomotor apraxia and progressive renal impairment was described in a single large Bedouin kindred with a homozygous *SLC30A9* in-frame deletion.(201) A single individual from a second family with compound heterozygous variants has also recently been reported.(202) This condition has been designated Birk-Landau-Perez syndrome (BLPS).(203)

We identified two additional families with biallelic *SLC30A9* variants through WGS or WES (one, DP11, was a "legacy" family who had had WES and homozygosity mapping some years before I joined the lab: they were not diagnosed at that stage but when I contacted Dr Rachele Danti to discuss reanalysis she identified their *SLC30A9* variant). After I presented these at a conference, an Australian team made contact to discuss a third possible family. This individual had a variant of uncertain significance for which I used splicing assays to confirm pathogenicity. Subsequently a fourth family was identified through diagnostic testing at Great Ormond Street – he had the same variant as one of the NMDC families, which led to it being treated as a serious candidate for pathogenicity. A manuscript describing these four families is currently

under review with Neurology.(204) Pedigrees of the two multiplex families are shown in **Figure 4.6**.

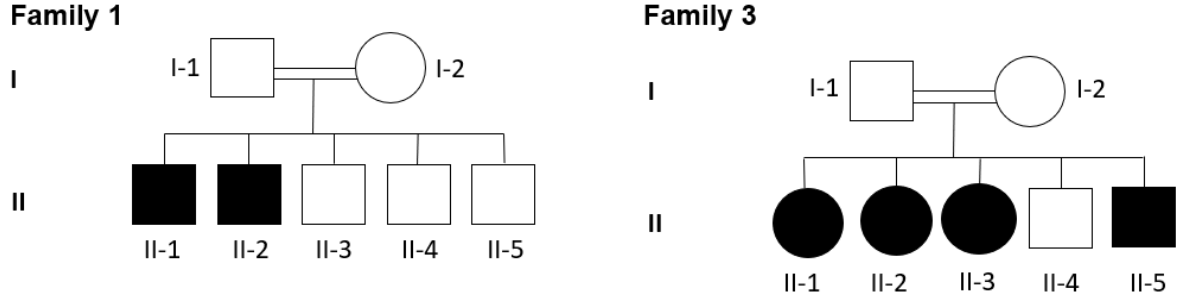
Case History: CD24 (Family 1)

CD24P and CD24S are the eldest two sons of a British family of Pakistani origin, in which the parents were first cousins. Both boys were born at term with unexplained low birth weight, <0.4th centile. Both achieved early neurodevelopmental milestones at the later end of the normal range, walking at 18 months, and both were diagnosed with moderate learning difficulties but were able to continue in mainstream education with specialised support.

CD24P, the second son, presented with a movement disorder at around seven years of age. CD24S did not develop a movement disorder until the age of 16, and it remains somewhat milder than his brother's. In both cases the movement disorder consists of continuous, irregular, low-amplitude choreiform movements and low-amplitude myoclonic jerks of all four extremities, together with some subtle perioral dyskinesia and mild dystonic posturing. Gait was effortful for both, and by 16 years old CD24P was only able to walk a few steps without assistance. He also developed a progressive thoracic kyphosis. Both also had musculoskeletal contractures: in the elder these involved the fingers bilaterally, and in the younger the elbows and the left foot. The elder had mild difficulties with saccadic eye movements and the younger developed clear oculomotor apraxia. Both had subtle bilateral ptosis. The elder also has moderate bilateral sensorineural hearing impairment.

Neither patient had any impairment of kidney function at the time of assessment, but CD24P was known to paediatric nephrology for unexplained small echogenic kidneys, and also under investigation for possible hypertension.

Figure 4.6: *Pedigrees of the two multiplex families. Circle: female; square: male; filled: affected; unfilled: unaffected. Double line indicates a consanguineous union.*



A table summarising genotypic and phenotypic characteristics of the families included in our series and those previously reported is shown below (**Table 4.1**). My work on this gene began with Family 1 (CD24), in which two brothers from a consanguineous family had a similar movement disorder, although only one had been sequenced. Once I had complied a shortlist of protein-altering homozygous variants, I was able to discount several which had no relevant physiological role and/or had a recognised disease phenotype incompatible with the presentation (see **Appendix 3**). Two homozygous variants remained: I tested the affected brother for both, and only the *SLC30A9* variant was homozygous in both siblings. Family 2 shared the same homozygous missense variant as Family 1, although no connection could be established between the two families, who originated from different regions of Pakistan.

Prof Maya Topf's team at the Institute of Structural and Molecular Biology kindly undertook *in silico* structural modelling of the missense p.Gly418Val variant and concluded that the amino acid substitution was likely to reduce conformational flexibility and affect helix packing. Family 3 (DP11) were homozygous for the previously reported pathogenic variant (c.1049delCAG, pAla350del). The child in Family 4 was compound heterozygous (in trans) for two variants: a frameshift (c.1083dup, p.Val362Cysfs*5) and synonymous variant which was close to a splice site (c.1413A>G, p.Ser471=). *In silico* tools did not predict a strong effect on splicing for the second variant so further investigation was required.

The patient's clinical team in Australia were able to share patient fibroblasts from a skin biopsy and these were cultured in Dr Derek Burke's laboratory at GOSH and subsequently by Dr Kimberley Reid, who also undertook RNA extraction and reverse transcription for this experiment. I designed primers to cDNA with the following sequences: GCATGGTCTCAGCATTCCTC (forward) and TTCTCAAGTCTATCTACTTCAGCTCCT (reverse); they were optimised and used in standard PCR conditions as described in Chapter 2 with an annealing temperature of 64°C.

Report	Perez et al						Kleyner et al	Steel et al Family 1		Steel et al Family 2	Steel et al Family 3				Steel et al Family 4
Variants	c.1049delCAG, pAla350del (homozygous)						c.40delA, p.Ser14Alafs*28 /c.86_87dupCC, p.Cys30Profs*13	c.1253G>T, p.Gly418Val (homozygous)		c.1253G>T, p.Gly418Val (homozygous)	c.1049delCAG, pAla350del (homozygous)				c.1413A>G, p.Ser471=/ c.1083dup, p.Val362Cysfs*5
Ethnicity	Bedouin						African American	British Pakistani		British Pakistani	Egyptian Palestinian				European Australian
Individual	V(9)	V(10)	V(11)	V(12)	V(13)	V(15)	Proband	F1 (II-1)	F1 (II-2)	F2	F3 (II-1)	F3 (II-2)	F3 (II-3)	F3 (II-5)	F4
Ever ambulant?	All remained ambulant though with difficulty						No	Yes	Yes	Yes	Briefly	No	No	No	Yes
Intellectual disability	All severe: limited speech						++	+	+	+	++	++	++	++	+
Dystonia/ choreoathetosis	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Age of movement disorder onset	1-2y in all						5-10y	16y	7y	1-2y	Under 2y in all				5y
Ptosis	?	+	?	+	+	+	-	+	+	+/-	+	+	+	+	+
Oculomotor apraxia	+	+	+	+	+	+	-	+	++	+	+	+	+	+	+
Renal abnormality	-	+	+/-	+	-	++	+	-	+/-	+	-	-	-	-	-
Hearing impairment	-	-	-	-	-	-	++	+	-	+	NA	NA	NA	NA	++
Low birth weight	-	-	-	-	-	-	+	+	+	+	NA	NA	NA	NA	NA
MRI abnormalities	NA	NA	-	-	+[a]	-	++[b]	NA	-	+/-[c]	NA	NA	NA	+/-[c]	+/-[c]

Table 4.1: *Genotype and phenotype of people with pathogenic* SLC30A9 *variants. +: present; ++: strongly present; -: absent; +/-: borderline; NA: information not available*

[a] *Periventricular white matter change*

[b] *Agenesis of corpus callosum; pachygyria; white matter loss; arachnoid cyst*

[c] *MRI showed borderline abnormalities such as small volume basal ganglia; minimal cerebellar volume loss; dysmorphic corpus callosum*

I ran the PCR products on a 2% agarose gel for one hour as described, alongside those from two unaffected controls. Unexpectedly, both proband *and* controls showed two main amplicons as shown in **Figure 4.7**. I therefore extracted the separated gel products, undertook a further PCR with the same primers for amplification, and then sequenced the products.

The smaller product (blue arrow) proved to be identical for both proband and controls: it consisted of exon 15 followed by exon 17 – i.e. exon 16 was skipped. However, the final 5bp of exon 15 were missing and were replaced (in both proband and control) by TCG – i.e. net loss of 2bp, which in combination with the loss of 130bp exon 16 implies an in-frame shortening of the polypeptide. The TCG codon appears several times in the 15-16 and 16-17 introns: it may represent a tiny and previously unrecognised exon. This transcript does not appear to have been documented previously and, interestingly, it implies that our proband's variant is immediately on a splice site rather than a few bases away from it.

However, the difference between proband and controls was found in the larger product (red arrow). In the controls, this included a normal exon 15-16 boundary as expected from published transcripts. In the proband, however, a double trace was seen in exon 15 until the 15-16 boundary (**Figure 4.8**). It was not possible to resolve the two products by gel electrophoresis as they were so similar in size, but by subtracting the wildtype sequence I could determine that the variant trace was missing the final 5bp of exon 15: that is, the 5bp immediately following the variant. In this transcript, this is anticipated to result in a frameshift and premature termination as follows: p.Val472Glyfs*4. This is shown schematically in **Figure 4.9**: the loss of the final 5bp of exon 15 (from the variant onwards) means that splicing to exon 16 is not possible in frame. The reduced brightness of proband's larger cDNA band could, speculatively, reflect partial nonsense-mediated decay of this transcript but confirming this would require quantitative testing.

Figure 4.7 *Gel electrophoresis of* SLC30A9 *cDNA from the proband (left) and an unaffected control (right). Red arrow: larger amplicon (which on sequencing proved to contain two different sequences in the proband); blue arrow: smaller amplicon (transcript lacking exon 16, found in both proband and normal control). Standardised 100bp "ladder" is shown for measurement. Second control not shown for simplicity.*

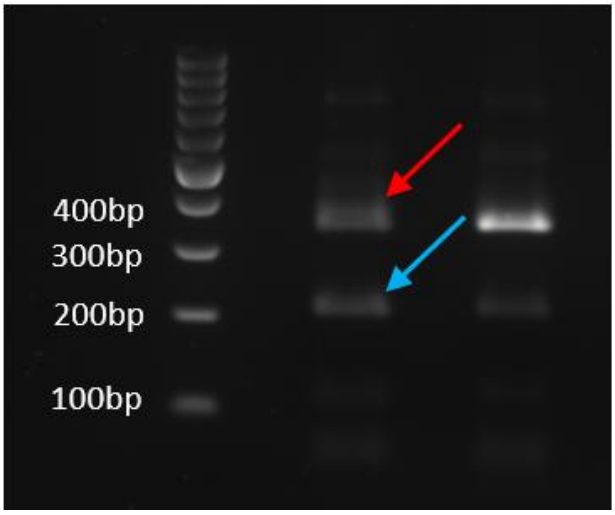


Figure 4.8 *cDNA sequences around the exon 15-16 boundary from the proband and control. Note the double trace starting at the exon-exon boundary. (NB The appearance of a single peak immediately to the left of the boundary is because the same base is found in the variant and wildtype transcripts).*

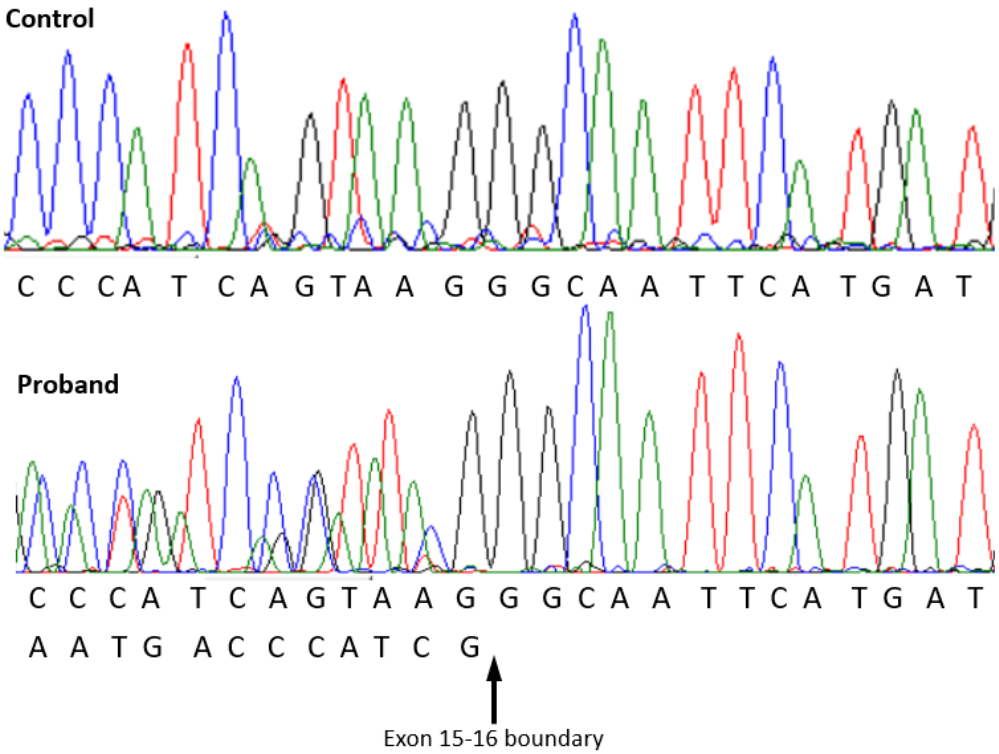
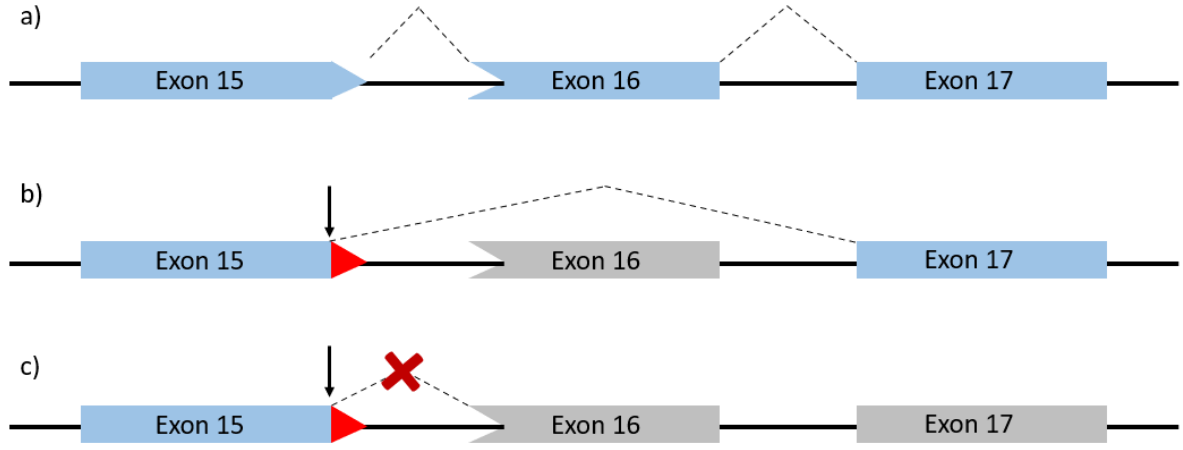


Figure 4.9 *Schematic representation splicing of* SLC30A9 *in a) the normal "main"/larger splicing product found in both patient and controls; b) the normal smaller product skipping exon 16, together with the final 5bp of exon 15 (marked in red, found in both patient and controls and c) the aberrant skipping of the final 5bp of exon 15 then proceeding into exon 16, resulting in a frameshift and premature termination (patient specific). Dotted lines indicate splicing.*



These four new families (eight affected individuals) enhance our knowledge of the phenotypic spectrum of this newly described neurogenetic disorder. All the participants in our series, and all those described in Perez et al's index report,(201) share the distinctive features of a progressive dyskinetic/dystonic movement disorder, intellectual disability, oculomotor apraxia and ptosis. The single case report by Kleyner et al is not known to have the ocular findings, but does have the movement disorder.(205) Two individuals from our cohort are under investigation for borderline abnormal renal findings but none (so far) have definite chronic kidney disease, hypertension or hyperkalaemia. Sensorineural hearing loss was not described in any of Perez et al's kindred, but was identified in Kleyner et al's proband and in at least three individuals, from different families, in our cohort, suggesting it is likely to form part of the phenotype rather than a coincidental finding. (Our Family 3, in the Gaza Strip, have not had a detailed audiological assessment.)

SLC30A9 encodes ZnT-9, a ubiquitously expressed transmembrane protein which is known to play a role as a zinc transporter. Perez et al, in the same paper in which they report on the disease, offer immunofluorescence-based evidence that the protein colocalises with cytosolic vesicles and especially the endoplasmic reticulum. In vitro, cells expressing the in-frame deletion variant have been found to have lower levels of cytosolic zinc but disturbances of zinc metabolism in vivo have not yet been

demonstrated, and notably, in those of our patients where testing was possible serum zinc levels were normal. Interestingly, people with BLPS do not manifest many of the classical signs of systemic zinc deficiency, such as dermatitis and diarrhoea.(23) Although this could reflect that fact that a transporter defect would only deplete intracellular, rather than total body, zinc, it is also possible that ZnT-9 has more physiological functions than have yet been identified. ZnT-9 is designated as a zinc transporter partly due to homology with other related proteins, but the possibility that it also plays a role in the transport of other metal ions, as is the case for other SLC30 transporters, has not yet been fully explored.(24)

Protein-truncating variants in *SLC30A9*, including the splicing variant in our Family 3, are likely to mediate a loss-of-function effect through nonsense mediated decay. Furthermore, both the initially reported in-frame deletion(201) and the novel missense variant reported here are also likely to exert a loss-of-function effect as they are predicted to affect the cation efflux domain of the protein and will potentially disrupt the structure of its transmembrane helices. We hypothesize that this would impair the regulation of zinc transport (and any other ions for which the protein may act as a transporter), leading to disturbances of metal ion homeostasis at the cellular or subcellular level.

To date, given the relative scarcity of reported cases, it is not possible to extrapolate strong genotype-phenotype correlations. Our Families 1 and 2, harbouring a missense variant, have a relatively milder phenotype and better neurodevelopmental outcome than those reported with either the in-frame deletion or the reported individual with biallelic protein-truncating variants, but on the other hand so does our Proband 4, who in effect also has biallelic truncating variants. Family 3, with an in-frame deletion which might be expected to be less damaging, all have severe intellectual disability and lost ambulation at an early age. One possibility, given the evidence from my splicing experiments that not all normal transcripts of *SLC30A9* are documented, is that this variant has an unpredicted effect on splicing. Identification of further cases of BLPS will help determine whether specific genetic variants are predictors of phenotypic severity.

BLPS is a newly recognised cause of complex early-onset hyperkinetic movement disorders. Clinicians should consider testing for *SLC30A9* variants in children with undiagnosed progressive dystonic or dyskinetic movement disorders, especially when accompanied by learning disability and oculomotor abnormalities, and particularly in families where a recessive mode of inheritance is suspected. For the time being, there is nothing to offer beyond symptomatic treatment of the movement disorder, together with regular surveillance of renal function, blood pressure and hearing. However, as the cohort of identified patients grows and the phenotypic spectrum is more fully described, it may be possible for future research to elucidate the underlying mechanisms by which dysfunction of the ZnT-9 channel brings about the manifestations of BLPS, and how these might be targeted by novel therapies.

4.5 New movement disorder presentations of other disorders: *GNAO1, YY1* and *STXBP1*

4.5.1 *GNAO1*

GNAO1 encodes the $G_o\alpha$ heterotrimeric G-protein subunit, which is abundant in the central nervous system but whose function is incompletely understood – the "O" stands, not very helpfully, for "other".(206) In 2013, *de novo GNAO1* variants were reported in a severe infant-onset developmental and epileptic encephalopathy, now designated DEE17.(207) Quite rapidly it became apparent that DEE was not the only, or even the most distinctive, phenotype associated with *GNAO1* variants: in 2016 several different teams published descriptions of children with severe neurodevelopmental impairment and a hyperkinetic movement disorder, vulnerable to life-threatening exacerbations.(208-211) OMIM considers this to be a separate condition, "neurodevelopmental disorder with involuntary movements" (NDIM), but in fact epilepsy and a paroxysmal movement disorder often co-exist.(212)

The classic DEE17 and NDIM phenotypes are both devastating disorders, both because of the impact of the epilepsy and/or paroxysmal hyperkinesia themselves, and because they are usually associated with severe to profound developmental and intellectual impairment. There is, however, a milder end to the *GNAO1* phenotypic spectrum which had not previously been fully recognised.

As part of an international collaboration, we compiled a case series of 24 individuals from 20 families who had pathogenic *GNAO1* variants and a relatively mild phenotype in which dystonia was the predominant symptom.(213) The majority (21/24) did not have epilepsy; half had no intellectual disability (the other 12 had mild to moderate impairment); and only 3/24 experienced acute exacerbations of their movement disorder. The mean age of symptom onset was 6.6 years, in contrast to the classical phenotypes, which present in the first year of life, and one participant experienced their first symptoms at the age of 47. The series included four families with more than one affected individual, and one family with an asymptomatic (apparently non-mosaic) carrier, in contrast to previous reports of *GNAO1* disease in which all variants have either occurred *de novo* or been inherited from a healthy mosaic parent.

Case History: DY120

DY120P is an Israeli girl with no relevant family history. Her parents became concerned that her development was delayed when she was two years old: she could walk, but unsteadily, and had not started talking. Her intellectual disability was mild but she had significant dysarthria which limited her ability to acquire speech, associated with some drooling.

When seen by a neurologist at the age of five years, she had mild generalised dystonia and poor balance. Deep tendon reflexes were brisk. Her dystonia was slowly progressive but she remained independently ambulant at 13 years old. MRI and EEG were normal and she never had seizures. She was noted to have an exaggerated acoustic startle response and might fall over in response to sudden noises.

Intriguingly, one family (consisting of an affected mother and daughter) had a nonsense variant, NM_020988.3:c.765dupT, p.Asn256*, and another proband had a 16q12.2 deletion encompassing the gene. The great majority of previously reported pathogenic *GNAO1* variations are missense changes: ClinVar does not include any pathogenic or likely pathogenic frameshift or nonsense variants at all. This suggests a possible genotype-phenotype correlation although it is not clear whether a dominant negative mechanism or haploinsufficiency is at work. It has previously been suggested that loss-of-function variants in *GNAO1* cause epilepsy while gain-of-function variants cause movement disorders,(214) but this appears likely to be an oversimplification.

Three of the families were identified through our movement disorders project, two of these (DY120 and DY72) through my analysis. Both of these had the variant c.724-8, p.?, a non-canonical splicing variant which had previously been reported as pathogenic and demonstrated to cause an in-frame insertion resulting in abnormal protein trafficking.(215) One additional family presented to the NMDC already diagnosed, and we recruited her to this study and shared data on her phenotype and genotype.

4.5.2 *H3.3*

Histone proteins are composed of a complex cluster of subunits which regulate DNA transcription in complex and incompletely understood way. H3.3 Histone gene nomenclature is exceptionally confusing, but *H3-3A* (alias *H3F3, H3F3*A or *H3.3A*) and *H3-3B* (alias *H3F3B* or *H3.3B*) encode H3.3 histones A and B. H3.3 is one of the principal forms of Histone H3. Somatic H3.3 variants have a well-documented role in cancer pathogenesis, especially glioblastoma.(216)

In late 2020 a paper was published describing a polymorphic neurodevelopmental syndrome involving variable congenital anomalies and in some cases spasticity and/or epilepsy, related to germline variants in *H3-3A* and *H3-3B*.(217) The authors described it as a "neurodegenerative" disorder but only a minority of patients showed developmental regression and OMIM has designated the condition "Bryant-Li-Bhoj neurodevelopmental syndrome 1".

Serendipitously, this coincided with me finally managing to obtain VCF files for the triome of a child seen in the NMDC (DY197P). Sequencing and initial analysis had been undertaken by a private company but I reanalysed the data at the family's request. This identified a novel *de novo* heterozygous variant in H3-3B, c.365C>G, p.Pro122Arg, with a CADD score of 32 (see **Appendix 3**). (NB I am using standard genetic position numbering here: histone biologists count the initial methionine as zero, so their p. number is always one less than expected.) I listed the variant on GeneMatcher (as the phenotype was not yet listed on standard databases) and very rapidly heard back from Dr Elizabeth Bhoj's team with details of their cohort, which included a patient with the same variant.

Case History: DY197

DY197P was born at term and in good condition but presented with seizures on the second day of life. These were brought under control with levetiracetam but by three months of age it was clear that he had global developmental delay. He continued to have infrequent tonic-clonic seizures also developed mild fluctuating generalised dystonia.

Examined at the age of 23 months, he could sit but not crawl or pull to stand, and he smiled responsively but did not use or understand words. Truncal tone was low and reflexes were normal. There were subtle dysmorphic features including mild bilateral ptosis, tented mouth, prominent forehead, and a low anterior hairline. MRI showed mild non-specific abnormalities including delayed myelination.

Case History: DY47

DY47P was adopted at the age of 16 months and therefore recruited as a singleton: full details of his birth family history are not known. He presented with gait difficulties from shortly after the onset of walking and was found to have slowly progressive bilateral lower limb spasticity. Over time he developed equinovarus deformities of both feet, successfully treated with plantar fascia release. He also had mild, non-progressive intellectual disability, for which he attended a specialist unit within a mainstream school, and had difficulties with fine motor tasks. MRI showed no abnormalities other than (on spinal imaging) slight prominence of the central canal.

Subsequently, we identified a novel heterozygous *H3-3A* variant (c.73G>A, p.Ala25Thr) in a teenage boy (DY47P) analysed as a singleton (see **Appendix 3**). This variant was in the histone tail and strongly suspected to be pathogenic although inheritance could not be confirmed. Both individuals were reported in a paper describing the spectrum of subtle malformative or dysmorphic features seen on brain imaging.(218)

4.5.3 *YY1*

YY1 encodes yin and yang 1, a zinc finger transcription factor. In 2017, heterozygous variants were reported to cause a syndrome of intellectual disability, feeding difficulties, subtle dysmorphic features and variable congenital anomalies in a cohort

of 23 patients.(219) The condition was named Gabriele-de Vries syndrome. Movement symptoms including gait abnormalities, tremor and dystonia were mentioned as occurring in several individuals but details were not provided.

Case History: CD39

CD39P presented in infancy with mild hypotonia, feeding difficulties, and repeated respiratory infections. She walked at 11 months but at 30 months developmental delay was noted. She was able to attend mainstream school with support but was found to have mild to moderate intellectual impairment and was diagnosed with attention deficit hyperactivity disorder.

At five years old she developed an upper-limb intention tremor, more marked on the left. Over the next five years this became more severe and generalised, and by 10 years old she had both resting and action tremor, myoclonic jerks involving the left arm, and ataxia. Due to these symptoms she lost the ability to write and to feed herself independently. There was no cognitive regression and she remained ambulant. Examination (by another team) at 21 years old showed both generalised dystonia and ataxia, with hypometric and delayed saccades.

A participant recruited to our study moved country, and thus ended up being investigated both by ourselves and by a team in Geneva led by Dr Julien Bally. Independently, we identified a novel *de novo* heterozygous *YY1* variant as follows: NM_003403.4:c.907T>C, p.Cys303Arg. The variant was unanimously predicted to affect protein function by *in silico* tools and was classified as pathogenic (Class 5) by the ACMG criteria.(220) She had a complex, progressive hyperkinetic disorder which was a much more significant cause of disability than her mild ID.

Several additional individuals with Gabriele-de Vries syndrome where dystonia is the most severe and prominent symptom have since been reported, including instances of an encouraging, though incomplete, response to DBS.(221-223) It has been noted that the condition has some phenotypic overlap with *KMT2B*-related dystonia 28, and dysfunction of *YY1* may also form part of the downstream pathophysiological mechanism of *THAP1* variants, in dystonia 6.(224)

4.5.4 *STXBP1*

Syntaxins are SNARE (“SNAp REceptor”) proteins involved in vesicle exocytosis and *STXBP1* encodes a syntaxin binding protein involved in the regulation of neurotransmitter release.(225) Heterozygous variants were linked to a developmental and epileptic encephalopathy in 2008.(226)

Like many other genes where the first described pathological association was a DEE – possibly because these conditions have a longer history of detailed genetic investigation – a wider phenotypic spectrum has since been revealed, including neurodevelopmental impairment apparently independent of the severity of epilepsy,(227) and subcortical myoclonus.(228) In a recent description of over 500 individuals, epilepsy prevalence was only 89%.(229)

Case History: TR5

Child TR5P presented with tremulous movements at six weeks old, followed by delayed motor development, balance difficulties, mild upper limb dystonia and action-associated upper limb tremor. He had no history of seizures and no epileptiform abnormalities on EEG. ID was mild to moderate and he was able to attend mainstream school with support. His birth father and paternal grandmother both had dystonia but were not available for examination or sampling. Neither had any known history of epilepsy.

I analysed TR5P as a singleton because he was not living with his birth parents and identified a heterozygous stop-loss variant in *STXBP1*: c.1783 T > C, p.*595Glnext*67. Stop-loss variants are an uncommon but recognised causes of disease and can act either through loss of function or dominant negative effects.(230, 231) No other variants were strong candidates for pathogenicity (see **Appendix 3**). The fact that the characteristics of the movement disorder in our patient are similar to those described in wider *STXBP1* cohorts increase our suspicion that it is likely to be pathogenic; we speculate that it may have been inherited from his affected father. We cannot definitively classify this variant as pathogenic in the absence of additional evidence, however. TR5P would clearly represent the mildest end of the described *STXBP1* spectrum: probably due to ascertainment bias, milder phenotypes are often later to be

described and published. A report on this case was written and published by Dr Robert Spaull, where I am co-author for undertaking the genetic analysis.(232)

4.6 Discussion

JPH3, RHOBTB2, SLC30A9, H3.3, GNAO1, YY1 and *STXBP1* are not "new genes" in the sense that all were already known to be associated with neurological disorders. The index report of a condition, however, is not the end of the story.

In some instances, as with *JPH3,* different classes of variant in the same gene can result in conditions which are entirely distinct, both in clinical and in pathophysiological terms. For genes, such as *RHOBTB2* or *YY1,* certain aspects of the phenotype have been well described but others – perhaps especially movement disorders, which are sometimes harder to categorise and quantify than, for example, epileptic seizures or ID – have been relatively neglected and their impact on patients' lives under-recognised. Others again, like *SLC30A9,* are simply so rare, and have been described in so few families, that corroborative reports can add substantially to what is known about the disorder.

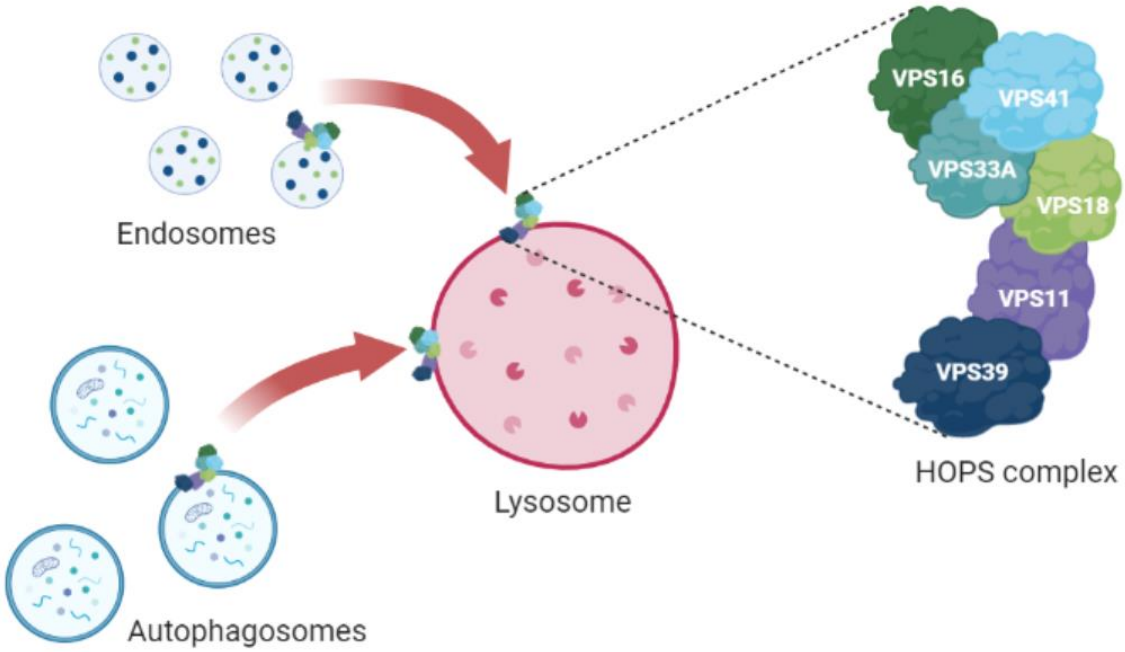
The benefits of understanding the phenotypic spectrum of a rare condition are manifold. Firstly, it improves prognostication of the future course of a disease, which is among families' most urgent questions: with *JPH3*, for example, the distinction between the apparently static loss-of-function condition and the progressive triplet-repeat expansion one is vitally important. Secondly, it can guide care and screening for predictable complications: with *SLC30A9,* now that it appears that sensorineural hearing loss is a common feature, clinicians will be aware that audiology monitoring should be considered. Thirdly, it allows diagnoses to be made across the full spectrum: for example, non-epileptic children with alternating hemiplegia-like phenotypes can only be diagnosed with *RHOBTB2* disorder now that this is recognised as one of its manifestations. Finally, detailed understanding of the range and variety of clinical courses seen in a genetic disorder is an important step towards understanding the pathophysiology, and thereby ultimately being able to modify it.

Chapter 5. HOPS complex variants: two novel causes of movement disorders

5.1 Introduction

The HOPS (HOmotypic fusion and Protein Sorting) complex is a protein complex with very high evolutionary conservation, found as far back in the phylogenetic tree as baker's yeast, with ubiquitous tissue expression.(233) It plays a key role in tethering vesicles to lysosomes, allowing them to fuse and deliver their cargo, and is required for the fusion of both late endosomes and autophagosomes.(234, 235) **(Figure 5.1.)** The HOPS complex achieves its role through a number of different interactions with SNARE proteins, particularly catalysis of SNARE complex formation(236) and protecting the SNARE complex against disassembly after formation.(237)

Figure 5.1: *Schematic showing the structure of the HOPS complex and its role in trafficking from endosomes and autophagosomes into lysosomes*



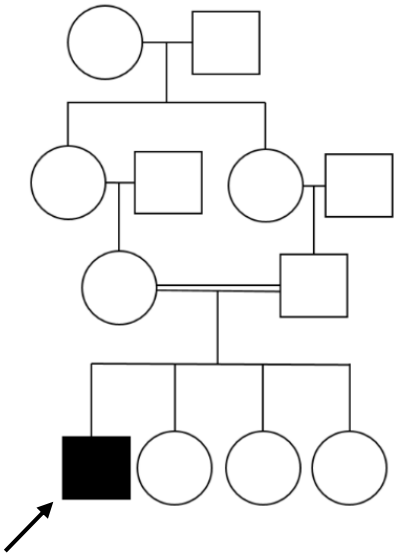
The mammalian HOPS complex consists of six proteins: VPS11, VPS16, VPS18, VPS33A, VPS39 and VPS41.(238) (VPS stands for “Vacuolar Protein Sorting.) Of these, autosomal recessive disease phenotypes had been described for *VPS11* (hypomyelinating leukodystrophy 12, MIM 616683)(239) and *VPS33A* (mucopolysaccharidosis-plus syndrome, MIM 61703).(240) A recessive mutation of *VPS16* had also been reported in a single family with dystonia.(241)

Through WGS of undiagnosed patients with movement disorders, we identified two new HOPS-associated movement disorders: an AR disorder of *VPS41* which has now been designated autosomal recessive spinocerebellar ataxia 29, and an AD disorder of *VPS16,* now known as dystonia 30. Work relevant to this chapter was published in the journal Annals of Neurology in 2020 and some elements including tables and figures are included.(77)

5.2 *VPS41*: the proband and family

We were referred a child with an undiagnosed dystonic disorder, pseudonymised as DY173P, for triome WGS. Clinical information was supplied in the form of a referral note, a clinical proforma, videos, and reproduced images from a hard-copy MRI brain scan.

Figure 5.2: *Pedigree for DY173P (proband indicated with arrow). Parents are consanguineous (first cousins).*



Case history: DY173

DY173P was born of first-cousin consanguineous parents ***(Figure 6.2)*** *and had one first cousin who was diagnosed with "cerebral palsy" but was not investigated. Antenatal and birth history was uneventful: the baby was born at home and did not require resuscitation. There was no history of neonatal encephalopathy or jaundice.*

From infancy onwards he manifested significant global developmental delay: aged five, he could sit with support, had voluntary limb movements including reaching and grasping, and was able to hold a crayon to scribble on paper. He learned to crawl at the age of four years but never walked. (Although ataxia was not reported on his neurological examination, the ability to crawl in the absence of the ability to sit stably is suggestive of this finding.) Starting at the age of three years he acquired a few words of spoken language. From six years old, however, his condition began to deteriorate, with decreased alertness and less ability to sit even with support.

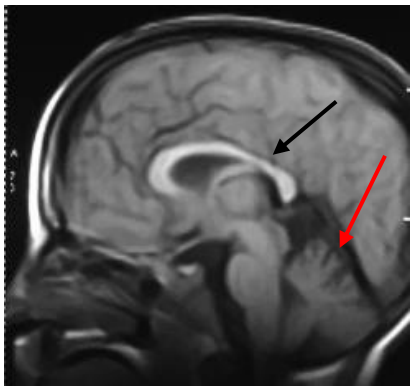
In the first year of life, he was noted to have generalised dystonia with intermittent abnormal posturing of all four limbs. (Video was provided but will not be included in this thesis as we do not currently hold written permission for publication.)

On examination aged five years he had some soft dysmorphic features including left-sided strabismus, a high forehead and slightly low-set ears. Head circumference was two standard deviations below the mean. Optic discs were noticeably pale. There were no signs of systemic disorder, specifically no organomegaly. Limb tone was variably increased with dystonic posturing noted, and deep tendon reflexes were present. On re-examination at six years old, however, tendon reflexes were decreased and at this stage nerve conduction studies revealed a generalised sensorimotor axonal neuropathy.

Lactate and tandem mass spectrometry screening for inborn errors of metabolism were normal. MRI brain scan was reported locally as normal, but review by a paediatric neuroradiologist (Dr Kshitij Mankad) at GOSH identified atrophy of the superior aspect of the cerebellar vermis and thinning of the posterior limb of the corpus callosum ***(Figure 6.3).***

A trial of treatment with L-dopa was ineffective.

Figure 5.3: *T2-weighted midline sagittal MRI brain image for DY173P. Black arrow indicates thinning of the posterior limb of the corpus callosum. Red arrow indicates atrophy of superior part of cerebellar vermis.*



5.3 Genetic analysis

DNA from DY173P and both parents was available. Sequencing and analysis took place as described in Chapter 2: panel-free analysis was undertaken for biallelic and *de novo* variants and no clearly pathogenic variant in any relevant disease-causing gene was found. One variant was identified in *ATP7A*, an X-linked gene with variants reported in Menke's disease(242) and in distal spinal muscular atrophy(243), but *in silico* predictors of pathogenicity were ambiguous and in any case neither condition was a close phenotypic fit. Inherited dominant variants were not analysed in this case because of (i) the severity of the phenotype, which makes inheritance from a wholly asymptomatic parent less likely, though not impossible and (ii) the history of consanguinity.

After analysis and variant prioritisation as described in Chapter 2, a shortlist of four homozygous variants which had no known disease phenotype but could plausibly be relevant was compiled **(Table 5.1).** Of these, the variant in *VPS41* (c.450+1G>T, p.?) was the preferred candidate because it is a canonical splicing variant, and therefore likely to lead to significant disruption of protein function.

Gene	cDNA change	Protein change	Genomic coordinate (GRCh38)	CADD	SIFT	Mutation Taster	PolyPhen-2	Align GVGD	Provean	Splicing predictors	gnomAD frequency	Function	Expression	Gene-Matcher match?
VPS41	NM_014396.3: c.450+1G>T	p.?	Chr7: 38817816	28.2	NA	NA	NA	NA	NA	MaxEnt: -100.0% NNSPLICE: -100.0% SSF: -100.0%	Not found	HOPS complex component	Ubiquitous	Y
SCOC	NM_ 001153484.1: c.234G>A	p.Met78Ile	Chr4: 140379173	28.3	Tolerated (0.11)	Disease causing (0.821)	HD: Prob Dam (0.993) HV: Prob Dam (0.968)	C0	NA	NA	0.0008%	Regulates starvation-induced autophagy	Widespread but highest expression in brain	N
PLCB2	NM_004573.2: c.1543del	p.Glu515 Argfs*26	Chr15: 40296578	NA	NA	Disease causing (1)	NA	NA	Deleterious -1193.151	NA	Not found	Mediates IP3 second-messenger signalling	Widespread but highest in platelets	N
FOXB1	NM_012182.: c.899C>A	p.Thr300Asn	Chr15: 60005862	25.3	Tolerated (0.06)	Disease causing (1)	HD: Poss Dam (0.947) HV: Prob Dam (0.950)	C0	NA	NA	0.0038%	Transcription factor	Highest in substantia nigra	N

Table 5.1: *Genomic and predicted characteristics of shortlisted homozygous variants identified in DY173P.*
Colour coding: Red: predictor supports pathogenicity; Yellow: predictor equivocal; Green: predictor opposes pathogenicity.

VPS41 appeared an intriguing candidate for neurological disease because previous work in invertebrate models of neurodegenerative disorders, especially those involving toxic accumulation of abnormal proteins, suggested that the healthy protein might play a neuroprotective role. *Caenorhabditis elegans* models exist for both Parkinson's and Alzheimer's diseases. For Parkinson's, indeed, there are two alternative models, one relying on overexpression of α-synuclein and the other using exogenous administration of the neurotoxin 6-hydroxydopamine (6-OHDA).(244) The Alzheimer's model relies on overexpression of β-amyloid: dementia in a nematode being difficult to demonstrate, its pathogenic effect is quantified either by the effect of β-amyloid expression on the worm's body wall muscles, where it induces paralysis, or by direct examination of the glutamatergic neurons, of which a healthy *C. elegans* has exactly five.(245)

In the Parkinson's disease worm, expression of the human form of *VPS41* led to reduced loss of dopaminergic neurons in both the α-synuclein and the 6-OHDA models. The same researchers also went on to demonstrate that *VPS41* overexpression in human cell-lines reduced the formation of insoluble alpha-synuclein in response to treatment with rotenone.(246)

For the Alzheimer's worm, researchers demonstrated that overexpression of *VPS41* ameliorated and delayed the paralysis phenotype *in vivo*, while depletion of endogenous *vps41* had the opposite effect. Worms overexpressing *VPS41* also underwent less degeneration of glutamatergic neurons. Interestingly, overexpression of *VPS41* was no longer able to rescue neurotoxicity in worms where a key mediator of autophagy, *lgg-2*, had been depleted by siRNA. This supports the hypothesis that autophagy forms an essential part of how *VPS41* accomplishes its neuroprotective role.(247)

Two of the other candidate genes in DY173P (*SCOC* and *FOXB1*) had physiological functions and expression patterns which could be consistent with a contribution to the proband's phenotype. However, their *in silico* predictors of pathogenicity were less convincing: in each case, at least one of the key tools predicted the variant to be tolerated. Unlike the *VPS41* variant, they also occurred at low frequency in population

databases, although they were not reported in homozygosity. The third candidate, *PLCB3*, was a frameshift variant with stronger predictors of pathogenicity but the gene product was known to be expressed primarily in platelets, making it a less probable candidate for a neurological phenotype with no known haematological features.

Nevertheless, all four candidate genes were listed on GeneMatcher.(124) No relevant matches were made for any of the other genes except *VPS41*. Through GeneMatcher we became aware of four other affected families with significant phenotypic overlap. All probands shared features of a movement disorder, neurodevelopmental impairment and (in most cases) retinal abnormalities, and all had homozygous variants in *VPS41*. In all the other cases these were missense variants. Although collaboration with the groups investigating these families was initiated it did not subsequently prove possible to pursue this to fruition, and therefore this thesis will not include a detailed discussion of these additional cases. They have subsequently been published in two separate reports.(248, 249) However, the knowledge that such families existed was enough to persuade us to pursue *VPS41* as the most probable culprit in DY173P.

Sanger sequencing confirmed that DY173P was homozygous for the variant, while both parents, as expected, were heterozygous **(Figure 5.4).**

In silico predictors agreed that the variant was likely to cause a complete loss of the splice site resulting in skipping of exon 7, so an experiment was undertaken to confirm this. Messenger RNA (mRNA) extraction and reverse transcription was undertaken in collaboration with Dimitri Budinger and the splicing experiment itself with Dr Katy Barwick. Culture of fibroblasts was undertaken by Dr Derek Burke and his team at the GOSH Enzyme Laboratory. The techniques are described in detail in Chapter 2.

PCR was carried out as described in Chapter 2 using sequence-specific primers (**Table 2.7**). The product was run on a gel, showing that, as expected, a product was formed using patient-derived cDNA but it was detectably smaller than the equivalent wild-type product, which used cDNA from a healthy control **(Figure 5.5).** The size difference was, as predicted, consistent with the loss of exon 7 from the variant transcript, although this cannot be accurately quantified on a gel of this sort. The

product was extracted from the gel and sequenced, revealing absence of exon 7 in the patient-derived cDNA **(Figure 5.6).**

Figure 5.4 *Sequencing chromatograms for DY173P and both parents. The affected base is highlighted in blue.*

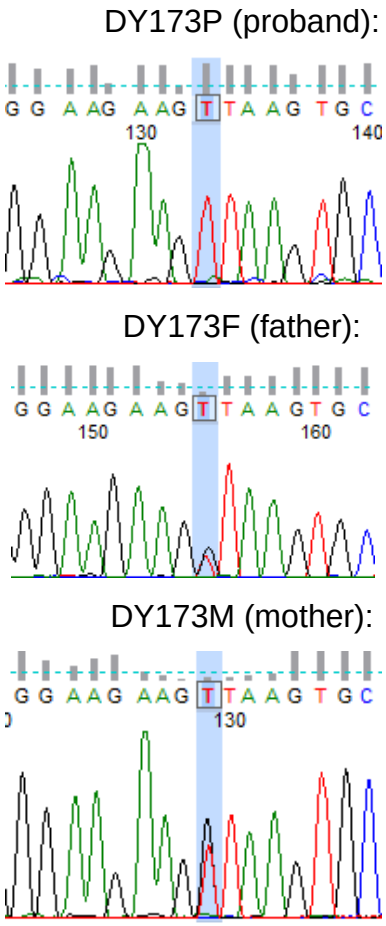


Figure 5.5: *Agarose gel following electrophoresis of patient and control cDNA for VPS41: note that the patient product is smaller by approximately 50-100bp, as predicted.*

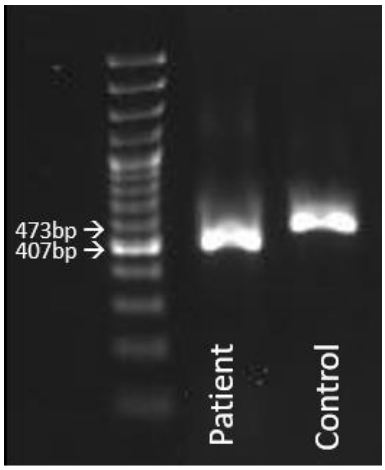
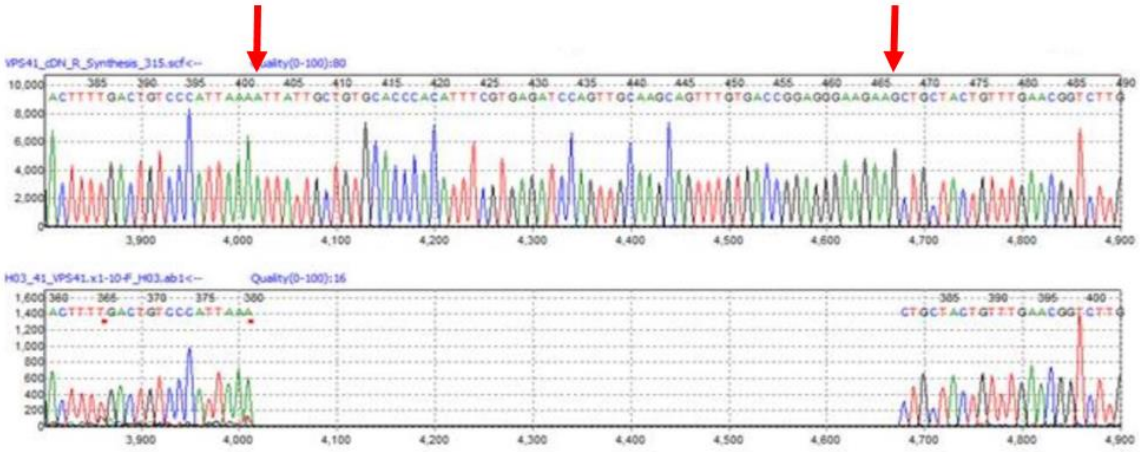
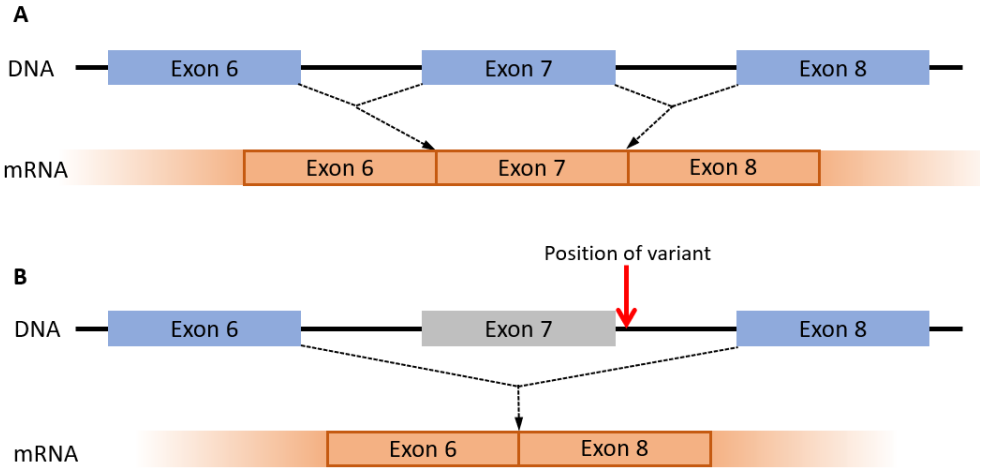


Figure 5.6: *Above: chromatograms showing the reference sequence for VPS41 (above) and skipping of exon 7 in DYP137P (below). Red arrows indicate first and last bases of exon 7. Below: schematic showing splicing in wildtype (A) and variant (B). Upper panel prepared by Dr Katy Barwick.*

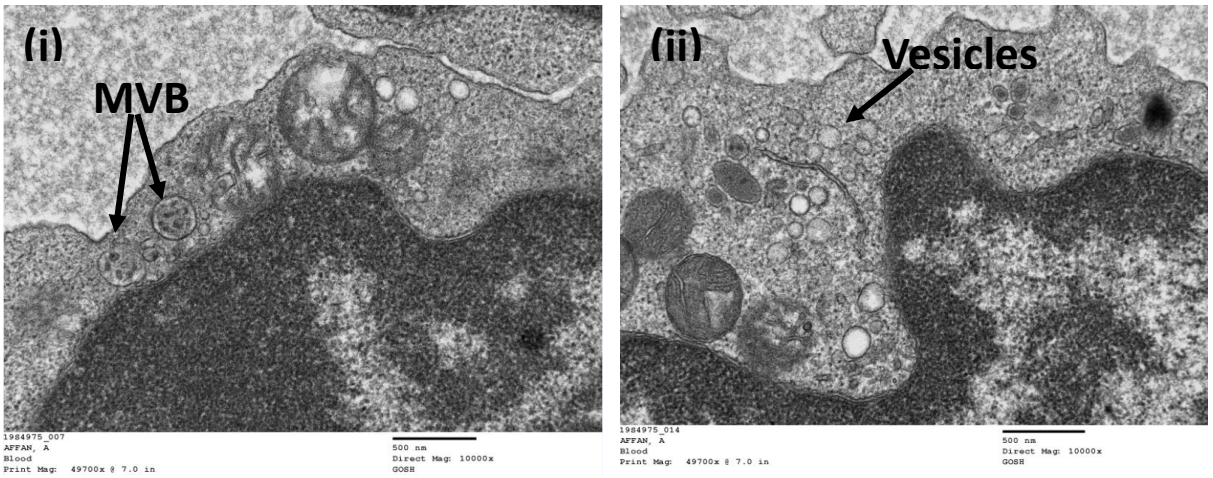




5.4 Electron microscopy findings

Both lymphocytes and fibroblasts were obtained from DY137P for microscopy by Mr Glenn Anderson. Compared with an age-matched control, significant abnormalities were seen in both cell types. Specifically, numerous membrane-bound vacuoles were seen, some of which contained granular material or fine strands of electron-dense laminar material, especially in the fibroblasts. Large numbers of small intracellular vesicles and smaller membrane-abutting vesicles were also noted. These were suspected to be pinocytic vesicles, but uptake marker studies were not undertaken to confirm this. A few multivesicular bodies were also seen. **(Figure 5.7, Figure 5.8)**

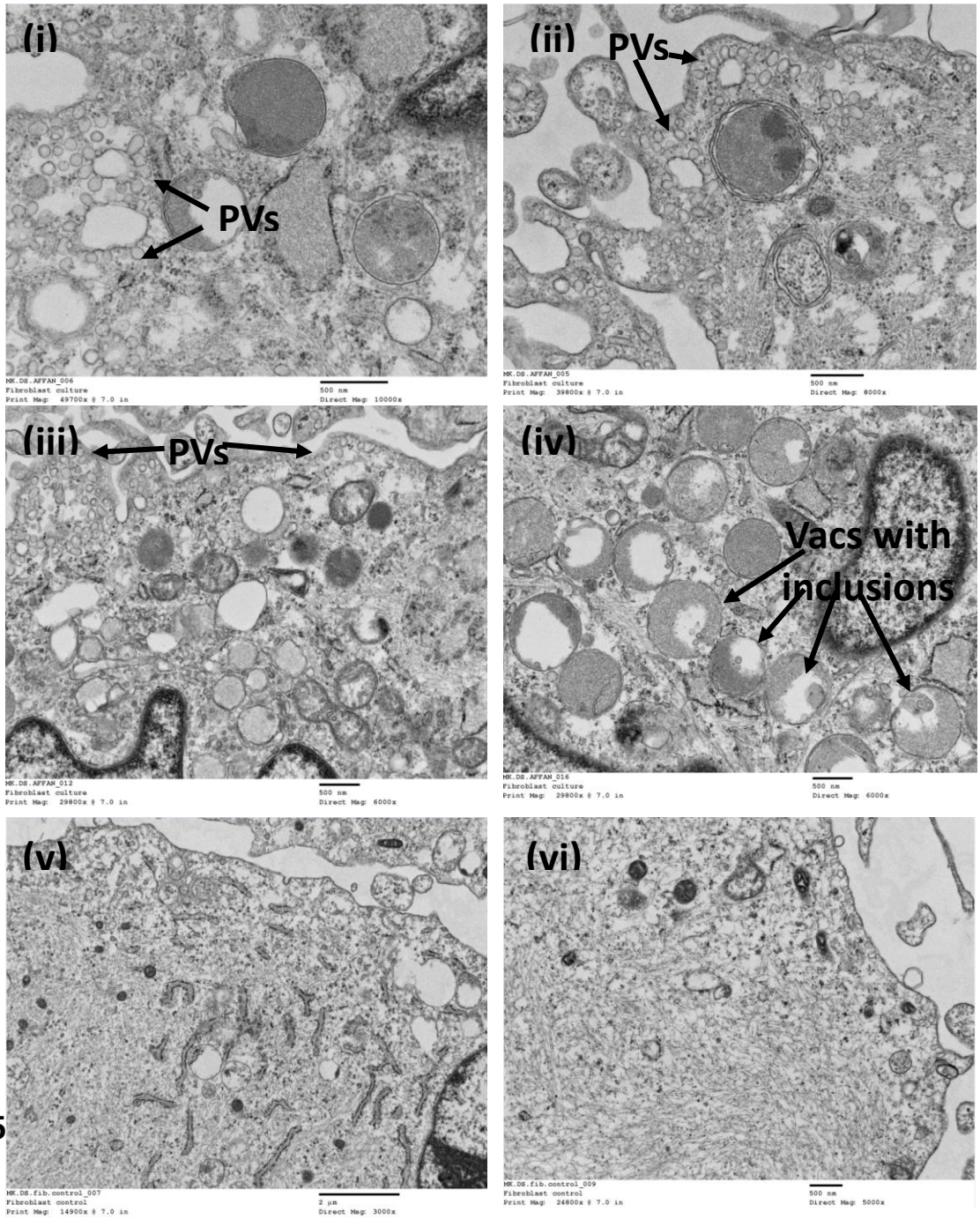
Figure 5.7: *EM images of patient-derived lymphocytes. MVB: multivesicular bodies.*



These findings were supportive of a pathogenic role for the *VPS41* variant because they were in keeping with previously reported results from both human cell lines and animal models. Firstly, Radisky et al demonstrated as long ago as 1997 that mutations in *VPS41* cause failures of vacuolar protein sorting in yeast and that, notably, this is accompanied by accumulation of multiple small membrane-bound compartments within the cell, much like those seen here.(250) The authors speculated that these compartments were “fragmented vacuoles” but it is equally plausible that they were in fact endosomes which had failed to fuse with and deliver their contents to the lysosome-like vacuoles. (Radisky et al were writing before the HOPS complex had been characterised so would not have interpreted their findings in this light.)

Secondly, *in vitro* work on human cell lines has shown that depleting *VPS41* using small interfering RNA (siRNA) results in impaired endosome-lysosome fusion. Lysosome numbers were reduced – although individual lysosomes remained apparently structurally and chemically normal – while late endosomes accumulated, their numbers almost doubled when compared with a control. Moreover, degradation of endocytic cargo was inhibited, as demonstrated by reduced fluorescence in cells treated with an endocytic probe which fluoresces on proteolytic cleavage.(251)

Figure 5.8: *EM images of fibroblasts. (i)-(iv) show patient-derived cells; (v)-(vi) are from an approximately age-matched normal control. Abbreviations: PV: pinocytic vesicle; Vac: vacuole.*

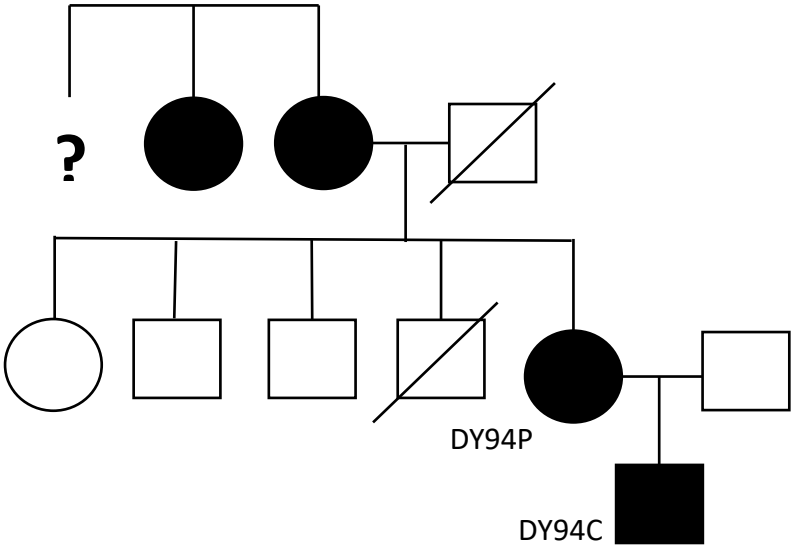


5

A collaborator referred a family, pseudonymised DY94, who were affected by dystonia with a pattern of inheritance strongly suggesting an AD variant **(Figure 5.9).** The family had some reluctance to engage with medical services so DNA was only available for

analysis from two members (DY94P and DY94C), but this was enough to conduct an analysis of shared variants.

Figure 5.9: *Pedigree for Family DY94.*



Case history: DY94

DY94P is one of five siblings. Her own mother had onset of progressive dystonia from 11 years old and used a wheelchair full-time by the age of 30. She also had some degree of both intellectual disability and psychiatric illness and made several suicide attempts. Her sister (DY94P's maternal aunt) also used a wheelchair, and her brother's son (DY94P's first cousin) had a motor impairment but full details were not available for these relatives.

DY94P herself had a mild intellectual disability and left school at 13 years old. As an adult she also experienced psychiatric and behavioural difficulties including difficulties with impulse control and a possible suicide attempt.

From 19 years old she developed dystonia, initially involving the right lower limb and spreading to involve the trunk within six months. Gradual generalisation ensued: the left lower limb was involved within a year, and she used a wheelchair from 20 years old. The upper limbs and neck became significantly involved in her 30s. MRI brain scan aged 34 showed significant symmetrical T2 hypointensity of the globi pallidi, the midbrain nuclei and the dentate nuclei of the cerebellum, consistent with iron deposition somewhat above the normal level expected for age (Figure 5.10).

Medications including L-dopa, baclofen, trihexyphenidyl and benzodiazepines had only limited impact. She underwent DBS in her mid-30s with a good but not complete response.

DY94C, DY94P's only son, is currently 27 years old. He was born moderately premature at 34 weeks and had mild to moderate intellectual disability, acquiring fluent speech but not literacy. He also had some behavioural difficulties.

He was first affected by dystonia at 11 years old: initially his neck and trunk were affected, and from 14 his upper limbs also became involved. Lower limb involvement was much milder and he remained ambulant, though with increasing difficulty and postural abnormality (Figure 5.11).

Medication had only limited benefits but he underwent DBS aged 17 with an excellent clinical response: the movement component of his Burke-Fahn-Marsden dystonia rating scale fell from 39.5/120 to 9/120 after treatment while the disability component fell from 6/30 to 2/30. Over the next few years his dystonia has started to progress again but both scores remain well below the pre-treatment baseline.

Figure 5.10: *Axial MRI images from DY94P. (i) Susceptibility-weighted images (SWI) sequences showing hypointensity of the globus pallidus (red arrow), midbrain nuclei (yellow arrow) and cerebellar dentate nuclei (blue arrow). (ii) T2-weighted image showing hypointensity of the globus pallidus (green arrow).*

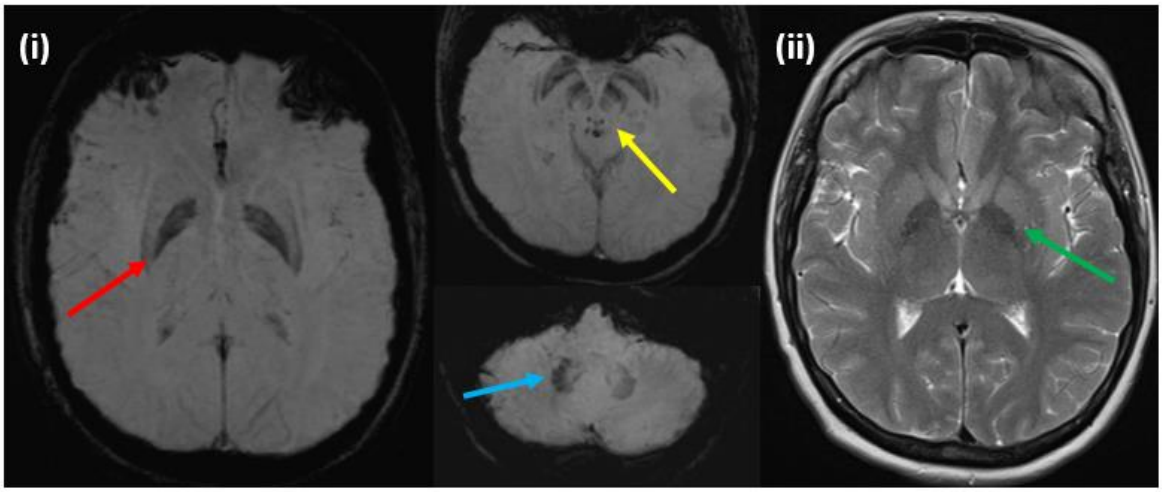
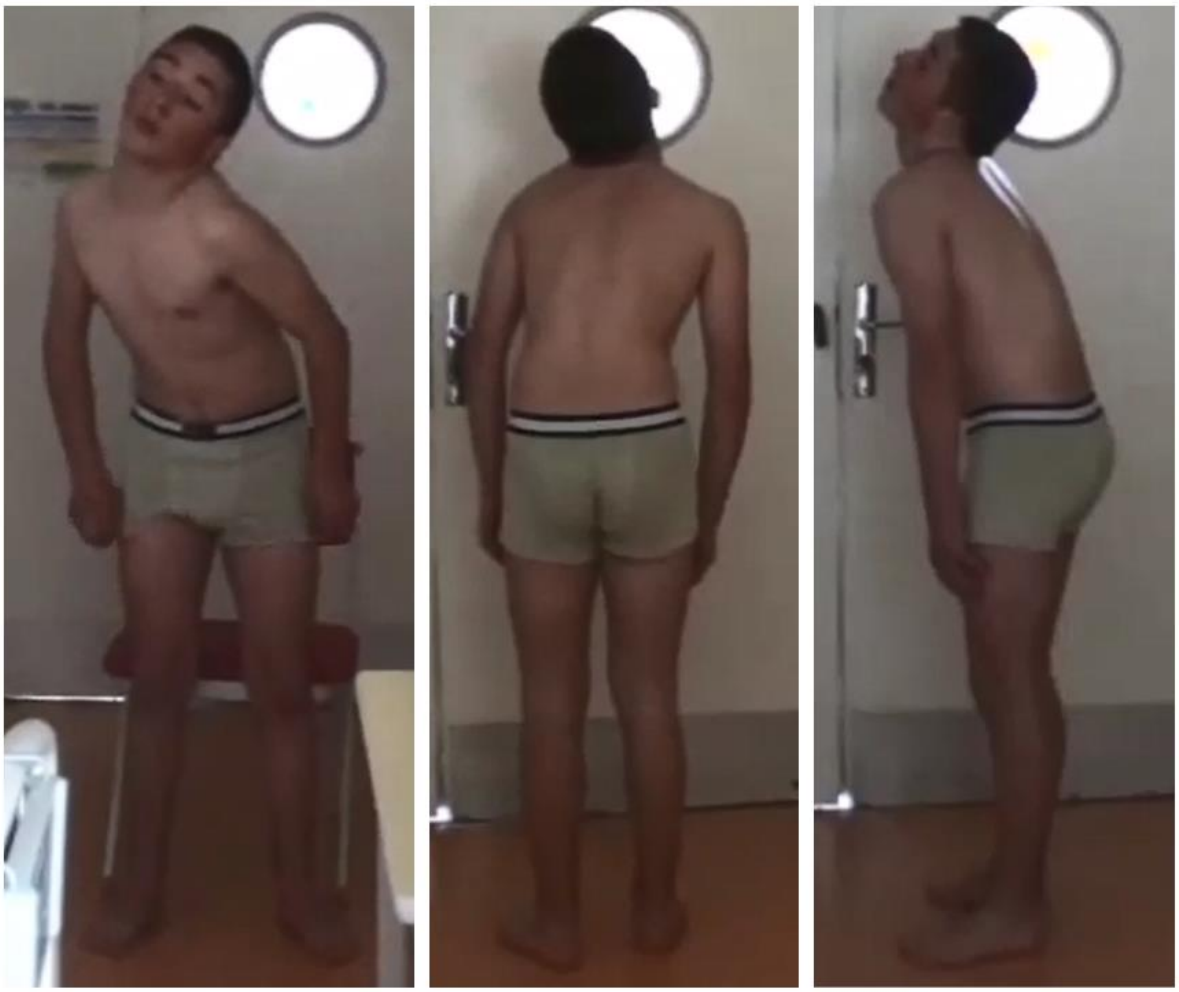


Figure 5.11: *Stills from a clinical video of DY94C aged 16 years (consent for publication in place). Note the marked laterocollis and anterocollis together with lateral and anterior flexion of the trunk. Upper limb posture is mildly dystonic but lower limb posture virtually normal.*



5.6 Genetic analysis

Analysis of WGS data focused on variants shared between DY94P and DY94C due to the strong phenotypic similarity. No obvious pathogenic variant being found in any known disease gene, a shortlist was compiled of variants with strong *in silico* predictors of pathogenicity and a biological role that was at least plausible (**Table 5.2**).

The three highest-priority variants were those resulting in loss of function: *MEF2D, PRRC2B, VPS16* and *ZNF493*. *ZNF493*, however, occurred at low frequencies in population databases which made it a less probable candidate for pathogenicity: the others were not previously reported. Variants were then ranked by pLI and loss-of-function constraint scores: on this, *MEF2D* and *PRRC2B* ranked the most highly, both

with pLI scores of 1, whereas that of *VPS16* is 0. On the other hand, the existence of a recessive dystonia phenotype (with a missense variant predicted to have only a slight impact on protein function) for *VPS16* was a point of interest.(241)

Less pre-existing laboratory evidence indicated a specific role for *VPS16* in neurological disease than was the case for *VPS41,* although it been shown to be involved in mediating the neuroprotective effects of hypoxic preconditioning in rats.(252)

All 12 shortlisted variants were confirmed by Sanger sequencing and listed on GeneMatcher. Over a period of several months, there were no convincingly positive matches for any of the genes.

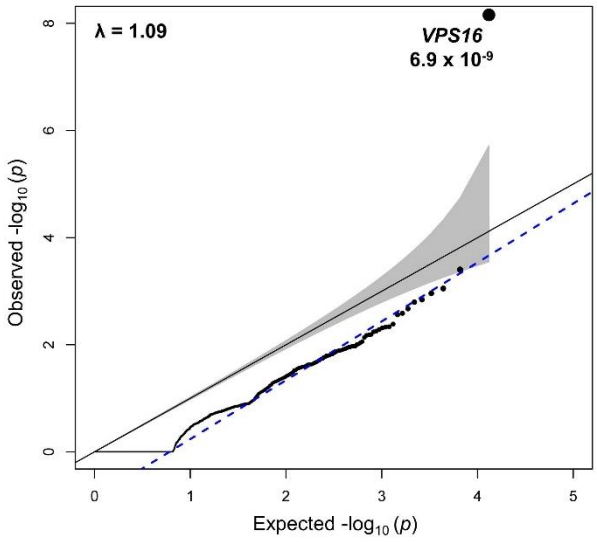
Gene	cDNA change	Protein change	Genomic coordinate (GRCh38)	CADD	SIFT	Mutation Taster	PolyPhen-2	Align GVGD	Provean	gnomAD frequency	Function/Expression
MEF2D	NM_005920.3: c.888dup	p.Arg297 Alafs*55	Chr1: 156475226	NA	NA	Disease causing (1)	NA	NA	Deleterious (-333.868)	Not found	Transcription factor involved in muscle and neuronal differentiation. Ubiquitous.
PRRC2B	NM_013318.3: c.5494dup	p.Arg1832 Profs*311	Chr9: 131484719	NA	NA	Disease causing (1)	NA	NA	Deleterious (-813.772)	Not found	Binds RNA. Ubiquitous.
VPS16	NM_022575.3: c.1094_1095dup	p.Tyr366 Serfs*12	Chr20: 2881955	NA	NA	Disease causing (1)	NA	NA	Deleterious (-1174.44)	Not found	HOPS complex subunit. Ubiquitous.
ZNF493	NM_001076678.2: c.631C>T	p.Arg211*	Chr19: 21423290	NA	NA	Disease causing (1)	NA	NA	Deleterious (-1822.586)	0.0018%	Zinc finger transcription factor. Ubiquitous.
MICAL1	NM_001286613.1: c.1280G>A	p.Arg427Gln	Chr6: 109450054	35	Tolerated (0.14)	Disease causing (1)	HD Prob dam (1.00) HV Prob dam (0.91)	C0	NA	0.002%	Cytoskeletal regulator. Widely expressed.
DHX57	NM_198963.2: c.4114G>C	p.Gly1372 Arg	Chr2: 38798346	33	Deleterious (0.01)	Disease causing (1)	HD Prob dam (0.99) HV Prob dam (0.94)	C0	NA	0.0032%	RNA helicase. Ubiquitous.
WDR91	NM_014149.3: c.715G>T	p.Asp239 Tyr	Chr7: 135205938	32	Deleterious (0)	Disease causing (1)	HD Prob dam (1.00) HV Prob dam (0.96)	C0	NA	0.003%	Modulate endosomal trafficking. Widely expressed.
TANC1	NM_001350064.1: c.2990C>T	p.Ala997Val	Chr2: 159196639	30	Deleterious (0.01	Disease causing (1)	HD Prob dam (1.00) HV Poss dam (0.85)	C15	NA	0.0004%	May form part of post-synaptic scaffold. Ubiquitous.
SEC22A	NM_012430.4: c.367A>G	p.Lys123Glu	Chr3: 123225123	28.7	Deleterious (0)	Disease causing (1)	HD Prob dam (0.99) HV Prob dam (0.98)	C55	NA	Not found	Involved in endoplasmic reticulum to Golgi transport. Widely expressed.
PLCL1	NM_006226.3: c.1930C>G	p.Pro644 Ala	Chr2: 198085447	24.3	Deleterious (0)	Disease causing (0.96)	HD Prob dam (1.00) HV Prob dam (1.00)	C25	NA	Not found	Mediates intracellular signalling. Widely expressed.
SMG8	NM_018149.6: c.410_430dup	p.Leu137_ Ser143dup	Chr17: 59210461	NA	NA	Polymorph (1)	NA	NA	Deleterious (-3.915)	Not found	Involved in nonsense-mediated decay of mRNA. Ubiquitous.

Table 5.2: *Genomic and predicted characteristics of shortlisted variants shared between DY94P and DY94C.*
Colour coding: Red: predictor supports pathogenicity; Yellow: predictor equivocal; Green: predictor opposes pathogenicity.

5.7 Case-finding and characterisation for *VPS16*

In October 2019, through GeneMatcher, I heard from researchers in Germany, Dr Michael Zech and Prof Juliane Winkelmann, who with their team had been conducting a burden analysis on a large cohort of patients with dystonia believed likely to be genetic in origin. This analysis had indicated that rare loss-of-function (LoF) variants (frameshift, nonsense and canonical splice variants) in *VPS16* were significantly enriched in this population compared with healthy controls. Because this analysis was undertaken by the German team without my involvement it will not be discussed in detail here, although a brief summary is included in Chapter 2; a graph of their results is shown for illustrative purposes only **(Figure 5.12).**

Figure 5.12: *Expected versus observed p-values of the loss-of-function model are exome-wide analysis in a cohort of 138 individuals with generalized dystonia and gnomAD controls (64,603 non-Finnish European subjects). 6.9X10-9 is the Fisher's exact test p-value for* VPS16.



Additional families with heterozygous LoF variants in *VPS16* and adolescent-onset progressive dystonia was identified in the Zech cohort, and subsequently through collaborators. Cases were identified from the following additional sources: the Kolling Institute of Medical Research (Sydney, Australia); the Carlo Besta Neurological Institute (Milan, Italy); the Koios Database of the Queen Square Genomics Group at University College London (London, UK); the Genomics England 100K Genomes Project dataset (UK) and Radboud University Medical Centre (Nijmegen, the Netherlands). Databases from Cardiff (Wales) and Dublin (Ireland) were also checked

but no additional cases were found there. Variants identified through WGS or WES and familial segregation were verified by Sanger sequencing at the relevant centres. Finally, 19 affected individuals from 14 families were identified.

Genotypic and phenotypic features of the members of the *VPS16* cohort are summarised in **Tables 5.3-6.**

Table 5.3: *Genomic characteristics of* VPS16 *variants identified. All variants are described with reference to build GRCh38, transcript NM_022575.3. CNV: copy number variant.*

Genomic position	cDNA variant	Protein effect	Variant type	Family
chr20:2835462-3974387	NA	Microdeletion	CNV	6
chr20:2860515	c.436del	p.Ile146Serfs*65	Frameshift	9
chr20:2860534-2860541	c.455_462dup	p.Leu155Alafs*59	Frameshift	13
chr20:2860792	c.559C>T	p.Arg187*	Stop-gain	4,10
chr20:2862601-2862602	c.1094_1095dup	p.Tyr366Serfs*12	Frameshift	7
chr20:2863068	c.1335T>G	p.Tyr455*	Stop-gain	3
chr20:2863102	c.1367+2T>C	p.?	Splice site loss	8
chr20:2864178	c.1612-1G>C	p.?	Splice site loss	14
chr20:2864288	c.1720+1G>C	p.?	Splice site loss	11
chr20:2864631	c.1903C>T	p.Arg635*	Stop-gain	1,2,12
chr20:2865039-2865040	c.1988_1989insG	p.Asn663Lysfs*2	Frameshift	5

Table 5.4: *Probands' variants, inheritance and demographic characteristics.*

Family	Patient	Mutation	Inheritance*	Sex	Current age (y)
1	1	p.Arg635*	NK	M	27
2	2	p.Arg635*	NK	F	42
3	3P	p.Tyr455*	I-AP	F	24
	3Aunt			F	50
4	4	p.Arg187*	I-AP	M	26
5	5	p.Asn663 Lysfs*2	NK	M	38
6	6	Micro-deletion: chr20:2835462-3974387	DN	F	21
7	7M	p.Tyr366 Serfs*12	I-SP	F	45
	7C			M	26
8	8	c.1367+2 T>C	I-SP	F	69
9	9F	p.Ile146 Serfs*65	I-SP	M	62
	9C			M	30
10	10	p.Arg187*	NK	M	32
11	11S1	c.1720+1 G>C	I-AP	M	30
	11S2			M	24
12	12	p.Arg635*	I-AP	M	33
13	13F	p.Leu155Alafs*59	I-SP	M	58
	13C			F	17
14	14	c.1612-1 G>C	NK	M	57

Patient identifiers: XC: child; XF: father; XM: mother; XP: proband; XS: sibling
Abbreviations: DN: de novo; F: female; I-AP: inherited from asymptomatic parent; I-SP: inheritance from symptomatic parent; M: male, NK: not known; y: years
** For multigenerational families, inheritance refers to the younger affected generation.*

Table 5.5: *Features of probands' dystonia. Abbreviations: Botox: botulinum toxin type A; DBS: deep brain stimulation; L: left; Lo: lower; R: right; TXY: trihexyphenidyl; U: upper; y: years*

Patient	Age of dystonia onset (y)	Initial movement disorder	Current movement disorder				Helpful interventions
			Axial	Limb	Facial/ bulbar	Ambulant?	
1	10	Cervical dystonia	Y	Y (R>L)	Y (severe speech involvement)	Y	DBS (partial improvement)
2	13	Writer's cramp	Y (mild)	Y (U>Lo)	N	Y	N
3P	17	Upper limb dystonia (R)	Y	Y	Y (severe speech involvement)	N (lost in adulthood)	N
3Aunt	4	Upper limb dystonia (R)	Y	Y (U only)	Yes (speech involvement)	Y	L-DOPA (minor improvement); Sensory trick
4	15	Speech involvement then cervical dystonia	Y	Y (U only)	Y	Y	N
5	30	Cervical dystonia	Y	Y (U only)	Y	Y	N
6	11	Upper limb dystonia (R)	Y	Y (U>Lo)	Y	Y	N
7M	19	Lower limb dystonia (R)	Y	Y	Y	N (lost in adulthood)	DBS (significant improvement)
7C	11	Facial + cervical dystonia	Y	Y (mild; U>Lo)	Y	Y (with difficulty)	DBS (significant improvement)
8	7	Cervical + upper + lower limb (R) dystonia	Y	Y (U and Lo)	Y (marked dysphonia)	Y	L-DOPA; Botox; Sensory trick
9F	14	Oromandibular dystonia	Y (cervical)	Y (U and Lo)	Y (dysphonia; blepharospasm)	Y	L-DOPA; Botox; Sensory trick
9C	Unknown	Oromandibular dystonia	Y (cervical)	N	Y (blepharospasm; oromandibular)	Y	N
10	19	Oromandibular dystonia	Y (cervical)	Y (mild; U only)	Y (laryngeal)	Y	TXY; Botox; Sensory trick
11S1	8	Writer's cramp	Y (cervical)	Y (U>Lo)	Y (severe dysphonia)	Y	Botox (larynx)
11S2	7	Writer's cramp	N	Y (Lo>U)	Y (dysphonia; oromandibular)	Y	N (DBS no use)
12	3	Speech involvement	Y	Y	Yes (anarthria)	N (lost age 20y)	N (DBS no use)
13F	50	Cervical dystonia	Y	Y (U>Lo)	N	Y (with difficulty)	.
13C	9	Lower limb dystonia (R)	N	Y	N	Y	TXY
14	14	Cervical dystonia + speech involvement	Y	Y (U>Lo)	Y	Y	Botox; Sensory trick

Table 5.6: *Additional clinical features of patients with* VPS16 *mutations.*

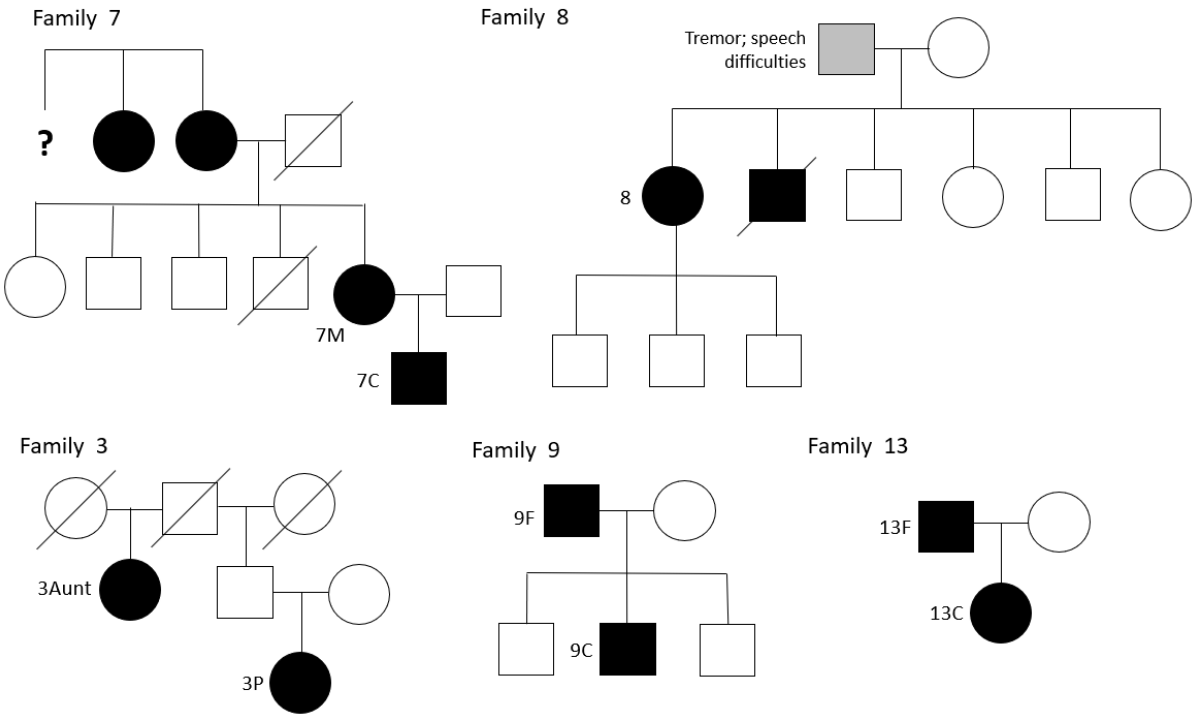
Family	Patient	Family History	Ethnicity	Cognition	Psychiatric disorder	Additional features	MRI brain findings (age at imaging)
1	1	No	Polish	Normal	No	.	Hypointensity of globi pallidi on T2 (10y)
2	2	Yes	German	Normal	No	.	Reportedly normal (NA)
3	3P	Yes	Austrian	Moderate ID	No	Mild congenital microcephaly; mild spasticity	Very subtle hypointensity of midbrain nuclei on SWI; mild cerebral atrophy (21y)
	3Aunt			Normal	No	.	Reportedly normal (NA)
4	4	No	Slovak	Moderate ID	Yes: mood disorder, impulsivity	Epilepsy	Result unavailable
5	5	No	German	Normal	No	.	Reportedly normal (NA)
6	6	No	German	Normal	No	.	Reportedly normal (NA)
7	7M	Yes	French	Mild ID	Mood disorder; impulsivity	.	Hypointensity of globi pallidi, midbrain and dentate nuclei on T2/SWI (34y)
	7C			Moderate ID	Behaviour difficulties	.	Slim putamina post-DBS (17y)
8	8	Yes	White Australian	Normal	No	FH of mental illness: brother had schizophrenia; father had behavioural anomalies	Result unavailable
9	9F	Yes	White Australian	Normal	No	.	Mild generalised atrophy; germinolytic cyst (NA)
	9C			Normal	No	.	Not done
10	10	No	British	Normal	Yes (mood disorder; aggressive behaviour)	Non-REM parasomnia	Paucity of white matter, mild hypointensity of globi pallidi and midbrain nuclei on SWI (32y)
11	11S1	Yes	Italian	Normal	No	.	Reportedly normal (10y)

	11S2			Normal	No	.	Reportedly normal (16y)
12	12	No	Italian	Mild ID	OCD (onset 13y)	Epilepsy (focal; onset 29y; controlled on carbamazepine)	Cerebellar cyst; pallidal atrophy post-DBS (20y)
13	13F	Yes	Dutch	Normal	PTSD	.	Mild atrophy, small/atrophic putamina, minor grey matter heterotopia (55y)
	13C			Normal	No	Prolactinoma	Hypothalamic hamartoma, minor grey matter heterotopia (16y)
14	14	No	British	Normal	Bipolar disorder	.	Reportedly normal (14y)

Abbreviations: DBS: deep brain stimulation; ID: intellectual disability; NA: not available; OCD: obsessive compulsive disorder; PTSD: post-traumatic stress disorder; REM: rapid eye movement (sleep); SWI: susceptibility weighted imaging; y: years

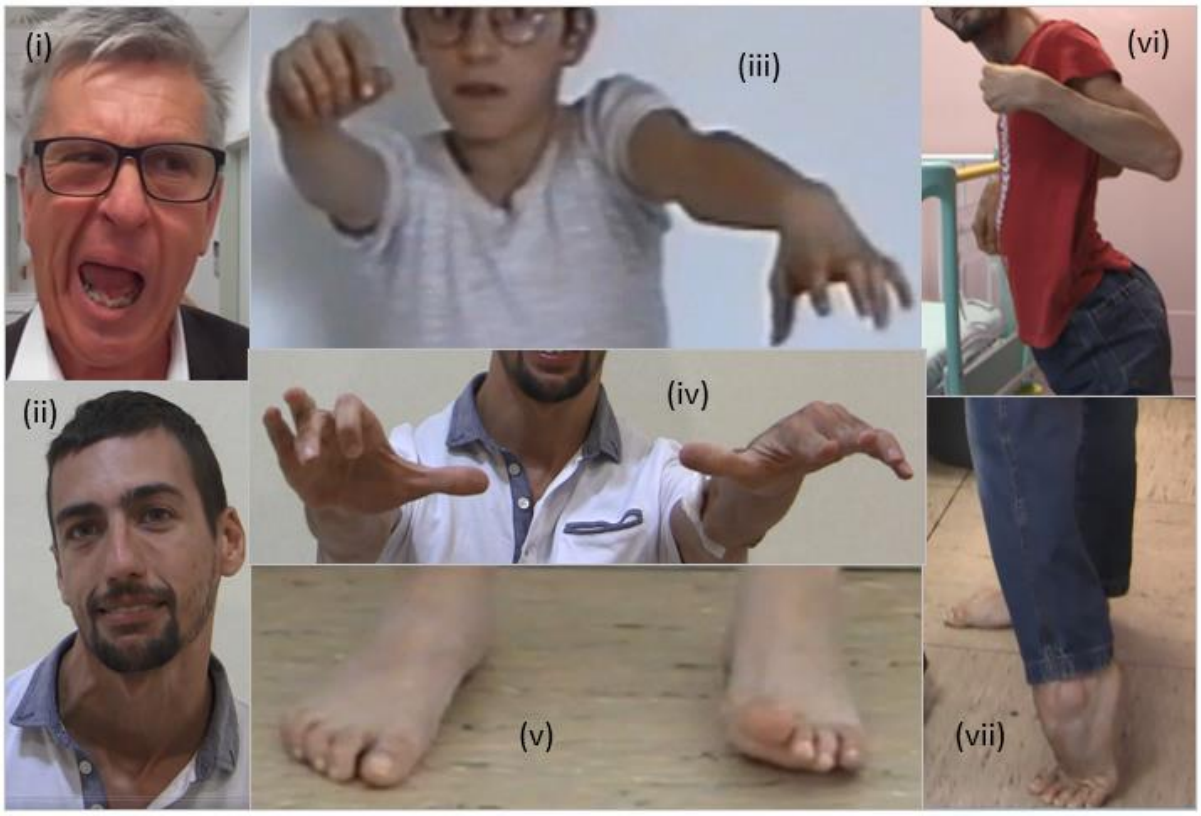
In five multigenerational families (3, 7, 8, 9 and 13) with more than one affected individual there was a clear pattern of AD inheritance **(Figure 5.13).** Segregation analysis was possible for nine families: of these, *de novo* occurrence was confirmed in one; inheritance from a symptomatic parent was found in four; and inheritance from an apparently non-manifesting parent in four, indicating incomplete penetrance **(Table 5.4).**

Figure 5.13: *Pedigrees of multigenerational families with* VPS16 *variants. Circle: female; square: male; black: affected; grey: possibly affected. Image appears in Steel et al., 2020.(77)*



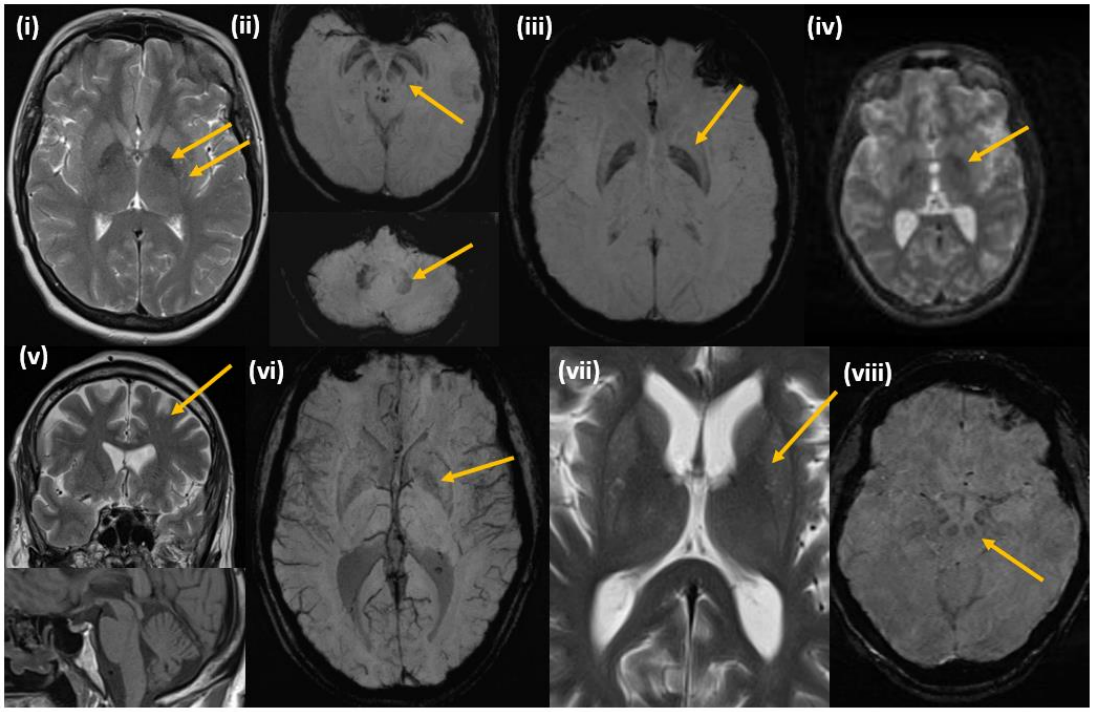
Affected individuals presented with a progressive early-onset dystonia (median age 12 years, range 3-50 years), with prominent oromandibular, bulbar, cervical and upper limb involvement **(Figure 5.14).** Progressive generalisation ensued in most cases, though most remained ambulant, and only a minority (16%) lost the ability to walk in adulthood **(Table 5.5).** Additional clinical features of mild to moderate intellectual disability and neuropsychiatric symptoms were present in approximately one-third of patients, and 50% of families had a positive family history of dystonia **(Table 5.6).** A degree of inter- and intrafamilial phenotypic variability was evident, both with regard to age of symptom onset and dystonia severity.

Figure 5.14: *(i) Patient 9F demonstrating orofacial dystonia elicited during speech; (ii) Patient 11S1 showing cervical dystonia; (iii) Patient 12 showing upper limb posturing; (iv) Patient 11S1 showing hand posturing; (v) Patient 12 showing spontaneous striatal toe on the left; (vi) Patient 12 as an adult, standing, showing exaggerated lumbar lordosis (vii) Patient 12 as an adult, standing, showing involuntary plantar flexion/tiptoe posture. Image appears in Steel et al., 2020.(77)*



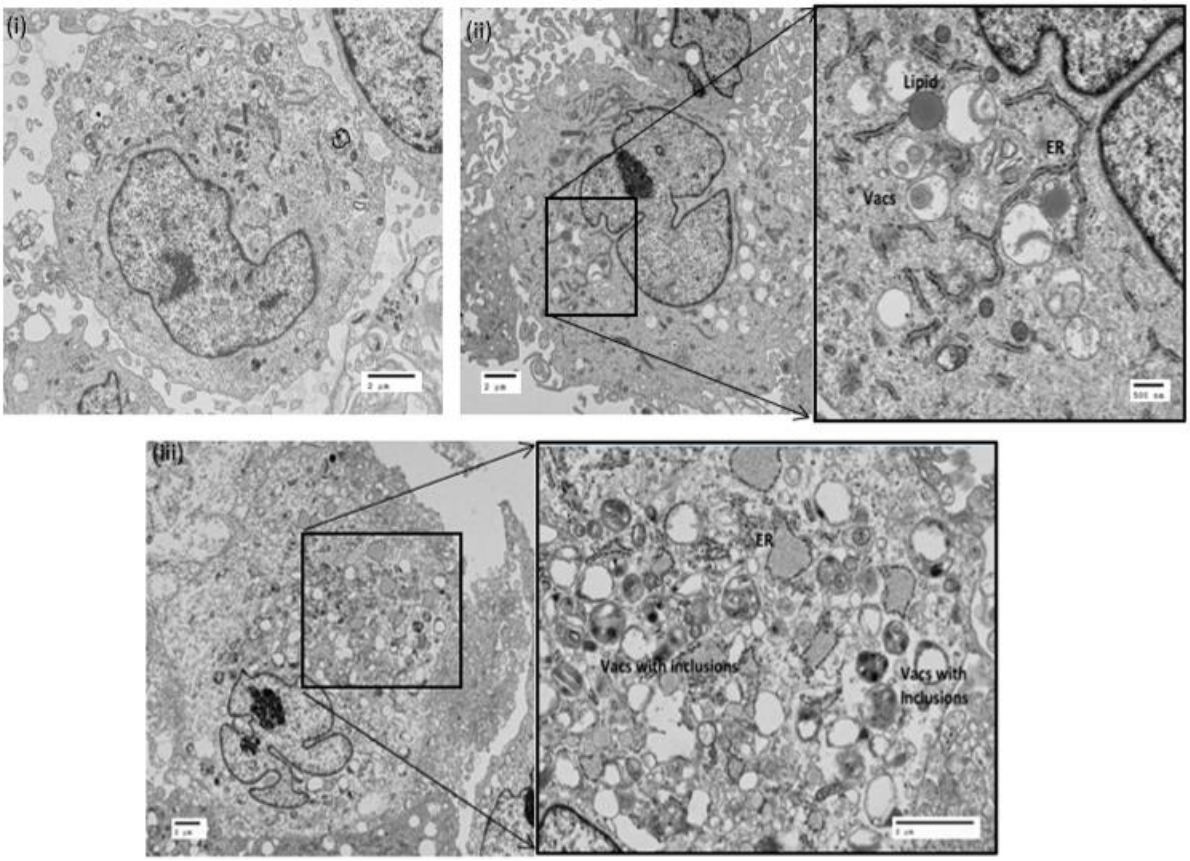
Routine diagnostic testing was unremarkable. In four individuals, magnetic resonance imaging (MRI) showed bilateral and symmetrical hypointensity of the globi pallidi and sometimes also the midbrain and dentate nuclei on MR sequences known to demonstrate susceptibility (T2-weighted, T2*-weighted, susceptibility-weighted datasets), suggestive of iron deposition. Mild generalised cerebral atrophy was also apparent in four individuals. Although not grossly abnormal, caudate nuclei and putamina appeared relatively small and bright on T2 (**Figure 5.15**). Some patients had a partial response to L-dopa, trihexyphenidyl and/or botulinum toxin type A injections. DBS was also beneficial for some, but not all, patients; sustained improvement in motor and disability scores for the Burke-Fahn-Marsden Dystonia Rating Scale were observed for Patient 7C.

Figure 5.15: *Selected MRI brain images for patients with VPS16 dystonia*
Abnormalities indicated by yellow arrows. (i) Axial T2 image from Patient 7M (aged 34 years, pre-DBS) shows hypointensity consistent with iron deposition in the globi pallidi; (ii) Susceptibility-weighted images (SWI) from Patient 7M showing hypointensity in the midbrain nuclei (above) and dentate nucleus of the cerebellum (below); (iii) Axial SWI from patient 7M showing hypointensity of the globi pallidi; (iv) Axial T2 image from Patient 1 (aged 10 years) showing hypointensity of the globi pallidi; (v) Subtle generalised atrophy in Patient 9F, demonstrated in a coronal T2 image of the cerebrum (above) and a sagittal T1 image of the cerebellum (below); (vi) Axial SWI from Patient 10, aged 32 years, showing hypointensity in the globi pallidi; (vii) Enlarged axial T2 image from Patient 13F, aged 55 years, showing relatively small, bright caudates and putamina; (viii) Axial SWI image from Patient 3P (aged 21 years) showing hypointensity of the midbrain nuclei. Image appears in Steel et al., 2020.(77)



EM of patient-derived fibroblasts and lymphocytes (again, undertaken by Mr Glenn Anderson), compared with healthy controls, showed increased clusters of vacuoles, with some containing inclusions in the form of particulate or laminated material (**Figure 5.16**).

Figure 5.16*: Representative EM images of patient-derived and control cells (i) Control fibroblast from a healthy individual. (ii): From Patient 8; (iii): From Patient 11S2; Black rectangles in (ii) and (iii) indicate the region enlarged in the following image.*



5.8 Splicing experiment in *VPS16* patient derived fibroblasts

Fibroblasts were available from two families with canonical splicing variants: Proband 8 (c.1367+2T>C, p.?) and Proband 11S1 (c.1720+1C>G, p.?). To investigate the impact of these variants I undertook a splicing experiment. RNA was extracted from fibroblasts and converted to cDNA by reverse transcription as described in Chapter 2.

Splicing primers were designed against the cDNA sequence to assess each splice site with sequences as shown in **Table 5.7.**

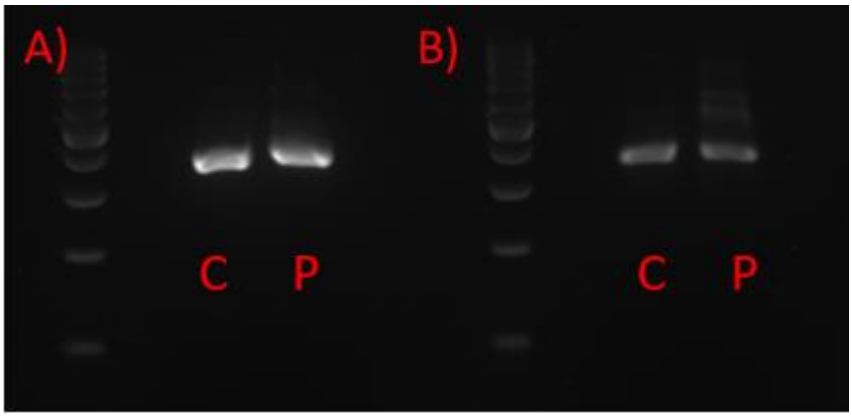
PCR with an annealing temperature of 62°C was undertaking using these primers and the product was run on a 3% agarose gel as described. cDNA from the fibroblasts of a healthy individual (kindly provided by Dimitri Budinger) was used as a control.

Table 5.7: *Splicing primers used in VPS16 experiments*

Variant	Forward Primer	Reverse Primer
c.1367+2T>C	GCGTGTGCTCAATGCTGTTC	GAGCGTGGCTCATACTCCAG
c.1720+1C>G	CGCCTGGTGTCTCTTACTCC	TTCCTGGTGATTGTCATCCT

I had expected at this stage to see a single PCR product for the control and double products for the two probands: one the same size as the control product, representing normal splicing of mRNA transcribed from the healthy allele, the other slightly smaller, representing mRNA transcribed from the variant allele, most probably with an exon skipped (exon 14 in the case of proband 8; exon 17 in the case of proband 11S1). In fact, however, a single product indistinguishable from the normal product was visualised (**Figure 5.17**).

Figure 5.17: *Electrophoresis (60 minutes on a 3% gel) of the relevant cDNA from A) Proband 8 and b) Proband 11S2. C: control; P: proband.*



(The very faint smears seen above the main product are *not* thought to reflect aberrant splicing. They are larger than the control product, and very faint, and therefore are more likely to represent a small amount of retained contaminating gDNA, despite the use of a DNA digestion step.)

cDNA was extracted from the gel for further analysis, put through a further PCR using the same primers, and sequenced. The fibroblasts had at this stage been in storage for over a year and had passed through the hands of several teams so I thought it conceivable that an accidental switch might have occurred: for this reason I ran both products with both sets of primers. This made no difference. I also confirmed the

presence of the expected variant in gDNA extracted from one set of fibroblasts. The sequencing results revealed normal exon-exon boundaries with no suggestion of a double trace.

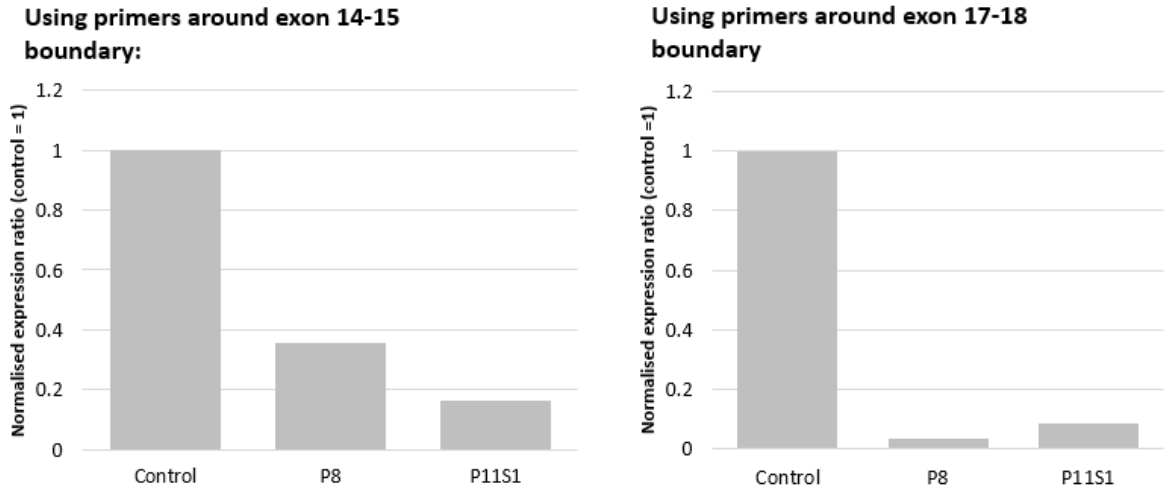
In view of the other evidence suggesting that these patients really did have *VPS16*-related disease (namely, unanimous *in silico* prediction of loss of the splice site; compatible familial segregation; absence of the variants from normal population databases; phenotypic correspondence with other cases; and EM evidence of vesicular abnormalities in fibroblasts from both patients) I hypothesised that the aberrant mRNA might be being degraded by nonsense-mediated decay. I therefore undertook an experiment to assess this, by performing qRT PCR on the cDNA from both probands and the healthy control, as described in Chapter 2. I used the same cDNA primers as above but, because the object was to confirm the overall level of *VPS16* mRNA, used both sets of primers for both probands.

QRT PCR showed a very substantial reduction (>50%) in the level of *VPS16* cDNA in both probands compared with the control. As predicted, results were similar no matter which set of cDNA primers were employed **(Figure 5.18).** Technical triplicates were used but because only one fibroblast sample was available for each proband no biological replicates were possible. The validity of considering two different primer sets as a biological replicate is questionable. I have therefore not measured statistical significance for these results and note that they cannot be considered statistically robust.

These results are consistent with both splicing variants resulting in aberrant transcripts which are destroyed by efficient nonsense-mediated decay, resulting in lowered overall levels of *VPS16* mRNA. This is a recognised difficulty in analysing aberrant splicing.(253, 254) Where it is necessary to analyse the impact of a splicing variant further and the aberrant transcript cannot be detected under normal circumstances, one option is to culture patient-derived cells in the presence of an inhibitor of nonsense-mediated decay such as anisomycin(253) or cycloheximide.(254) This allows the aberrant transcript to escape degradation and so be detected and

sequenced. In this instance, however, I did not proceed to this experiment due to lack of time.

Figure 5.18: Normalised expression ratios for *VPS16* mRNA in fibroblasts derived from probands with splice-site variants compared with a healthy control.



5.9 HOPs-Associated Neurological Disorders (HOPSANDs)

Since the publication of our paper, at least nine additional patients with DYT-30 have been described in the literature in five separate reports.(255-259) One additional patient with a missense variant which the authors considered likely to be pathogenic has also been reported.(260) In addition, I have subsequently diagnosed two other members of our own cohort with DYT-30 (unpublished). This suggests that pathogenic *VPS16* variants may be less rare than we had assumed.

Additionally, an autosomal recessive phenotype with features of a classic lysosomal storage disorder with multisystem features including skeletal, brain and visceral involvement has been described independently by two groups.(261, 262) In two families a missense variant was implicated; in the other a non-canonical splicing variant: parents were asymptomatic. Presumably such variants cause only moderate disruption to VPS16 function and can be tolerated in the heterozygous state. It is notable that the only previously reported family with biallelic *VPS16* variants had a far milder phenotype consisting of dystonia only.(241) This could be explained by the fact

that the variant in this kindred, c.156C>A, p.Asn52Lys, is predicted to have only mild effects on protein function – indeed, the majority of *in silico* tools assessed it as harmless.

For SCAR29, 12 additional patients in seven families have been described.(248, 249) These publications were slightly later than ours, but independent, and van der Welle et al's work was deposited on bioarXiv, a pre-publication server for biological research, while we were still preparing our manuscript. The symptoms of the additional individuals overlap closely with our proband: ataxia, global developmental delay and in some cases regression, dystonia, and retinal abnormalities. MRI findings were also similar to our proband, showing hypoplasia or atrophy of the corpus callosum and cerebellar vermis. The clinical descriptions are all somewhat brief and a more detailed, clinically focused case series would undoubtedly help characterise the disorder.

Using patient-derived fibroblasts, van der Welle *et al* demonstrated failure of endocytic cargo to reach lysosomes. The lysosomes themselves were small, numerous and enzymatically active. This is consistent with the absence of clinical features of a classical lysosomal storage disorder observed in our patient.

With co-immunoprecipitation techniques, they demonstrate that the ability of both the mutant alleles to interact with other subunits of the HOPS complex was reduced. They were also able to demonstrate a failure of delivery of endocytic cargo to lysosomes in patient-derived fibroblasts as compared with wild-type, by measuring reduced and delayed co-localisation of gold-labelled bovine serum albumin with lysosomal membrane proteins.

In nutrient-starved patient-derived fibroblasts, they showed an impaired autophagy response as indicated by increased levels of LC3-II, a marker of autophagic activity which is normally recycled through autophagosome-lysosome fusion. They found similar changes in HeLa cells with *VPS41* knocked out using CRISPR, and found that the phenotype could be rescued by wild-type *VPS41* but not by either of the patient mutant alleles.

They further demonstrated that patient-derived fibroblasts displayed reduced recruitment of mTORC1, the mammalian target of rapamycin complex 1, to lysosomal membranes, and that the normal starvation response – in which mTORC1 colocalisation with lysosomes is reduced – was abolished. Once again, the phenotype could be replicated in *VPS41* knockout HeLa cells and rescued by wild-type, but not mutant, *VPS41* expression.

As mTORC1 acts as a transcriptional regulator, this could open the door to understanding how a specific defect of endosomal trafficking and autophagy could have broader effects, through disrupting normal adaptive mechanisms of gene expression.

They also used the α-synuclein-based *C. elegans* model of Parkinson's disease to show that the mutant alleles failed to have the neuroprotective effect described above.

Thus multiple lines of evidence converge to indicate that 1) *VPS41* is essential for normal endosome/autophagosome-lysosome interactions and 2) this function permits it to protect neurons against the toxic effects of abnormal protein accumulation. We hypothesize that the large numbers of membrane-bound vacuoles seen in patient-derived cells represent endosomes and/or autophagosomes which have failed to fuse with lysosomes, meaning their contents cannot be correctly processed. The combination of a neurodegenerative phenotype and clear vacuolar abnormalities in our patient suggest that this mechanism may underlie the pathology of autosomal recessive *VPS41* disease.

Including *VPS16* and *VPS41*, four of the six HOPS subunits now have a described disease phenotype. *VPS39* and *VPS18* – so far – do not, but it is reasonable to hypothesise that damaging biallelic variants would be likely to cause autophagic and/or lysosomal dysfunction. Knockout of either gene is lethal in mice; conditional CNS ablation of *VPS18* in mice using a Cre-lox system causes neurodegeneration with evidence of impaired endocytosis and autophagy.(263) *VPS16, VPS41, VPS33A* and *VPS11* all have severe AR phenotypes with clear features of lysosomal storage defects.(77, 239, 240, 262) For *VPS33A* and *VPS16* there is obvious multisystem

involvement; for *VPS41* and *VPS11* CNS pathology is much the most prominent but there is more subtle evidence of wider involvement such as visceromegaly or peripheral neuropathy.

VPS16 and *VPS11* both have additional AR phenotypes – reported in only one family each – consisting of apparently isolated dystonia.(76, 241) These may reflect less damaging variants as in each case the heterozygous parent carriers are asymptomatic.

So far, only *VPS16* has a known AD phenotype. It would be interesting to know whether heterozygous LoF variants in any other HOPS subunits were associated with variably penetrant phenotypes and/or acted as risk factors for neurodegenerative disorders. Such work would require statistical analysis of larger cohorts of patients. Intriguingly, heterozygous *VPS39* variants have been reported to be enriched in people with schizophrenia.(264)

The prevalence of CNS involvement, and specifically dystonia, in disease involving HOPS subunits, has led to the suggestion of creating a category of “HOPSANDs” – HOPS-Associated Neurological Disorders” to reflect their common underlying pathways.(76) It should be remembered, however, that VPS16, 18, 11 and 33A are also essential subunits of the CORVET complex, a related but distinct complex involved in early endosomal tethering.(233) The extent to which dysfunction of one complex rather than the other leads to the observed manifestation of disease remains to be explored.

5.10 Discussion

In this chapter I have discussed two newly described movement disorders, both due to defect involving subunits of the HOPS complex, the index cases for which I identified through my project. Proving pathogenicity of the variants identified required collaboration between a large team of people, including statisticians, electron microscopists, geneticists and clinicians.

DYT-30, the *VPS16*-related disorder, is an AD adolescent-onset progressively generalising dystonia, sometimes with associated neuropsychiatric features, and is variably penetrant. This last feature is probably what led to it being discovered relatively late, as it is hard to demonstrate pathogenicity convincingly when a variant does not segregate clearly with disease within a family, unless a larger cohort can be assembled.

SCAR29, related to *VPS41*, is a progressive infant-onset neurodegenerative condition characterised by developmental regression, dystonia, ataxia, optic atrophy, neuropathy and some systemic features. All cases described so far have involved homozygous variants in consanguineous families, although compound heterozygosity for damaging variants would doubtless also be pathogenic.

The conditions are clinically similar only in that they involve early-onset progressive movement disorders, but at the pathophysiological level they may be more closely linked. *VPS16* and *VPS41* are both subunits of the HOPS complex, vital for autophagy and lysosomal function, processes which are firmly implicated in other neurodegenerative conditions. Full understanding of the aetiology of these disorders will require more research into both the functions performed by each individual protein, and the reasons why disrupting a protein expressed in almost all cells manifests with a primarily neurological phenotype. Patient-derived induced pluripotent stem cells (iPSCs) might prove useful here, permitting techniques such as single-cell RNA sequencing to be used to assess which downstream genes and pathways in which cell types are disrupted in HOPSANDs.

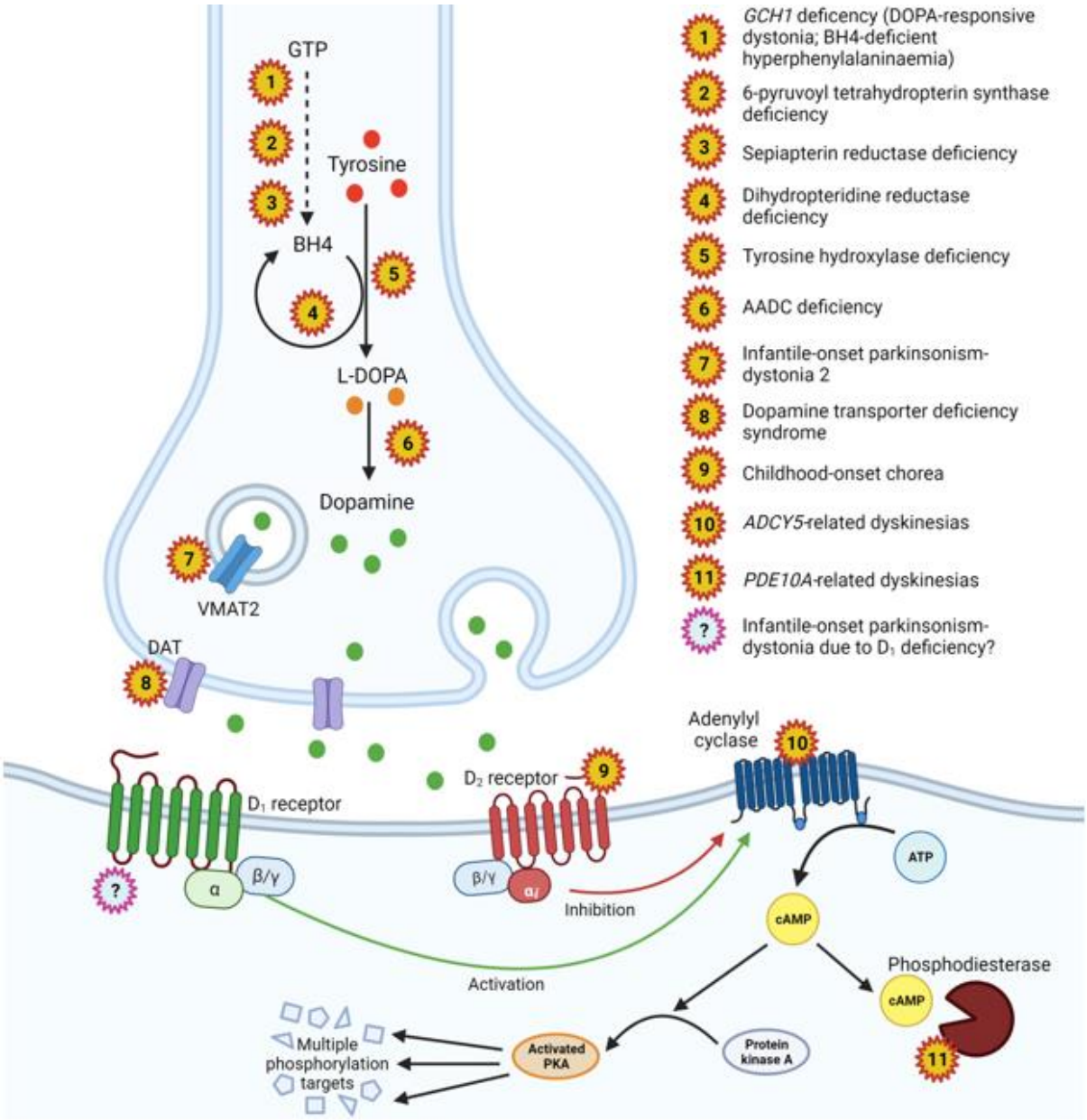
Chapter 6: *DRD1* and disorders of dopamine receptors

6.1 Introduction

Dopamine is a key monoamine neurotransmitter that plays a central role in the initiation and control of voluntary movement, motor learning, motivation and executive function.(265) Within the CNS, dopamine is primarily synthesised by and expressed in neurons of the midbrain (substantia nigra and ventral tegmental area) with extensive projections to the dorsal and ventral striatum and the prefrontal cortex.(266). Dopamine is a product of the monoamine synthesis pathway, in which tyrosine is converted to levodopa by tetrahydrobiopterin-dependent tyrosine hydroxylase, before conversion to dopamine by aromatic L-amino acid decarboxylase (AADC). Dopamine is then packaged into synaptic vesicles by the vesicular monoamine transporter 2 (VMAT2) for release into the synaptic cleft **(Figure 6.1)**. In the dorsal striatum, dopamine acts on cell surface dopamine receptors of medium spiny neurons (MSNs), GABAergic inhibitory neurons which form the major efferent projections of the basal ganglia, in the direct pathway (reducing inhibitory output of the globus pallidus interna and substantia nigra pars reticulata to the thalamus) and the indirect pathway (increasing inhibitory output)(267, 268) (see **Chapter 1, Figure 1.2**).

There are five dopamine receptor subtypes, a group of G-protein coupled receptors (GPCRs) divided into two classes: D_1-like (D_1 and D_5) and D_2-like (D_2, D_3 and D_4). D_1-like receptors are coupled with $G_{s\alpha}$ proteins, where dopamine binding stimulates adenylyl cyclase to increase cyclic adenosine monophosphate (cAMP) levels, leading to activation of protein kinase A and generally an excitatory response.(269, 270) D_2-like receptors are coupled with G_i proteins, which inhibit adenylyl cyclase and reduce cAMP, generally exerting an inhibitory effect.(270) Striatal MSNs preferentially express either D_1 or D_2 receptors. The direct (prokinetic) pathway consists primarily of D_1-expressing MSNs while the indirect (antikinetic) pathway has mainly D_2-expressing MSNs.(9) Broadly speaking, D_1 activation increases motor activity whereas D_2 activation reduces it .

Figure 6.1 *Schematic representation of a dopaminergic synapse, indicating the location of genetic defects leading to various childhood-onset neurotransmitter diseases.*



Defective dopaminergic homeostasis is associated with a broad range of neurological and neuropsychiatric diseases. A number of rare monogenic neurotransmitter diseases attributed to impaired dopamine synthesis, metabolism, transport or signalling have been described **(Figure 6.1):** these childhood-onset diseases are typically characterised by neurodevelopmental delay, complex hyperkinetic and hypokinetic movement phenotypes, autonomic features and mood/ behaviour symptoms.(11, 271-276) Dopaminergic dysregulation is also implicated in a wide

range of common human diseases from Parkinson's disease (PD)(265) and restless legs syndrome(277) to neuropsychiatric disorders such as depression, schizophrenia,(278) and attention deficit hyperactivity disorder (ADHD).(279) The specific contribution of dopamine receptors to disease is increasingly recognised: polymorphisms in dopamine receptor genes have been implicated in addiction, psychiatric illness and PD complications.(280-283)

In 2021, a hyperkinetic disorder associated with gain-of-function variants in D_2 receptor was described:(284) following the report of the index kindred, we identified and reported two more unrelated affected individuals, in collaboration with Dr Niccolò Mencacci and his team.(285) I will describe this in section 6.2. In the rest of this chapter, I will describe our identification of a new dopamine receptor disorder associated with recessive loss-of-function variants in the gene encoding the D_1 receptor, *DRD1*. Although we have so far only been able to identify a single patient with this condition, there is considerable supporting evidence for the pathogenicity of the variant.

6.2 Childhood-onset chorea associated with a heterozygous *DRD2* variant

In 2021, van der Weijden *et al* described Dutch kindred in which eight people over four generations were affected by a choreiform disorder with onset in childhood or adolescence. They shared a heterozygous *DRD2* variant: c.634A>T, p.Ile212Phe. The authors demonstrated that, in a HEK293 cell overexpression model, the variant showed increased basal activation and increased response to agonists, consistent with gain of function.(284)

At the point this study was published, we were already aware of a suspicious *de novo* *DRD2* variant in CD57: c.1121T>G, p.Met374Arg, whom I had analysed on a triome basis and who had no other variants likely to be pathogenic. The variant is absent from gnomAD and *in silico* predictors of pathogenicity were very strong (CADD 32; Polyphen-2 predicted "probably damaging" by HDiv and HVar; "Deleterious (95|5)" on MutationTaster v.2021; "Deleterious (0.03)" on SIFT), and the amino acid residue in

question is conserved back to zebrafish. We had not published it as we had not yet identified a convincing second case despite listing on GeneMatcher or had a chance to undertake functional studies. Shortly afterwards, however, we liaised with Dr Mencacci's team, who were involved in the care of a young man who had similar presentation of an infant-onset, nonprogressive movement disorder and intellectual disability, and who had the same variant, also *de novo*.

Case History: CD57

CD57P is a three-year-old girl from Chile who presented at four months old with choreiform movements involving all four limbs and her tongue. She had global developmental delay, achieving independent sitting at 15 months and walking at two years. Expressive language was severely impaired but receptive language relatively preserved. MRI, EEG, and neurometabolic investigation including CSF neurotransmitter assays were unremarkable. She continued to make steady developmental progress with no apparent progression of her movement disorder.

Serendipitously, a different variant affecting the same residue (Met374Leu) had already been modelled for studies the function of the D_2 receptor when embedded in a lipid membrane and had been shown to increase basal activation of the receptor.(286) Of note, *DRD2* is moderately constrained for truncating variants (pLI 0.75, o/e 0.17) but not significantly constrained for missense variants (o/e 0.56).(103) Because gain-of-function variants are much less common compared to loss-of-function, constraint scores are not very informative concerning them. We therefore suspect that the p.Met374Arg variant is likely to cause gain of function, like that described in the van der Weijden paper.

Gain of D_2 receptor function is expected to decrease activity in the indirect pathway (because the effect of D_2 stimulation is inhibitory) and therefore to reduce inhibitory inputs to the motor cortex, which would be consistent with a hyperkinetic phenotype. Conversely, variants causing loss of function in either D_1 or D_2 might be expected to result in parkinsonism – which leads on to the next finding.

6.3 A child with unexplained infantile parkinsonism-dystonia

Case History: DP35

The proband is the first child of healthy Afghani parents who are first cousins (Figure 6.2). There was no relevant family history. There were no complications in pregnancy or birth other than a suspicion of reduced foetal movements, and she was born at full term by forceps-assisted vaginal delivery, in good condition and not requiring any resuscitation or special care. Birth weight was 25th centile for gestation and head circumference 50th centile.

By six months of age, developmental delay was apparent: she never acquired the ability to sit, roll or babble, although she was visually alert and had a social smile. From infancy onwards she experienced recurrent episodes of severe generalised dystonia: these, sometimes triggered by pain, anxiety or excitement, were characterised by neck arching and limb posturing, accompanied by upward eye deviation, and could persist for hours. They could generally be terminated by sleep. Other symptoms included constipation, excessive sweating and chronic nasal congestion. Enteral feeding due to unsafe swallow was commenced at three years old and she required treatment for gastrointestinal reflux and went on to have difficulties with feed tolerance due to abdominal pain.

On examination at four years old, she had no dysmorphic features and was normocephalic. Antigravity movements were present in all four limbs but there was overall a very marked paucity of spontaneous movement. She was also hypomimic but appeared visually alert, able to fix and follow, and demonstrated a responsive smile. Truncal tone was low, with head lag, but limb tone was variable: fluctuating generalised dystonia was present but no frank spasticity. Tendon reflexes were brisk but plantar reflexes downgoing. Striatal toe was present. The reported dystonic episodes were characteristic of oculogyric crises (Figure 6.3).

There was no clinical response to levodopa at doses of up to 10 mg/kg/day. Only modest improvements in her dystonia were achieved with trihexyphenidyl, gabapentin, chloral and diazepam. There was some response to clonidine, especially delivered via a transdermal patch, in terms of reduced frequency and severity of exacerbations of dystonia.

Given the clinical presentation of infantile-onset parkinsonism-dystonia, recurrent oculogyric crises, global developmental delay and dysautonomia, there was a strong impression that she was likely to be affected by a monoamine neurotransmitter

disorder, such as AADC deficiency. Surprisingly, CSF monoamine neurotransmitter levels were all individually in the normal range when checked at the age of 18 months, although the ratio of homovanillic acid to 5-hydroxyindoleacetic acid was mildly elevated. This normal result was so unexpected that the test was repeated when she was 24 months old; this time it was entirely normal (**Table 6.1**).

Table 6.1: *Results of CSF neurotransmitter assays in DP35P on two occasions. Normal ranges as per Hyland et al.(287)*

Parameter (normal range)	Result (18 months)	Result (24 months)
5-HIAA (89-367)	138	182
HVA (154-867)	603	499
HVA:5-HIAA (1.0-3.7)	**4.4** ↑	2.7
Dihydrobiopterin (0.4-13.9 nmol/l)	5.7	8.9
Tetrahydrobiopterin (8-57 nmol/l)	33	22
Total neopterin (7-65 nmol/l)	10	17
Pyridoxine (11-64 nmol/l)	38	29
5-MTHF (<2 years: 72-305 nmol/l; 2-5 years: 52-178nmol/l)	74	86

Plasma AADC enzyme activity was also tested and found to be within the normal range. Prolactin was elevated on one occasion but was normal on repeated subsequent testing. Sequencing of a panel of genes associated with childhood-onset movement disorders was normal. Extensive neurometabolic investigations were normal: urine organic acids, glycosaminoglycans and amino acids; blood ammonia, urate, copper, caeruloplasmin, zinc, vitamin B12, folate, creatine phosphokinase, transferrin glycoforms, guanidinoacetate, very long chain amino acids, thyroid function, urea and electrolytes, liver function tests, calcium and full blood count; CSF lactate, glucose and amino acids. Electroencephalography was normal for age.

Magnetic resonance imaging of the brain (at 13 months old) and spine (at 19 months) was wholly unremarkable. Brain MRI was repeated at two years old with magnetic resonance spectroscopy: minimally increased prominence of the ventricular and extraventricular cerebrospinal fluid spaces for age was noted, but there were no major

or specific findings (**Figure 6.3**). A DaTscan (^{123}I-Ioflupane uptake scan) at 23 months old was also normal, indicating normal presynaptic uptake of dopamine (**Figure 6.4**).

Figure 6.2 *Pedigree of DP35 family.*

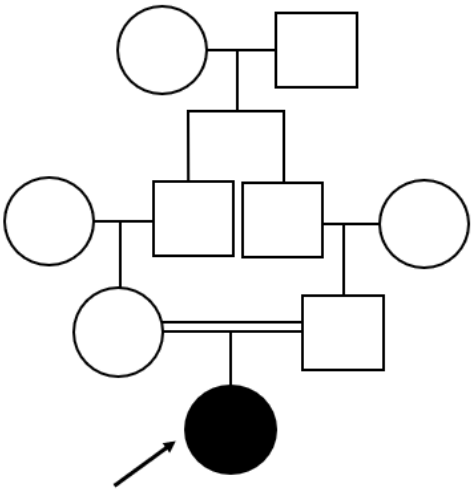
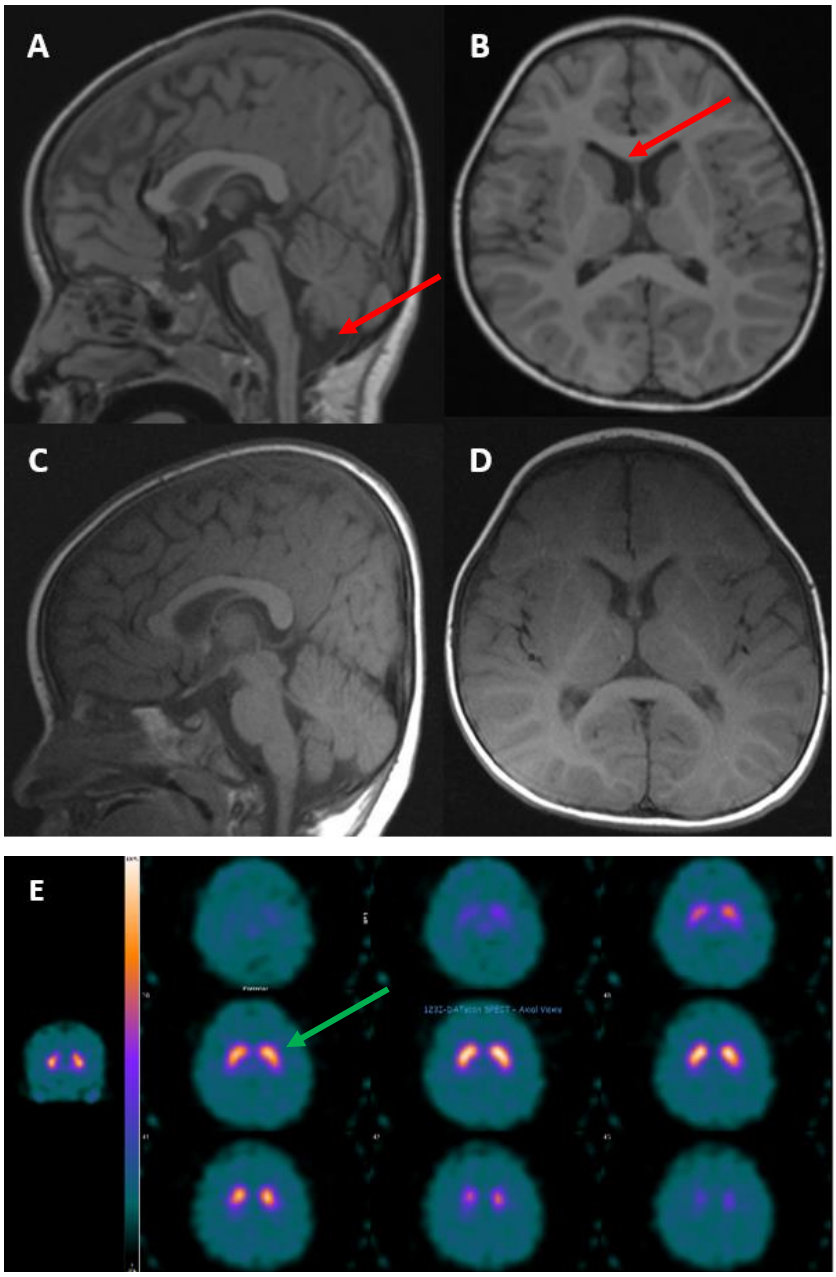


Figure 6.3 *Still images taken from videos of DP35P during exacerbations of dystonia. Note striatal toe (top left); limb posturing, neck arching and dystonic mouth opening (top right) and sustained upwards eye deviation (below).*

Figure 6.4 *Imaging results for DP35P. A-D: T1-weighted MRI images, sagittal (left) and axial (right). A-B : age 36 months; C-D: aged 13 months. Minimally increased prominence of CSF spaces was noted at 36 months (red arrows). E: DaT scan images, showing normal striatal tracer uptake (green arrow).*



6.4 Genetic investigation

As DP35P remained undiagnosed after intensive standard diagnostic investigations, we enrolled the family for research WGS on a triome basis. (Had I been aware then of the existence of VMAT2 deficiency, for which her presentation was an excellent match, I would have requested testing of that gene as a preliminary, but I wasn't – and we did not subsequently find any variants of interest in *SLC18A2*.)

Initial analysis did not reveal any variants likely to be pathogenic in genes known to be associated with neurological disorders. A heterozygous *de novo* variant in *PABPC1* was identified (NM_002568.4:c.388-1G>A, p.?). I felt this was unlikely to explain the proband's phenotype because it occurred at low frequencies in normal population databases (130 occurrences in ExAc,(288) although absent from gnomAD (https://gnomad.broadinstitute.org/)). Missense variants in *PABPC1* have been reported in developmental disorders in a small number of patients, but these are clustered in a different region of the gene and most affected patients appear to have a much milder phenotype characterised by moderate developmental delay but with achievement of ambulation in most cases, and epilepsy in some instances.(289)

Panel-free analysis did not identify any other *de novo* or compound heterozygous variants likely to be relevant. 19 variants which were homozygous in the proband, were predicted to affect the protein either through altering the coding sequence or occurring near a splice site, and had a read depth of ≥20X and a MAF of <1%, were detected. Of these, 12 had a CADD score <15, indicating very low likelihood of pathogenicity. The remaining 7 variants (in *ATXN3, CIC, DRD1, GPR4, MUC16, ZAN* and *ZNF354C*) are detailed in **Table 6.2**.

The variants in *CIC* and *MUC16* could be confidently excluded as candidates for severe childhood-onset disease because both occur in homozygosity in normal population databases including gnomAD. *ZAN* is expressed exclusively in spermatozoa.(290) The *ATXN3* variant reflects the variability of the CAG repeat region of this gene: frameshifts here are common in normal population databases and the insertion length does not approach the threshold for a pathological triplet repeat expansion as found in Machado-Joseph disease (which would also not match the proband's presentation).(291) All the remaining three variants (*DRD1*, *GPR4* and *ZNF354C*) must be considered as possible candidates for pathogenicity but we considered *DRD1* to be the most probable. *ZNF354C* is a ubiquitously expressed zinc finger transcription repressor: it is known to be involved in control of bone development endothelial sprouting but has relatively low brain expression and no known specific contribution to neurological functioning.(292) *GPR4* is a proton-sensing G-protein coupled receptor: overall brain expression is very low although it is expressed in a

specific population of medullary neurons which help regulate breathing.(293) The variant occurs close to the 3' terminus of the gene, where it is postulated that truncating variants may have a reduced impact on protein function.(94)

The *DRD1* variant, NM_000794.4:c.110C>A, p.Thr37Lys, was felt to be a strong candidate for two reasons. Firstly, it appeared likely to impact protein function: it affects a highly conserved amino acid residue, showing species conservation to zebrafish, and all *in silico* tools used predicted a damaging effect (**Table 6.2**). It is absent from normal population databases such as gnomAD. Secondly, the physiological function of the gene product, D_1, was highly relevant to the proband's presenting symptoms, which before any genetic analysis was undertaken had already been thought to reflect a deficiency of dopaminergic function. (I had written to one of my supervisors, Dr Katy Barwick, before any genomic data was available, that I believed this patient had a defect affecting a dopaminergic receptor. I did this because I erroneously believed at the time it would count as a point in favour of classifying the variant as pathogenic, were one to be found: as I discuss in Chapter 2, I continue to think that variant classification systems inadequately capture and weight pre-test probability.)

Sanger sequencing of the *DRD1* p.Thr37Lys variant confirmed that it was homozygous in the proband and heterozygous in both parents (**Figure 6.5**).

Figure 6.5: *Sanger sequencing results showing the c.110C>A variant (highlighted in blue) heterozygous in both parents and homozygous in the proband.*

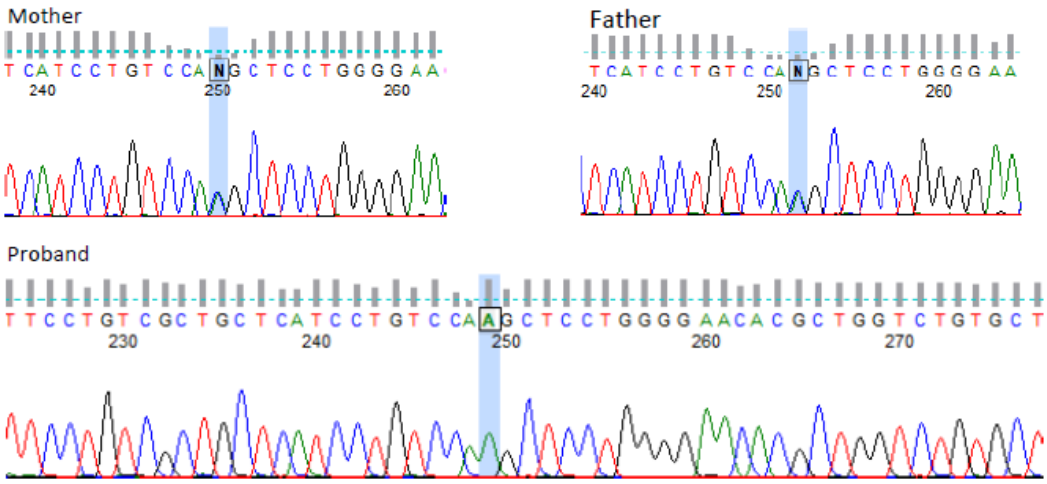


Table 6.2: *Characteristics of homozygous candidate variants identified in DP35P.*

Gene Transcript	Variant	Frequency (homozygote count)	CADD	PolyPhen2 (HumVar)	SIFT	Mutation Taster (v2021)	Function	Expression	Disease phenotype (inheritance)	Notes
ATXN3 NM_004993.6	c.915_916ins17 p.G306Qfs*31	Absent	19.2	n/a	n/a	Benign (82|118)	Deubiquitinating enzyme	Ubiquitous	Machado-Joseph disease, MIM109150 (AD)	Adult-onset cerebellar degeneration; triplet repeat disease
CIC NM_001386298.1	c.4394C>A p.T1465N	0.048% (1)	21.9	Benign (0.021)	Tolerated (0.18)	Benign (14|86)	Transcriptional repressor	Ubiquitous	Intellectual developmental disorder, MIM617600 (AD)	Only truncating variants reported as pathogenic
DRD1 NM_000794.4	c.110C>A p.T37K	Absent	27.5	Probably damaging (0.99)	Deleterious (0.02)	Deleterious (89|11)	Dopamine receptor	Brain, especially basal ganglia	Nil	See text (section 6.4)
GPR4 NM_005282.3	c.864_871 delinsACGT p.V289Rfs*49	Absent	n/a	n/a	n/a	Deleterious (168|32)	Proton-sensing GPCR; role in breathing control	Highest in adipose, kidney, lung, thyroid; low in brain	Nil	Close to 3' terminus
MUC16 NM_024690.2	C19511C>A p.S6504Y	0.39% (3)	15.64	Unavailable	n/a	Benign (0|100)	Mucus component	Mucosal membranes	Nil	.
ZAN NM_003386.3	c.5768del p.C1923Ffs*18	Absent	33	n/a	n/a	Deleterious (165|35)	Allows sperm to bind zona pellucida	Testis only	Nil	Expressed exclusively in sperm
ZFN354C NM_014594.2	c.55G>A p.V19M	0.0011% (0)	23.5	Probably damaging (0.968)	Deleterious (0)	Benign (8|92)	Transcriptional regulator: involved in vascular and bone development	Ubiquitous	Nil	.

We looked for possible additional cases by interrogating DECIPHER (the Deciphering Developmental Disorders database)(294) and the 100,000 Genomes Study data(48) through the Genomics England research portal. We also asked collaborators who maintain specialist genomic databases of patients with undiagnosed neurological disorders to check their data (specifically Dr Niccolò Mencacci, who collaborates with the Carlo Besta Neurological Institute in Milan and Dr Michael Zech at the Human Genetics Institute at Munich) but no additional cases were found. This suggests that (if the association is confirmed to exist) autosomal recessive illness related to *DRD1* deficiency is extremely rare.

6.5 Investigating the effect of the variant on gene expression and protein localisation

Having identified that DP35P had a homozygous *DRD1* variant which was novel and predicted to be damaging, and that her clinical presentation would fit well with a defect of the D_1 receptor, the next step was to test whether the variant did actually affect protein function, and if so, how.

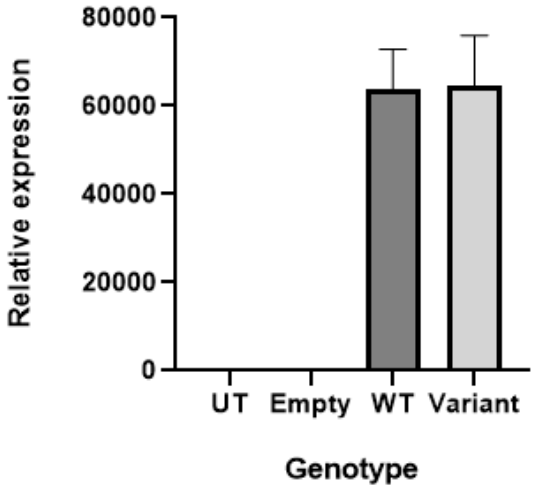
A genetic variant may disrupt a protein at any point along the complex chain of transcription, translation, folding and formation of secondary/tertiary/quaternary structure, maturation, trafficking, and function. We sought first to check whether the p.Thr37Lys variant reduced either gene transcription or translation. Dr Reid prepared an overexpression model of *DRD1* in HEK293T cells as described in Chapter 2, using a vector containing the wildtype (WT) sequence, the variant sequence or with no added sequence (“empty vector”).

To check whether the variant interfered with transcription and/or induced high levels of nonsense-mediated decay, I used qRT PCR to measure levels of mRNA, as described in Chapter 2. There was no significant difference between *DRD1* mRNA levels (normalised to levels of *GAPDH*, a ubiquitously expressed housekeeping gene) in WT or variant expressing cells. Reassuringly, the levels in untransfected cells, or cells transfected with an empty vector, were extremely low (**Figure 6.6**).

Having confirmed that there was no impact on transcription, the next stage was to check whether protein expression was reduced, which could reflect either reduced translation or increased degradation. This was done by Dr Reid using Western blotting: she found no difference in D_1 expression in the overexpression model between cells expressing WT and variant protein. As I did not perform this experiment, I will not present these results in detail.

Next, we looked at whether the protein was undergoing appropriate subcellular localisation – that is, to the cell surface membrane. To do this, we used two independent tests. Dr Reid performed biotinylation studies (i.e. measuring the proportion of the protein which is biotinylated, indicating that it is on the cell surface): these found no significant difference between WT and variant protein. I undertook immunofluorescence using a primary antibody (mouse anti-Myc IgG) which bound the Myc tag co-expressed from our plasmid with *DRD1*, and a secondary antibody (goat anti-mouse) conjugated with Alex Fluor 594, a fluorescent red dye. Cells were labelled with DAPI, a nuclear stain, and WGA, a cell-surface stain. I then looked for colocalization between D_1 (red) and surface membranes (green) (**Figure 6.7**). This showed that D_1 appeared to reach the cell surface membrane appropriately in both WT and variant-expressing cells, with no obvious visual differences between the two, meaning abnormal trafficking to the cell surface was not a likely mechanism for pathogenicity.

Figure 6.6: *qRT PCR results showing* DRD1 *expression (normalised to* GAPDH*) between HEK293T cells untransfected (UT), transfected using an empty vector, or transfected with WT or variant* DRD1.



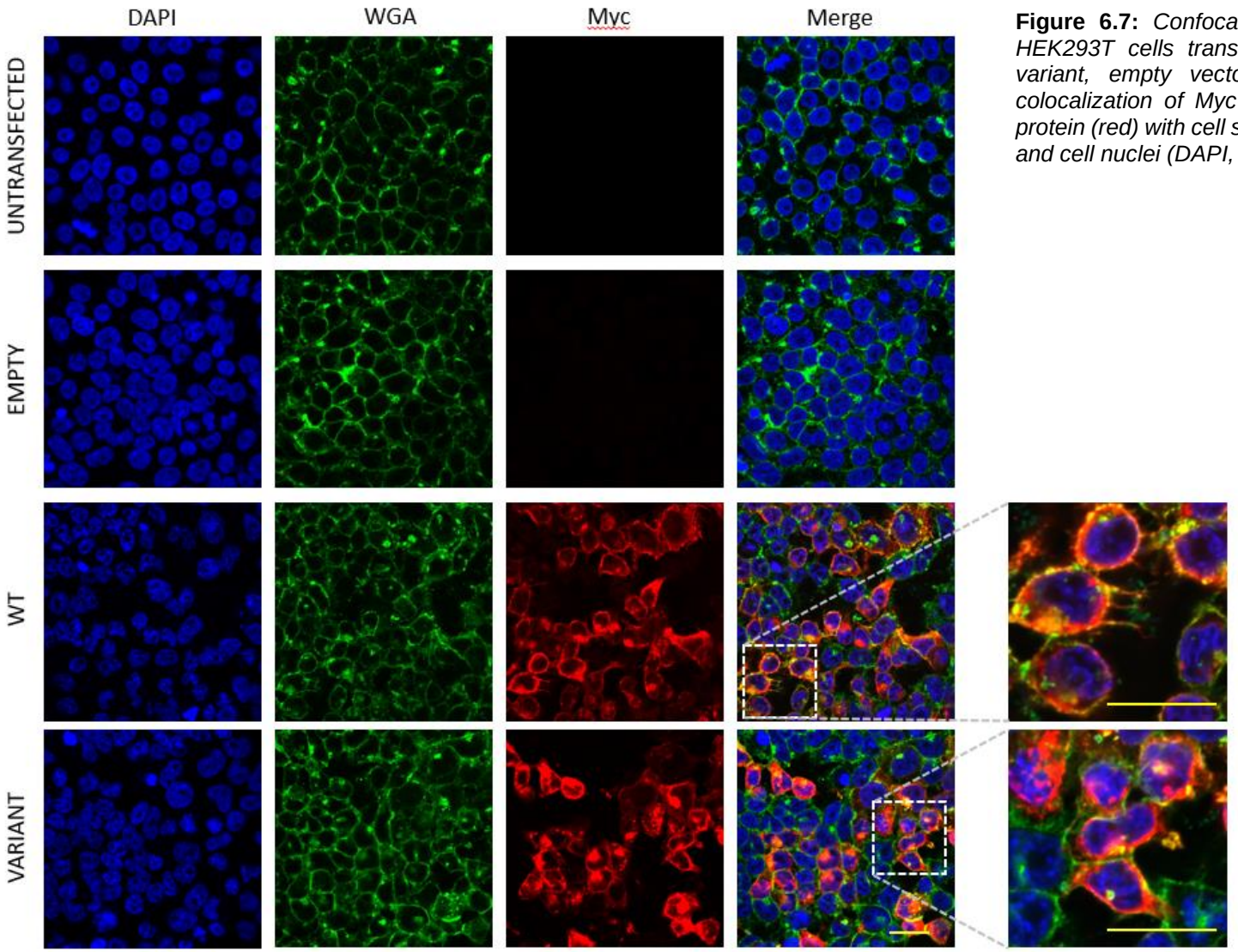
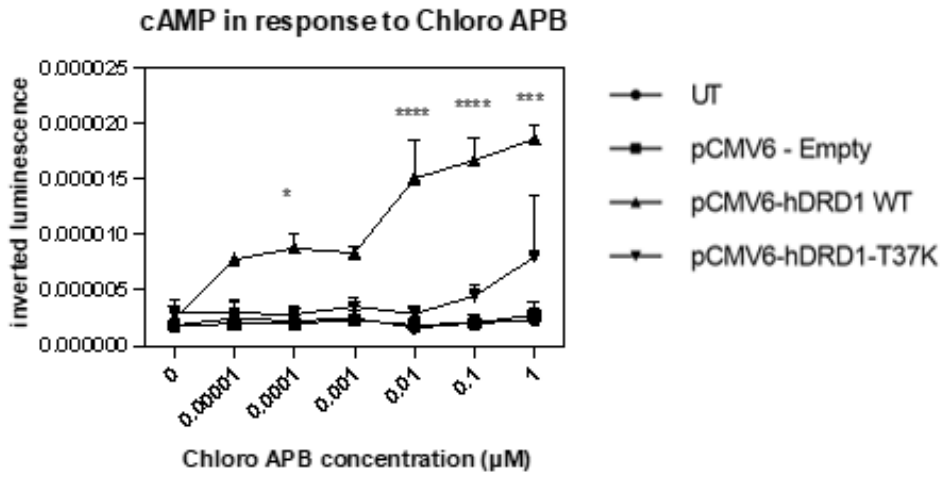


Figure 6.7: *Confocal microscopy images (X63) of HEK293T cells transfected with DRD1-WT, DRD1-variant, empty vector or un-transfected, showing colocalization of Myc-tagged wild-type or variant D_1 protein (red) with cell surface membrane marker, WGA and cell nuclei (DAPI, green) . Scale bar = 20μm.*

6.6 Effect of the variant on receptor function

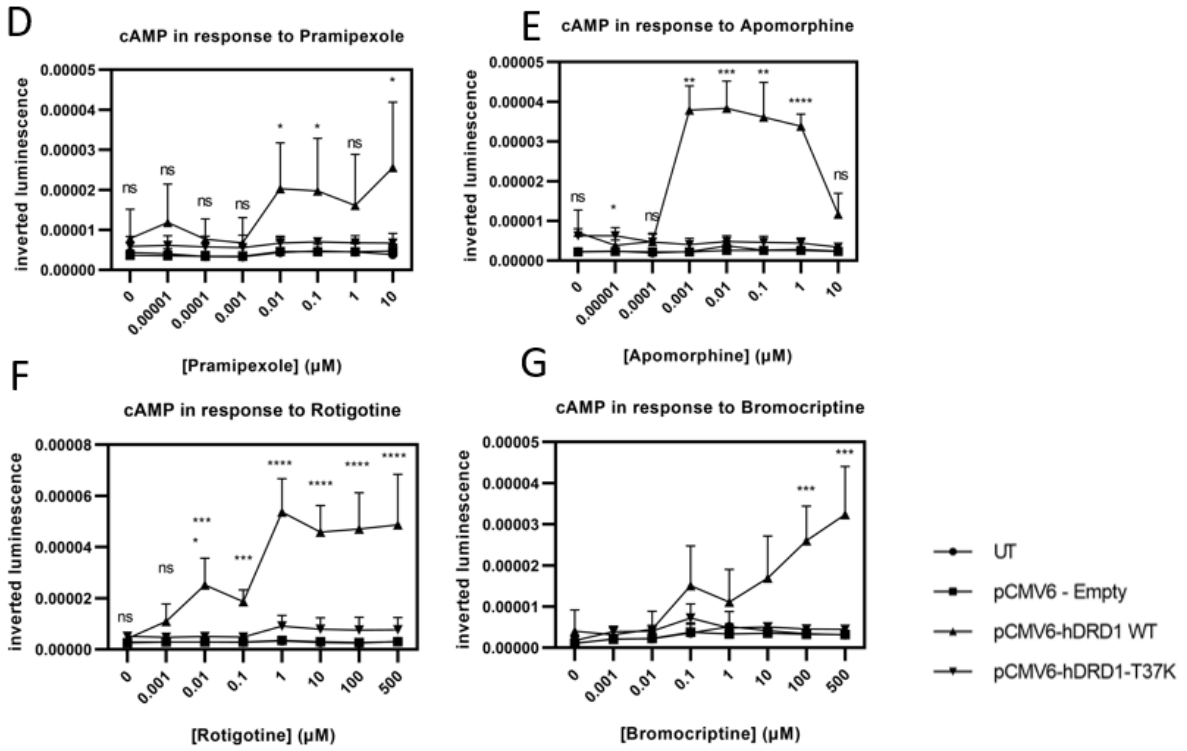
The experiments in this section were undertaken by Dr Reid, but I will briefly summarise her results as they are essential to the conclusions we drew. Using the overexpression model, Dr Reid exposed cells transfected with *DRD1*-WT, *DRD1*-variant, an empty vector or untransfected to the D_1 agonist Chloro APB and measured the resulting change in intracellular cAMP levels using a luminescence-based assay as described in Chapter 2. As discussed in Chapter 1, agonist binding at the D_1 receptor is expected to result in increased intracellular cAMP. As shown in **Figure 6.8**, cAMP increased in *DRD1*-WT expressing cells from a concentration of 0.00001µM Chloro-APB upwards, but a detectable response was only seen in the *DRD1*-variant expressing cells at a concentration 10,000-fold greater, indicating grossly reduced receptor sensitivity.

Figure 6.8: *cAMP changes in response to exposure to varying levels of Chloro APB, a D_1 agonist. An inverted luminescence scale is shown for clarity because in the assay used, increased cAMP levels result in* lower *levels of luminescence. pCMV6 is the vector name. T37K: p.Thr37Lys variant; UT: untransfected; WT: wildtype. *: statistically significant difference, p<0.05; ***: <0.005; **** p<0.001. Figure prepared by Dr Reid.*



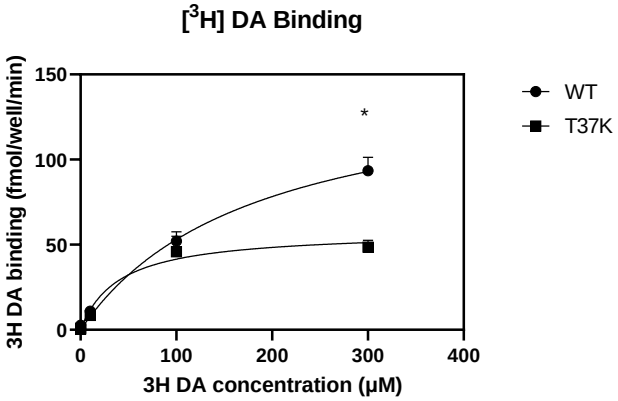
To see whether the receptor could be stimulated by other agonists, Dr Reid exposed it to a range of agents. We chose clinically relevant dopaminergic agonists. Unfortunately, none elicited any increase in cAMP levels in the presence of the variant receptor (**Figure 6.9**).

Figure 6.9: *cAMP changes in response to exposure to varying levels of dopaminergic agonists. *: statistically significant difference, p<0.05; **: p<0.01; ***: <0.005; **** p<0.001. Figure prepared by Dr Reid.*



Thus it appeared that the variant receptor was insensitive to a range of (orthosteric) agonists. We wondered whether this was due to impaired ligand binding or to a failure of signal transduction. Molecular modelling, through a collaboration with Prof Maya Topf and her team at the Institute of Structural and Molecular Biology, London, suggested the former but did not rule out the latter: as well as altering the hydrophobic pocket of the ligand binding site, the angle of two of the seven transmembrane helices was predicted to be altered. To answer this question experimentally, Dr Reid performed a ligand binding assay using tritiated dopamine, as described in Chapter 2. This showed that ligand binding was indeed significantly reduced, although apparently less drastically than cAMP response (**Figure 6.10**). This suggests that the impairment of receptor function may result from elements of both reduced ligand binding and an additional contribution from reduced signal transduction.

Figure 6.10: *Binding of tritiated dopamine ([³H] DA) to WT or variant receptors at varying concentrations.*



6.7 Discussion

No gene-disease association identified in a single patient can be considered confirmed, and this rule certainly applies to our putative "D_1 receptor deficiency syndrome". However, in this instance several lines of evidence support the pathogenicity of the biallelic *DRD1* variant.

Firstly, the proband's phenotype of classical infantile parkinsonism-dystonia with oculogyric crises and dysautonomia is strongly suggestive of a dopaminergic deficiency. The D_1 receptor is (among other roles) an essential element of the cortico-basal ganglia-thalamic network, specifically the direct pathway by which the basal ganglia promote motor activity. Loss of function in this pathway would be expected to result in a paucity of spontaneous movement, just as seen in our profoundly parkinsonian proband. The hyperkinetic elements of her condition, such as oculogyric crises, are clinically indistinguishable from those seen in known disorders involving loss of dopaminergic function such as AADC deficiency – although their detailed pathophysiology remains incompletely understood.(295)

This complex motor phenotype is also reflected in animal models of loss of D_1 function. Abnormalities of motor function and/or learning are features of multiple different mouse knockout (KO) models, but their nature varies. In *Drd1* KO mice generated by using homologous recombination techniques to replace 95% of the gene with a neomycin cassette, Xu et al reported reduced growth and increased locomotor activity

(measured by photobeam interruptions, which were more than doubled compared with WT controls).(296) In contrast, two other studies of KO mice created using a similar technique but by an independent group reported severe feeding difficulties in infancy, resulting in death soon after weaning unless special softened feed was provided, with subtle deficits in motor activity in adulthood (reduced rearing activity with an impression of overall reduced activity,(297) or reduced movement initiation measured by time to move from starting point in an arena, together with inability to improve performance over time in a Morris water maze(298)). Another KO, generated in 2005 by similar methods, showed severely reduced and slowed locomotor activity combined with slower improvement in a place learning task.(299) However, in 2014, using the same model, Nakamura et al(300) reported that the *Drd1* KO showed increased spontaneous locomotor activity compared with WT, although it had significant impairments in rotarod and step-wheel tasks. Using an alternative approach, Durieux et al(301) selectively ablated D_1-expressing striatonigral MSNs in mice using a diphtheria toxin receptor-mediated approach and again showed reduced locomotor activity and motor learning deficits on a rotarod task.

Drosophila models of dysregulated dDA1 (Drosophila's D_1 equivalent) function also manifest motor and learning abnormalities, but again there is an inconsistent pattern between different genetic models. Two strains generated either by insertion of transposable elements or gene inversion showed impaired responses to conditioning stimuli, rescued by reinstating dDA1 expression.(302) In contrast, RNA interference strategies to reduce dDA1 expression resulted in increased larval locomotion,(303) whereas a similar approach significantly reduced motor activity in adult flies.(304) Despite these phenotypic differences, mouse and Drosophila models collectively appear to support an essential role for *DRD1* in motor development and motor control.

We have shown that although the variant D_1 protein is synthesised at normal levels and is appropriately trafficked to the cell surface membrane (at least in a non-neuronal human cell model), its response to agonists is grossly reduced. This is at least in part due to impaired ligand binding but may also reflect additional impairments impacting signal transduction and G-protein activation. This could explain the totally absent

cAMP response to agonist exposure until very high concentrations, while ligand binding is more moderately reduced.

The non-responsiveness of our variant D_1 was conserved across five different agonists, four of which (apomorphine, rotigotine, pramipexole and bromocriptine) could have been considered for prescription had there been any reason to think they would help. From the clinical point of view, this is disappointing. All these agonists rely on the orthosteric binding site, so it is unsurprising that their binding may also be disrupted.(305) A number of positive allosteric modulators (PAMs) of dopaminergic receptors are under development,(306) but as far as I am aware no independently-effective allosteric agonists are known. However, given that dopamine binding does not appear to be as thoroughly abolished as receptor activation in our variant, theoretically a trial of PAMs, possibly in combination with a supraphysiological levodopa dose, could be considered.

We were only able to examine the variant receptor in a non-neuronal human cell overexpression model. In real life, D_1 receptors exist as part of a complex synaptic ecosystem: they form heteromers with other receptor molecules including D_2;(307) are modulated by other neurotransmitters including adenosine and endocannabinoids;(308, 309) and participate in positive and negative feedback loops.(310) A more detailed and accurate examination of variant D_1 receptor function would require system which more closely approximates this, for example, using neurons differentiated from patient-derived induced pluripotent stem cells, preferably in an organoid or assembloid model. Alternatively, an animal model (preferably mammalian) would allow investigation of receptor function within a natural system – although the difficulties of interpreting results from the existing mouse models described above highlights the difficulty of comparing movement patterns between humans and non-primate mammals.(311)

Despite our efforts, we have not been able to identify any additional cases of D_1 receptor deficiency to date. *DRD1* is a moderately small gene (the main transcript consists of a single exon of 447 amino acid residues) and much of it is highly conserved at the species level. It is strongly constrained for protein truncating variants

(pLI 0.96; o/e 0.000) and, interestingly, the only truncating variants found in gnomAD occur close to the C-terminus, from residue 361 onwards.(103) This suggests that even heterozygous variants which result in full loss of function of one allele may be subject to significant adverse selection. Given this, and our proband's severe and early-onset phenotype, it is possible that foetuses homozygous for damaging variants have an increase rate of prenatal demise. I will return to the subject of ultra-rare recessive disorders in my final chapter.

In the future, it is possible that this single gene disorder would be a candidate for precision gene replacement therapy, a strategy that has already been employed for other similar disorders of monoamine neurotransmission. Gene replacement therapy for AADC deficiency has recently been approved by the European Medicines Agency(312): stereotactic injection is used to insert an AAV2 viral vector containing human *DDC* into either the putamen(313) or, in an alternative approach currently in clinical trials, the midbrain.(314) In theory, this genetic condition might also be amenable to a similar approach The small size of the *DRD1* gene coding regions (447 amino acid residues, comparable to *DDC* at 480 residues(315)) would make gene replacement using an AAV viral vector a theoretically practical option. Although expression of *DRD1* is less anatomically concentrated than that of *DDC*, it is strongly enriched in the basal ganglia, especially the striatum, which would almost certainly be the preferred target, as restoration of D_1 function there might restore the key direct pathway of the cortico-basal ganglia-thalamic network. This would not, however, be expected to rescue any defects resulting from absence of functional *DRD1* in the cortex and elsewhere.

I would speculate that, given the evidence for an AR loss-of-function disorder of *DRD1* and an AD gain-of-function disorder for *DRD2*, the "mirror image" conditions may also exist, or at least could potentially occur: that is, that a biallelic variant causing loss of function of D_2 would also be expected to cause a parkinsonian disorder, as has already been shown in mice,(316) due to removing inhibitory input from the indirect pathway, while a variant increasing D_1 activation compared with the wildtype would be likely to cause a hyperkinetic condition. The differing roles and expression profiles of the two receptors, however, suggests that the conditions would not be perfect mirror images:

for example, a D_2 deficiency disorder would be likely to include hyperprolactinaemia. As genetic disorder identification continues to improve, it will be interesting to see whether instances of these hypothetical conditions appear in the coming years.

There is good reason to think that DP35P's severe neurodevelopmental and movement disorder is due to a monogenic deficiency of the D_1 receptor and that this may ultimately form part of a "family" of genetic disorders related to aberrant dopaminergic receptor function. This association remains unproved, however, pending the description of additional affected individuals. I am hopeful that publication of our work(317) may prompt other groups to identify more people: this could confirm and expand the genotype-phenotype association, but more importantly it would pave the way to consideration of developing better precision disease-modifying treatments, which could be a realistic option for a single-gene deficiency disorder such as this.

Chapter 7: Discussion

7.1 Introduction

In this thesis so far, I have discussed the results of genomic analysis of a cohort of 168 people with childhood-onset movement disorders. For about one third I was able to make a diagnosis: this is comparable with results from other published cohorts, although for reasons discussed in Chapter 3, I consider the usefulness of direct comparisons limited. More surprisingly, I identified nine patients with previously undescribed genetic disorders (six disorders: four participants had *VPS16*-related dystonia). This equates to a rate of over one participant with an identifiable undescribed disorder per 19 genomes (or triomes) analysed, or one new disorder per 28 genomes.

This rate of new disease gene identification is far higher than I would have anticipated at the start of my project. Much as I would like to ascribe it to my own talents for genomic analysis, this is clearly not the explanation: I will discuss what I think are the reasons for it in section 7.2. I will also look at the reasons why the diagnostic rate is not higher: while a diagnostic rate of one in three is respectable within the parameters of the field, it should be remembered that the index of suspicion for participants in my project having a genetic disorder was very high. My impression is that the large majority probably do have a genetic explanation for their symptoms and therefore, for one reason or another, my analysis (and other people's) is still missing a significant proportion of these.

I consider the detailed delineation of phenotypes of equal importance with the identification of new disease-gene relationship: a genetic label is not, in itself, much use unless it is attached to an accurate idea of its implications. This is closely related to the question of what, in an era of molecular testing, should constitute a "diagnosis", which I will explore in section 7.3.

For patients and families who do receive a genetic diagnosis, the utility is considerable, though not always in straightforward ways. In a setting where the false negative rate

of a test may be over 50% (in that fewer than half of people with likely genetic movement disorders have a molecular diagnosis made after maximal investigation), the utility of the test itself is much less obvious. Design of clinical systems needs to reflect not only how to manage increasing number of patients with rare and ultra-rare disorders, but how to ensure that those left undiagnosed are not disadvantaged. In section 7.4 I will examine how this might be achieved.

Gene identification is most valuable when it forms not the conclusion of a process but the starting point. Ideally, finding the genetic cause of a disorder should be the first step in a pathway of research that runs through understanding of physiology and phenomenology to targeted disease-modifying treatment, or even cure. In my last section, 7.5, I will discuss this process and whether there are ways it could be expedited.

7.2 There's a lot still to be found

My results make it very clear that in-depth genomic analysis, combined with detailed clinical phenotyping, can still have a high yield of new disease genes in rare disease populations. Importantly, this is true even *without* using cutting-edge approaches to genomic analysis and interpretation. Clearly, the yield would be even higher with a more comprehensive approach to analysis.

I started this project as a complete novice in genomic analysis and am keenly aware that, at the end of it, my analytic pipeline still has considerable room for improvement. I have never developed adequate systems for the analysis of copy number variants, structural variants or short tandem repeats: this can be done,(318, 319) but it requires more sophisticated bioinformatic skills than I currently have. I have no accurate way of assessing variants in deep intronic, promoter or other regulatory regions short of undertaking experiments such as splicing assays and qRT PCR: for specific variants of interest I can and have employed these techniques, but it would be impractical to use them to screen a large number of variants.

My results suggest that many unidentified neurogenetic conditions remain so not because they are genetically complex – although this is undoubtedly the case for some – but simply because they are rare. Identifying candidate variants and then case-finding and/or undertaking functional work to prove or disprove pathogenicity remains relatively labour-intensive, although far less so than a decade or two ago. It also requires complex collaboration, both between people with different skillsets (clinical, bioinformatic and wet-lab based) within research groups, and between research groups who may each have identified one individual with an ultra-rare condition.

Technical changes in WGS, notably PCR-free library generation and long-read sequencing,(320, 321) will improve researchers' ability to identify structural and copy number variants within user-friendly analysis pipelines, as too will improvements in bioinformatic techniques and software. As in many other branches of computing, novel analysis techniques tend to start life as specialist applications, usable only by those with specific coding expertise (and Unix operating systems) and gradually become more user-friendly and accessible to a wider pool of researchers, often involving commercial adoption. The software on which I have depended for my analysis, such as Alamut™ Visual Plus and QIAGEN Clinical Insight, are good examples of this. A few years ago, even analysis of SNPs and small indels on a large scale required extensive bioinformatic training. Now it is available to anyone with genetics knowledge and an ability to learn to use software packages.

This is not to say that the process is (yet) close to full automation, or even to delegation to an entirely technical workforce. Integrating detailed clinical information with genomic interpretation remains a challenge to automation, and systems such as HPO terms only partially remedy it.(130) They cannot compare with the richer detail available from an actual clinical assessment. It is also valuable to be able to return to the clinic to obtain specific data that can support interpretation of a variant: for example, when I found a novel *ARID1B* variant in DY175P, at her next appointment we were able to check whether she had hypoplastic fifth digit nails, supporting a diagnosis of Coffin-Siris syndrome.

Prioritisation of candidate variants in novel genes also involves integration of information from a range of sources: variant characteristics; gene constraint; known animal and/or cellular phenotypes; expression patterns; and physiological roles. This means that even a preliminary check of an interesting gene could involve me looking at gnomAD,(103) GTEx,(290) Uniprot,(322) OMIM,(36) the Jax MGI database(323) and PubMed. There are many efforts to bring these types of information together but none of them, as yet, are wholly satisfactory – *vide* the often bizarre disease associations suggested by otherwise useful database sites such as GeneCards.(324)

In section 1.5, I listed reasons why disease genes might evade detection, as follows:

- Digenic or polygenic inheritance
- Incomplete penetrance
- Phenotypic pleiotropy
- Structural and copy number variants
- Intronic variants
- Variants in intergenic regions
- Extreme rarity
- Mosaicism

Neither I nor, to my knowledge, anyone else has undertaken research which would allow quantification of the relative contribution of these factors to the burden of unidentified disorders. However, to some extent, the findings of my project support the relevance of several of them.

VPS16-related dystonia (see Chapter 5) shows **variable penetrance**, which may explain why (though apparently less rare than some other disorders occurring in my cohort) it was not identified before. The participant in whom I most recently made this diagnosis, DY225P, has a heterozygous *VPS16* splice site variant inherited from an asymptomatic parent: had I undertaken their analysis before researching the gene in other families, I would probably have overlooked it.

DRD1-related disorder (see Chapter 6) is so **rare** that, despite extensive interrogation of other databases, we have not yet been able to identify a second case. The only reason that we feel any degree of confidence in attributing pathogenicity to the variant

is the strong pre-test clinical impression that the proband had a deficiency of dopaminergic transmission: without this, we would not have considered proceeding to the considerable time and expense of functional assays.

The neurodevelopmental disorder related to *TNPO2* variants (see Chapter 3) occurred in our participant in a **mosaic** state. (I have a hunch – which I am unable at present to substantiate – that a proportion of our undiagnosed patients with paroxysmal movement disorders may have tissue-specific mosaicism, possibly related to somatic mutations like those found in Sturge-Weber syndrome.(325) This could explain the highly focal nature of many of their symptoms and the fact that a high proportion seem to remain undiagnosed, and without strong candidate variants, despite a high index of suspicion for genetic pathology.)

Disorders related to *H2AC6* (see Chapter 3) appear to be **phenotypically pleiotropic**. At first, when I made contact with a team assembling a case-series of individuals with variants and neurodevelopmental impairments, I felt it was unlikely that our participant's variant could be pathogenic as the phenotypic fit was poor. Only lately, when it became apparent that a subset of patients had a combination of cerebellar atrophy and dystonia without major intellectual disability, did we feel more convinced of the variant's relevance.

Thus instances from my own work support the importance of incomplete penetrance, mosaicism, extreme rarity and phenotypic pleiotropy to the difficulty of identifying disease genes. The fact I did not find instances of pathogenic structural, intergenic or deep intronic variants does not imply that these are lesser contributors: it simply reflects that I lacked the tools within the remit of this PhD to look for them efficiently. (Identification of polygenic inheritance is a different challenge: it would require analysis of larger population datasets.)

Compared to my expectations when I started the project, I now feel that rarity alone is responsible for a higher proportion of unidentified conditions than I had expected: this was a factor for *VPS41* and *INTS11* as well as *DRD1*. As WGS and WES become available to populations who previously had little access to them, especially those

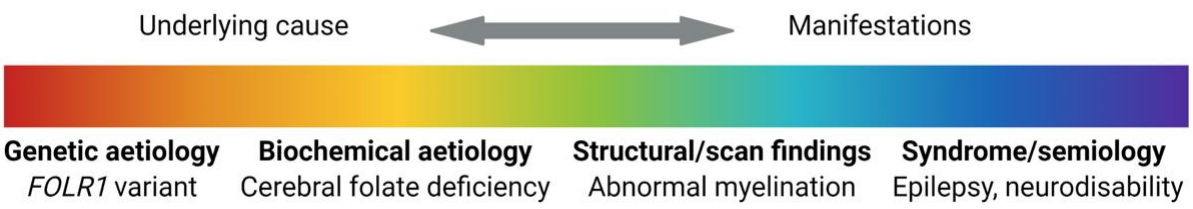
where consanguinity remains common, I suspect identification of ultra-rare AR disorders will accelerate.

7.3 What is a “diagnosis” in the era of molecular testing?

In this thesis I have sometimes referred to “undiagnosed movement disorders”. This term is problematic: the movement disorder is not undiagnosed; its genetic basis is unknown. The distinction may seem pedantic but it is not trivial.

Diagnosis has its own well-established sociology, into which I will not trespass.(326) However, at the simplest level the word can denote an entity positioned virtually anywhere on a spectrum from molecular aetiology to clinical phenomenology **(Figure 7.1)**. Anecdotally, patients and families may conceptualise their diagnosis at any point along the spectrum and are sometimes concerned that the different levels of description imply multiple diagnoses, asking questions such as “Dystonia or a genetic disorder – which is it?”. In the field of epilepsy this complexity has been recognised and formalised in the International League Against Epilepsy’s classification of the epilepsies, which separates diagnosis into three “levels” (seizure type; epilepsy type; epilepsy syndrome) together with acknowledging aetiology and comorbidities as additional aspects of its overall diagnostic framework.(39) Movement disorders do not currently have a corresponding agreed framework, but a similar approach of acknowledging both aetiology and clinical manifestations or syndrome is sometimes used in practice to good effect.

Figure 7.1: *“Diagnosis” can refer to anywhere on a spectrum of descriptions of a disease, as illustrated below using the example of cerebral folate deficiency due to* FOLR1 *variants*



Any system of diagnostic labelling for genetic movement disorders must reflect that fact that there is no 1:1 correspondence between genotype and phenotype: a single

clinically identifiable syndrome may have multiple genetic causes ("genetic heterogeneity"), while variants in a single gene may result in a range of different phenotypes ("phenotypic pleiotropy"). Some current naming conventions fail to reflect this. For example, the OMIM designation "developmental and epileptic encephalopathy 64" may poorly describe a person with *RHOBTB2*-related disorder who has neurodevelopmental impairment and a movement disorder but no epilepsy. "Alternating hemiplegia of childhood" is a useful descriptive label but is sometimes interpreted as implying the presence of an *ATP1A3* or possibly *ATP1A2* variant, whereas the clinical criteria may sometimes also be met by people whose conditions are due to *RHOBTB2* variants,(87) *JPH3* haploinsufficiency,(197) GLUT1 deficiency,(327) *TBC1D24* variants,(328) or the *CACNA1A* spectrum.(329) These overlaps can be confusing both for families and for non-specialist health professionals, and careful use of diagnostic labels is required to avoid misunderstandings. This is one reason why accurate delineation of the phenotypic spectrum associated with variants in particular genes is important: it helps clinicians both to understand what the implications of a genetic finding are for the patient, and to identify which clinical findings can, or cannot, be ascribed to a genetic diagnosis – that is, where it may be necessary to seek an additional "label". For example, one of the patients described in our *SLC30A9* case series has just been referred for investigations into the aetiology of his hearing impairment: in my view this is probably unnecessary as it is likely to reflect his Birk-Landau-Perez syndrome, and hopefully our description of the phenotype will help avoid unnecessary tests for future patients.

"Plain-language" description has a great deal to recommend it as an approach to recording phenomenology, but for families it can be less effective as a way of concisely conveying a medical issue which will be recognised and accommodated by the people and services around them. I have taken to using encouraging families to use "medical"-sounding labels, preferably Latinate or Greek in origin, as this seems to evoke a more positive response in terms of educational and financial support. It would be an interesting experiment to submit two applications for Education, Health and Care Plans and/or Disability Living Allowance, differing only in that where on used "intermittent involuntary movements", the other would read "paroxysmal non-kinesigenic dyskinesia". Although I have not had a chance to test my hypothesis, I predict that the

latter would have a better success rate, despite the two terms being essentially synonymous.

One aspect of the ontology of diagnosis which is specific to movement disorders is the question of what is, and what is not, cerebral palsy. There is, as yet, no consensus on the answer. A range of terminological options CP-like conditions with a suspected genetic basis have fallen in and out of use over time **(Table 7.1)**. Although many genetic disorders, if they are not progressive, do technically meet Bax et al's definition of CP ("a group of disorders of the development of movement and posture….attributed to non-progressive disturbances that occurred in the developing fetal or infant brain"(330)), this usage is far from universal. CP has sometimes functioned as a label of last resort: conditions may be included within it when no convincing alternative explanation presents itself, but removed into separate categories as their aetiology becomes better understood. For example, twenty years ago "ataxic cerebral palsy" was described as accounting for up to 10% of all CP,(331) but since then the category has gradually shrunk as various genetic syndromes presenting with infant-onset ataxia have been described: many of these are no longer habitually referred to as CP.

Table 7.1: *First occurrences on PubMed of various terms used to describe cerebral-palsy-like disorders of genetic or suspected genetic origin(137, 332-335)*

Term	First PubMed occurrence
Hereditary CP	Blumel et al, 1957
Idiopathic CP	Hutton et al, 1994
Atypical CP	Nygaard et al, 1994
Mimic of CP	Gucuyener et al, 1998
Cryptogenic CP	Segel et al, 2015
Genetic CP	*No occurrences*

Retaining the term "cerebral palsy" as part of the diagnostic label for people with a genetic disorder of motor function has the important advantage that there is evidence that it is preferred by patients, since it is well understood and may facilitate access to services and support.(336) This question evokes the "lumpers and splitters" debate, which dates back to Darwin(337) if not earlier: diagnostic precision (such as identifying a specific genetic aetiology) clearly has advantages, but there are also advantages to

allowing patients to retain “membership” of a larger, better-known and arguably better-served category. My own preference is therefore to use “genetic CP” for people with non-progressive disorders meeting the Bax definition, and “genetic mimics of CP” for conditions which appear similar but are progressive, as this would violate a key part of the definition. Interestingly, however, searching the phrase “genetic cerebral palsy” in PubMed gives no exact results at all.

7.4 Reaping benefits and minimising harms

For the successful incorporation of genomics into paediatric neurology practice, our systems will need to evolve. Clinical Genetics services are quite obviously not resourced to deal with the growing demand for genetic testing, even in high income countries such as the UK: if plans to introduce WGS as a neonatal screening tool(338) come to fruition, this will be greatly exacerbated. Levels of competence and confidence in requesting and understanding genetic results among other professional groups, including paediatric neurologists, vary.(339) The technological improvements which allow WGS to be widely implemented will only deliver real benefits if they are accompanied by development of adequate capacity for interpretation and discussion of results. This will require either a large and rapid expansion of Clinical Genetics services, or training and resourcing non-specialist clinicians to integrate genomic testing (including appropriate requesting, interpretation of ambiguous results, and discussions with families) into the care they provide. Even if counterbalanced by savings in other forms of diagnostics, neither option will be cheap.

In Chapter 1, I discussed the benefits of receiving a genetic diagnosis: access to specific treatments if they exist; access to research; genetic counselling; accurate prognostication; ability to join condition-specific support networks; and an end to invasive diagnostic testing. This is all very well for the patients who *do* receive a diagnosis, but there are at least two important problems.

Firstly, many patients with genetic movement disorders never do receive a molecular diagnosis. Remaining (on the genetic level) “undiagnosed” can bring emotional and practical difficulties for families. To quote SWAN UK (“Syndrome Without A Name”, an

advocacy group for families of children with undiagnosed rare disorders), "*Having no diagnosis leaves families in limbo. Children with syndromes without a name are impossible to track in European healthcare systems. Governments do not keep records of people who have not been given a confirmed diagnosis. This means they can have difficulties accessing appropriate health care and support."*(340) Although rates of diagnosis are increasing steadily, they will not reach 100% any time in the foreseeable future and meanwhile it is vital that those left undiagnosed do not become "second class citizens" in healthcare.

Secondly, the benefits of diagnosis are attenuated if you are among the first people to be diagnosed with your disorder. Several of the key benefits of diagnosis – especially knowing what the future holds, and connection with other families – may be unavailable. Having said that, family-led support groups do now start to form very rapidly: I am aware of at least two support groups started for genes identified since the beginning of my project: The TNPO2 Foundation (www.tnpo2.org); and a Bryant-Li-Bhoj neurodevelopmental syndrome Facebook group. For those who have no disease-specific advocacy group, umbrella organisations such as Genetic Alliance(341) and RareConnect(342) play a useful role as a clearing-house, but they cannot provide all the support that an established condition support group can offer.

That curious phrase, "diagnostic odyssey", is perhaps relevant here. The original Odyssey was notable for lasting many years and involving a series of mostly unpleasant and frightening encounters, a reasonable analogue for a prolonged process of investigation, sometimes painful and invasive. We should remember, though, that when Odysseus finally reached home he found not a peaceful resting place but further threats, doubt and unwelcome surprises. To receive a genetic diagnosis about which virtually nothing is yet known is a mixed blessing: no prognosis, little support, and limited understanding from other medical professionals and services.

Alongside developing health systems to cope with increased use of broad-spectrum genetic testing, then, we must consider how best to support people diagnosed with ultra-rare disorders, and their families. No clinician can now reasonably aim to "know" every genetic disorder, and for many conditions patients must expect to be the only

one ever encountered by their paediatrician and even their neurologist or geneticist. There have been some successful instances of centralisation of care for rare disorders in one or a few national centres, such as the national ataxia-telangiectasia clinic and the highly specialised service for mitochondrial disorders. In some cases, though, the interface between such highly specialised centres and the local and regional services on which families must still depend for much of their care remains challenging. As the number of identified ultra-rare diseases, and people affected by them, continues to climb, we must think about systems which are geared not only to diagnosis but also towards integration of specialist expertise with local and generalist services.

Increased genomic testing has benefits for patients but also for research and, just as for individual patients, these will only be fully realised once our systems adapt to take advantage of them. Since the inception of genomics, freely available online databases, often with multiple contributors, have been vital to the field. I would cite here gnomAD,(103) Ensembl,(343) Uniprot,(315) ClinVar(344) and OMIM(36) but there are many more. These allow ready comparison of sequence and protein data, variants and reported phenotypes. Tools such as GeneMatcher,(124) which allows researchers to make contact with others interested in the same gene – essential in ultra-rare disease research – are also a hugely important resource.

Sharing repositories of actual genomic data, however, presents grave difficulties for both practical and regulatory reasons. There is currently no straightforward way to access data linked to identifiable patients from new NHS diagnostic genomics service for clinical reanalysis, let alone research: I am told (personal communication) that this reflects a lack of resources to support constructive and appropriate use by teams with varying experience. The 100,000 Genomes dataset is more accessible, but at the time of writing I am several months into the process of applying to review (for diagnostic purposes) the data of patients enrolled outside my own hospital. This leads to unnecessary duplication of effort and expenditure. Various projects, including the those led by healthcare agencies (100,000 Genomes Research Environment,(48) from Genomics England), research bodies (NIHR Bioresource(345)) a private enterprises (RARE-X(346)) are working on this, but access to patient-level data remains, understandably, challenging. As WGS becomes cheaper, families may

increasingly consider paying for, and taking ownership of, their own analyses: the positive and negative consequences of this will unfold over the next few years.

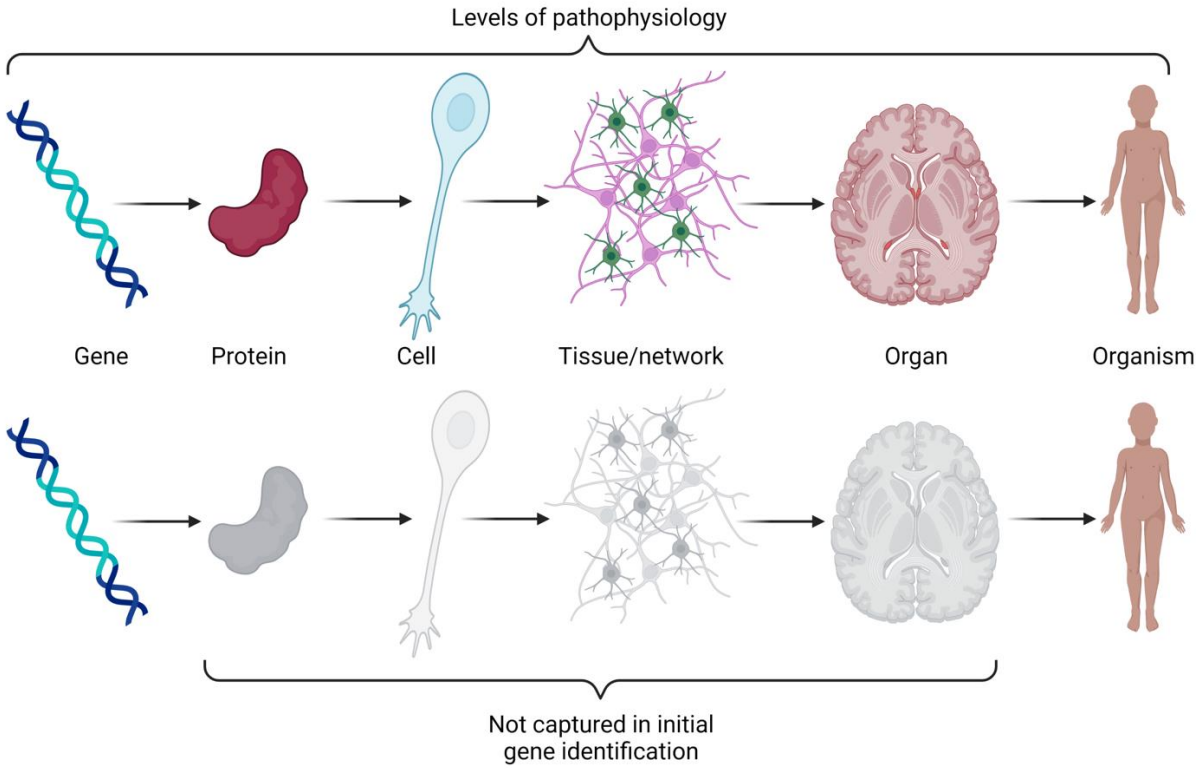
7.5 Gene identification is a beginning, not an end

Gene identification has always existed within a wider framework of research into understanding genetic diseases. Although it does have value in itself, particularly allowing genetic counselling and testing, it is more importantly the starting point for other research.

This research can proceed in any of several directions: pathophysiological; phenomenological; and therapeutic. Pathophysiological research aims to understand the mechanism by which a genetic defect exerts its effects. A genetic disorder can be envisaged as a chain of causation from a nucleic-acid level defect through a dysfunctional protein, changes in cell behaviour, impacts on the tissue or (in the case of neurons) network in which the affected cell participates, organ-level effects and ultimately the symptoms experienced by the patient **(Figure 7.2)**. Of these, only the two extremes – the gene-level defect and the whole-patient presentation – are captured in the initial "gene identification" stage, leaving all the intermediate levels still to be explored. Which level is prioritised for research will depend on many factors, including predictions of where therapeutic targets are likely to be found and the skillset of the researchers.

(In a few situations it may theoretically be an option to bypass a full understanding of pathophysiology and proceed directly to developing therapies. This could apply where, for example, a disease is clearly due to loss of function of a specific gene in a particular tissue: if it is possible artificially to replace the gene or its product, it may be justifiable to trial this approach even if the gene product's physiological role remains incompletely understood. Generally however, it will be preferable – even from the strictly utilitarian point of view, leaving aside scientific curiosity – to seek to understand the disease process. Even where the route to a therapy appears clear-cut, establishing biomarkers and identifying possible toxicity risks, such as the impact of gene overexpression, are invaluable.)

Figure 7.2: *Pathophysiology of a genetic condition can be described at many levels from the gene to the whole organism and in between. While "gene identification" captures the two extremes, it often gives no information regarding the intermediate levels.*



Phenomenological research involves mapping the symptomatology, semiology and natural history of a condition. This allow clinicians to recognise and diagnose the disorder, to be aware what care and monitoring may be needed, and to give accurate prognostic information to patients and families. Importantly, it also facilitates future clinical trials. Placebo-controlled designs can be problematic for trials involving rare, severe disorders, both because of the challenge of reaching adequate statistical power in very small populations and the ethical difficulty of withholding a potentially beneficial treatment in a situation where non-treatment is always associated with a bad outcome. Use of a historical control, although not without difficulties, can mitigate these issues(347): the better the quality of data on a condition's natural history and the more robust the biomarkers, including markers of progression, the more satisfactory a historical control can be. "Trial readiness" is increasingly recognised as an important pre-therapeutic target for rare disease populations.(348)

Therapeutic research is the ultimate aim of most study of genetic diseases. I have grown to dislike the phrase "personalised medicine": everyone's medical care should

be personalised to many features of the both the person and the disorder, not just with reference to a molecular-level diagnosis. Nor is it the case that a therapy guided by molecular understanding of the disease process will necessarily be superior to one that is not. Nevertheless, identifying the genetic cause of a disorder does undeniably open up new therapeutic possibilities. These can be divided into three categories: rational physiology-based treatment choice; empirical identification of effective options; and design of new precision medicines.

Firstly, clinicians' choice from the existing menu of therapeutic options can be guided by their knowledge of the underlying gene or variant and its pathophysiological implications. Examples could include choosing a sodium channel blocking antiseizure agent such as carbamazepine for epilepsy associated with a gain-of-function *SCN1A* variant, or rejecting it in Dravet syndrome,(186) using levodopa for dopa-responsive dystonia,(24) following the ketogenic diet for conditions in the GLUT1 deficiency spectrum (59) or the use of metal chelation in hypermanganesaemia syndromes.(349)

Secondly, it may become clear serendipitously that a particular approach works well for a particular genetic aetiology, whether or not the therapeutic mechanism is understood. For example, dystonia related to *KMT2B* variants appears to respond particularly well to deep brain stimulation(179) – an observation none the less clinically valuable because it remains obscure why this should be so. Another example is caffeine for the treatment of dyskinesia related to gain-of-function variants in *ADCY5*, the effectiveness of which was first noted by members of an affected family who found their symptoms improved on drinking coffee.(350) In this instance, a plausible mechanism for caffeine's therapeutic effects is known – caffeine is an A2A receptor antagonist, and A2A stimulates the neuronal adenylate cyclase encoded by *ADCY5* – but this was worked out *post hoc*, after the initial observation of its efficacy.(25)

Thirdly, a new treatment can be designed to target a specific genetic fault, whether actually at the genetic level or elsewhere along the pathophysiological chain. This category includes "designer" small molecules, enzyme replacement therapy (ERT) and nucleic acid therapies.

A relatively small number of specifically designed small molecules have reached clinical trials for childhood movement disorders, including a vitamin B5 derivative in pantothenate kinase associated neurodegeneration.(351) Miglustat, an inhibitor of glycosphingolipid synthesis used in Niemann-Pick disease type C, though not technically a ‘designer molecule’, also falls within this category.(352)

ERT is more established in treatment of childhood neurological disorders and is used for a number of lysosomal storage disorders. More recently cerliponase alfa, an analogue of tripeptidyl peptidase 1, has been approved as a treatment for neuronal ceroid lipofuscinosis 2, in which it can significantly slow disease progression.(26) Difficulties include the need for ongoing treatment at regular intervals, and the inability of enzymes to cross the blood-brain barrier, which necessitates intraventricular infusions via a ventricular access device. While ERT can have transformative benefits, it is far from a cure.

Nucleic acid therapies can be divided into three main categories: gene replacement, gene editing and RNA-based approaches. Gene replacement approaches aim to supply a healthy copy of a defective gene to the tissue where it is needed using a vector such as virus or lipid nanoparticle. It is best suited to diseases which are:

- Caused by loss of function of the gene product;
- Amenable to treatment of specific sites or tissues, rather than the whole body or brain and
- Caused by genes which are small enough that they can be carried by a virus or other suitable vector.

In the field of paediatric movement disorders, the condition for which gene replacement therapy is furthest advanced is AADC deficiency, which meets all these criteria. A healthy copy of the human *DDC* gene, encoding AADC, is inserted into a plasmid which is taken up by a viral vector (adeno-associated virus serotype 2, or AAV2). Stereotactic neurosurgical techniques are used to inject these into a site where endogenous dopamine synthesis should happen: the midbrain and the putamen have both been targeted by different research teams (55, 314). Because the virus is taken

up by post-mitotic cells, stable expression of the transfected gene can be achieved over a period of years (353). The results, though not curative, have been impressive, with marked reduction or elimination of oculogyric crises and significant neurodevelopmental gains.

Another monogenic neurotransmitter disorder for which gene therapy is in development is dopamine transporter deficiency syndrome, which is caused by autosomal recessive variants in *SLC6A3* and is another cause of severe infantile dystonia-parkinsonism (354). The proposed technique is similar to that used in AADC deficiency: stereotactic delivery of and AAV2 vector containing a healthy copy of the gene to the midbrain. This has already achieved marked improvements in a knockout mouse model (355) and clinical trials are planned to start in the near future.

When a disease results from absence or inadequacy of a specific protein, gene replacement is an obvious potential approach, but often this is not the case. If pathogenesis involves an abnormal or toxic protein, increasing production of the healthy form may not be helpful, whereas gene editing may allow correction of the faulty protein. Currently the most widely used gene-editing system is CRISPR-Cas9, which uses a synthetic RNA guide to direct the Cas9 nuclease to a specific point in the DNA. This process has to a large extent supplanted other methods which required creation of a whole synthetic nuclease protein for each target sequence, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). (356) Gene editing may enable modification of, for example, the gene for alpha-synuclein, which is implicated in the damage to dopaminergic neurons in PD.(357) Gene editing is already in clinical use (though not yet for movement disorders) in the form in *ex vivo* techniques. These involve removal of patient cells, most commonly bone marrow, which are then edited in a laboratory, and returned to the patient. *In vivo* techniques, where the editing agent is administered directly to the patient using a viral vector or lipid nanoparticle delivery system, have now entered clinical trials.(358, 359)

An alternative family of approaches involve the use of RNA. RNA is a versatile molecule with several distinct physiological roles, and accordingly several different

approaches have been explored, involving small interfering RNA (siRNA), small activating RNA (saRNA), microRNA (miRNA) and antisense oligonucleotides (ASOs).

siRNAs are short double-stranded RNA molecules which bind to specific mRNA sequences and prevent them being translated, thereby silencing expression of the corresponding gene.(360) saRNAs, by contrast, target gene promoter regions to enhance expression.(361) miRNAs include naturally occurring molecules: dysregulation of miRNA has been identified in an increasing range of neurological disorders including Parkinson's disease (PD) and perinatal brain injury, (362, 363) meaning that unlike siRNA or saRNA they are a potential target as well as a treatment. They are also under investigation as a therapeutic tool, however: artificial miRNAs, which unlike siRNAs are delivered to cells by viral vectors and expressed there under tissue-specific promoters, may enable longer-lasting gene silencing due to ongoing expression.(364)

Antisense oligonucleotides (ASOs) are short synthetic single-stranded RNA-like molecules, usually with a chemically modified backbone to improve their stability, which can alter the fate of mRNA in any of several ways: triggering its degradation; stopping translation; blocking mRNA binding proteins; modifying splicing, or increasing translation by blockading upstream inhibitory regions.(365) They are already in clinical use for conditions including spinal muscular atrophy and are in clinical trials for forms of PD related to mutations in specific genes (*LRRK2* and *GBA1*).(366, 367) However, a phase 3 trial of tominersen, an ASO which aimed to treat Huntington's disease by inducing degradation of huntingtin mRNA, was stopped in 2021 due to adverse reactions and lack of efficacy,(368) illustrating the potential risks of off-target effects in this type of therapy even where early results may seem promising.

Genetic and nucleotide-based therapies offer ground for optimism in a range of conditions, but they do have limitations. The problem of tissue targeting is a major one: for example, in gene replacement therapy for AADC deficiency, only a single anatomical site is targeted. This restores dopamine synthesis in key basal ganglia circuits but does not restore the dopaminergic systems of the brain (or the enteric nervous system) in full, and 5-HT synthesis remains grossly deficient, with unknown

implications for mood and behaviour in the longer term. AADC deficiency, though, provides a comparatively “easy” target: for many movement disorders, and most neurodevelopmental disorders, there is a much less clear anatomical site for treatment: whole-brain targeting would be required. With current vector technology, that remains out of reach for now – although advances such as capsid modification,(369) peptide shuttle systems to overcome the blood-brain barrier,(370) and delivery via novel sites such as the cisterna magna(371) offer hopeful prospects.

Another serious issue is the timing of treatment: even a biochemically curative therapy for a severe neurodevelopmental disorder cannot hope to reverse the impact of lost years (or even months) of normal development prior to treatment. There may be only a narrow chronological therapeutic window during which treatment can minimise permanent developmental impacts. Here, early diagnosis and accurate knowledge of the natural history of movement disorders will be vital in allowing us to make the best use of novel therapies.

Detailed study of every rare disease is limited by resources, but rare disorders research may increasingly offer synergistic benefits for other rare disorders. For example, gene therapy for AADC deficiency, itself an ultra-rare disorder, paves the way for clinical trials of gene therapy for even rarer conditions using some of the same expertise. Once such a pathway is well-established, many aspects, including vector design, manufacturing and surgical techniques, may be generalisable to other conditions. The marginal cost of producing a gene replacement therapy for an amenable novel condition might therefore fall dramatically. *DRD1*-related disorder will probably never be a candidate for commercial drug development, but I am hopeful that one day stereotactic gene replacement therapies will be so well-developed that switching a different gene into an established vector will become an option. There are already moves afoot in the USA to offer “hyperpersonalised” ASO therapies to one or a very few patients who may benefit based on their individual variants:(372) if this is successful, it could be the start of a great change in how we approach ultra-rare disorders.

Gene identification, then, opens the door to targeted pathophysiological, phenomenological and therapeutic research, each of which can reinforce the others. Giving a disorder a genetic label does bring benefits in its own right, but more importantly it is a first step towards theoretical understanding and practical treatment.

7.6 Conclusions

When asked to describe my PhD project, I tell people that doing rare-disease genomics today is a little like doing natural history in the early 19th century. It is not (primarily) hypothesis driven research: instead I am surrounded by a rainforest of incompletely explored data in which I can hunt for “new” genes as the old-style naturalists hunted and catalogued new kinds of frog. A new kind of frog is a charming and interesting thing, but studying it may also provide new insights in a range of areas: ecology,(373) pharmacology,(374) microbiology,(374) cell biology(375) and, incidentally, genetics.(376) Similarly, identifying a new gene (or gene-phenotype relationship), or improving understanding of an old one, is a step forward in understanding physiology. For example, identification of diseases due to variants in HOPS gene subunit genes such as *VPS16* and *VPS41* can help elucidate the mechanisms of the endosomal-lysosomal pathway, while a condition involving the total loss of function of a specific receptor such as D_1 throws light on its role in human movement and neurodevelopment which cannot be provided by animal models alone.

Modern-day naturalists must travel to remote and inaccessible areas if they want to find new species, but the genomics “rainforest” is still thronged with undiscovered genetic disorders. New tools – cheap and rapid WGS, efficient and user-friendly bioinformatics, and improving rare disease population datasets – will increase our ability to find them even further. There will be roles both for clinician-scientists analysing genomic data in depth using conventional methods and detailed phenotyping, as I have, and for bioinformaticians finding new ways to screen and interpret complex variant types such as deep intronic, intergenic and structural changes.

It is a truism that rare diseases are collectively common. In paediatric neurology, very rare diseases are very common, and as well as being one of the most interesting aspects of the specialty this does present challenges. Genomic investigation is rapidly becoming a staple of clinical practice and I anticipate that within 10 years it will be commonplace for families to arrive at their initial consultation with their own WGS data on a memory stick. Paediatric neurologists working in movement disorders and beyond will need to become either more comfortable getting the best out of genomic testing, or more closely integrated with our geneticist colleagues, or preferably both. An ecosystem of patient- and family-led support and advocacy, data sharing, condition registries and research and clinical collaborations is starting to spring up around ultra-rare genetic diseases, including but not limited to movement disorders. Together, these may allow the prospects of those diagnosed with what are (for now) dangerous, distressing and disabling conditions to be transformed.

References

1. Caligiore D, Arbib MA, Miall RC, Baldassarre G. The super-learning hypothesis: Integrating learning processes across cortex, cerebellum and basal ganglia. Neurosci Biobehav Rev. 2019;100:19-34.
2. Courtemanche R, Cammalleri A. Basal Ganglia: Striosomes and the Link between Motivation and Action. Curr Biol. 2019;29(2):R62-R5.
3. Leisman G, Melillo R. The basal ganglia: motor and cognitive relationships in a clinical neurobehavioral context. Rev Neurosci. 2013;24(1):9-25.
4. Hammes J, Theis H, Giehl K, Hoenig MC, Greuel A, Tittgemeyer M, et al. Dopamine metabolism of the nucleus accumbens and fronto-striatal connectivity modulate impulse control. Brain. 2019;142(3):733-43.
5. DeLong MR. Primate models of movement disorders of basal ganglia origin. Trends Neurosci. 1990;13(7):281-5.
6. Albin RL, Young AB, Penney JB. The functional anatomy of basal ganglia disorders. Trends Neurosci. 1989;12(10):366-75.
7. Lester J, Fink S, Aronin N, DiFiglia M. Colocalization of D1 and D2 dopamine receptor mRNAs in striatal neurons. Brain Res. 1993;621(1):106-10.
8. Lanciego JL, Luquin N, Obeso JA. Functional neuroanatomy of the basal ganglia. Cold Spring Harb Perspect Med. 2012;2(12):a009621.
9. Kravitz AV, Freeze BS, Parker PR, Kay K, Thwin MT, Deisseroth K, et al. Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature. 2010;466(7306):622-6.
10. Sheean G, McGuire JR. Spastic hypertonia and movement disorders: pathophysiology, clinical presentation, and quantification. PM R. 2009;1(9):827-33.
11. Ng J, Zhen J, Meyer E, Erreger K, Li Y, Kakar N, et al. Dopamine transporter deficiency syndrome: phenotypic spectrum from infancy to adulthood. Brain. 2014;137(Pt 4):1107-19.
12. Odding E, Roebroeck ME, Stam HJ. The epidemiology of cerebral palsy: incidence, impairments and risk factors. Disabil Rehabil. 2006;28(4):183-91.
13. Balint B, Mencacci NE, Valente EM, Pisani A, Rothwell J, Jankovic J, et al. Dystonia. Nat Rev Dis Primers. 2018;4(1):25.
14. Lumsden DE, King MD, Allen NM. Status dystonicus in childhood. Curr Opin Pediatr. 2017;29(6):674-82.
15. Tomic S, Petkovic I, Pucic T, Resan B, Juric S, Rotim T. Cervical dystonia and quality of life. Acta Neurol Belg. 2016;116(4):589-92.
16. Rinnerthaler M, Mueller J, Weichbold V, Wenning GK, Poewe W. Social stigmatization in patients with cranial and cervical dystonia. Mov Disord. 2006;21(10):1636-40.
17. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health. 2015;18(6):906-14.
18. Bax M, Goldstein M, Rosenbaum P, Leviton A, Paneth N, Dan B, et al. Proposed definition and classification of cerebral palsy, April 2005. Dev Med Child Neurol. 2005;47(8):571-6.
19. Rare Disease UK: What is a rare disease? 2020 [Available from: https://www.raredisease.org.uk/what-is-a-rare-disease/.
20. Torres-Russotto D, Perlmutter JS. Task-specific dystonias: a review. Ann N Y Acad Sci. 2008;1142:179-99.
21. Jinnah HA, Berardelli A, Comella C, Defazio G, Delong MR, Factor S, et al. The focal dystonias: current views and challenges for future research. Mov Disord. 2013;28(7):926-43.
22. Brooks JC, Strauss DJ, Shavelle RM, Tran LM, Rosenbloom L, Wu YW. Recent trends in cerebral palsy survival. Part II: individual survival prognosis. Dev Med Child Neurol. 2014;56(11):1065-71.

23. McGovern E, Pringsheim T, Medina A, Cosentino C, Shalash A, Sardar Z, et al. Transitional Care for Young People with Neurological Disorders: A Scoping Review with A Focus on Patients with Movement Disorders. Mov Disord. 2021;36(6):1316-24.
24. Wijemanne S, Jankovic J. Dopa-responsive dystonia--clinical and genetic heterogeneity. Nat Rev Neurol. 2015;11(7):414-24.
25. Meneret A, Mohammad SS, Cif L, Doummar D, DeGusmao C, Anheim M, et al. Efficacy of Caffeine in ADCY5-Related Dyskinesia: A Retrospective Study. Mov Disord. 2022;37(6):1294-8.
26. Schulz A, Ajayi T, Specchio N, de Los Reyes E, Gissen P, Ballon D, et al. Study of Intraventricular Cerliponase Alfa for CLN2 Disease. N Engl J Med. 2018;378(20):1898-907.
27. Rothwell JC. Overview of neurophysiology of movement control. Clin Neurol Neurosurg. 2012;114(5):432-5.
28. Borges L. El idioma analitico de John Wilkins. Otras Inquisiciones (1937-1952). Buenos Aries: Sur; 1952.
29. Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW. The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nat Genet. 1993;5(4):327-37.
30. Ichinose H, Ohye T, Takahashi E, Seki N, Hori T, Segawa M, et al. Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. Nat Genet. 1994;8(3):236-42.
31. Walker RH. Thoughts on selected movement disorder terminology and a plea for clarity. Tremor Other Hyperkinet Mov (N Y). 2013;3.
32. Caviness JN. Treatment of myoclonus. Neurotherapeutics. 2014;11(1):188-200.
33. Merchant SHI, Vial-Undurraga F, Leodori G, van Gerpen JA, Hallett M. Myoclonus: An Electrophysiological Diagnosis. Mov Disord Clin Pract. 2020;7(5):489-99.
34. Albanese A, Bhatia K, Bressman SB, Delong MR, Fahn S, Fung VS, et al. Phenomenology and classification of dystonia: a consensus update. Mov Disord. 2013;28(7):863-73.
35. Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. Nat Genet. 2020;52(8):754-8.
36. Online Mendelian Inheritance in Man (OMIM) [Internet]. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. 1966 [cited 16/11/2022]. Available from: https://omim.org/.
37. Mencacci NE, Jinnah HA. Naming Genes for Dystonia: DYT-z or Ditzy? Tremor Other Hyperkinet Mov (N Y). 2019;9.
38. Marras C, Lohmann K, Lang A, Klein C. Fixing the broken system of genetic locus symbols: Parkinson disease and dystonia as examples. Neurology. 2012;78(13):1016-24.
39. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):512-21.
40. Sheffield Children's Hospital Diagnostic Genetics Service: Neurogenetic Disorder Service NGS Panels [updated 03/10/2018. Available from: www.sheffieldchildrens.nhs.uk/sdgs.
41. Ameur A, Che H, Martin M, Bunikis I, Dahlberg J, Hoijer I, et al. De Novo Assembly of Two Swedish Genomes Reveals Missing Segments from the Human GRCh38 Reference and Improves Variant Calling of Population-Scale Sequencing Data. Genes (Basel). 2018;9(10).
42. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473-8.
43. Gross AM, Ajay SS, Rajan V, Brown C, Bluske K, Burns NJ, et al. Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. Genet Med. 2019;21(5):1121-30.
44. Glusman G, Severson A, Dhankani V, Robinson M, Farrah T, Mauldin DE, et al. Identification of copy number variants in whole-genome data using Reference Coverage Profiles. Front Genet. 2015;6:45.

45. Duan M, Chen L, Ge Q, Lu N, Li J, Pan X, et al. Evaluating heteroplasmic variations of the mitochondrial genome from whole genome sequencing data. Gene. 2019;699:145-54.
46. Daber RD, Conlin LK, Leonard LD, Canevini MP, Vignoli A, Hosain S, et al. Ring chromosome 20. Eur J Med Genet. 2012;55(5):381-7.
47. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet. 2020;21(10):597-614.
48. Investigators GPP, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N Engl J Med. 2021;385(20):1868-80.
49. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78-81.
50. Modi A, Vai S, Caramelli D, Lari M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. Methods Mol Biol. 2021;2242:15-42.
51. Harrington CT, Lin EI, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. Arch Pathol Lab Med. 2013;137(9):1296-303.
52. Diekstra A, Bosgoed E, Rikken A, van Lier B, Kamsteeg EJ, Tychon M, et al. Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. Clin Chem. 2015;61(1):154-62.
53. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348-65.
54. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N Engl J Med. 2017;377(18):1723-32.
55. Kojima K, Nakajima T, Taga N, Miyauchi A, Kato M, Matsumoto A, et al. Gene therapy improves motor and mental function of aromatic l-amino acid decarboxylase deficiency. Brain. 2019;142(2):322-33.
56. Markun LC, Starr PA, Air EL, Marks WJ, Jr., Volz MM, Ostrem JL. Shorter disease duration correlates with improved long-term deep brain stimulation outcomes in young-onset DYT1 dystonia. Neurosurgery. 2012;71(2):325-30.
57. Koy A, Cirak S, Gonzalez V, Becker K, Roujeau T, Milesi C, et al. Deep brain stimulation is effective in pediatric patients with GNAO1 associated severe hyperkinesia. J Neurol Sci. 2018;391:31-9.
58. Meyer E, Carss KJ, Rankin J, Nichols JM, Grozeva D, Joseph AP, et al. Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nat Genet. 2017;49(2):223-37.
59. Klepper J. Glucose transporter deficiency syndrome (GLUT1DS) and the ketogenic diet. Epilepsia. 2008;49 Suppl 8:46-9.
60. Brunklaus A, Dorris L, Ellis R, Reavey E, Lee E, Forbes G, et al. The clinical utility of an SCN1A genetic diagnosis in infantile-onset epilepsy. Dev Med Child Neurol. 2013;55(2):154-61.
61. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.
62. Barrett JH, Teare MD. Linkage analysis. Methods Mol Biol. 2011;760:19-33.
63. Alkuraya FS. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Hum Genet. 2013;132(11):1197-211.
64. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508(7497):469-76.
65. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306(5940):234-8.
66. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell. 1993;72(6):971-83.

67. Kramer PL, de Leon D, Ozelius L, Risch N, Bressman SB, Brin MF, et al. Dystonia gene in Ashkenazi Jewish population is located on chromosome 9q32-34. Ann Neurol. 1990;27(2):114-20.
68. Ozelius LJ, Hewett JW, Page CE, Bressman SB, Kramer PL, Shalish C, et al. The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. Nat Genet. 1997;17(1):40-8.
69. Kwiatkowski TJ, Jr., Orr HT, Banfi S, McCall AE, Jodice C, Persichetti F, et al. The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps centromeric to D6S89 and shows no recombination, in nine large kindreds, with a dinucleotide repeat at the AM10 locus. Am J Hum Genet. 1993;53(2):391-400.
70. Banfi S, Servadio A, Chung MY, Kwiatkowski TJ, Jr., McCall AE, Duvick LA, et al. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat Genet. 1994;7(4):513-20.
71. Charlesworth G, Angelova PR, Bartolome-Robledo F, Ryten M, Trabzuni D, Stamelou M, et al. Mutations in HPCA cause autosomal-recessive primary isolated dystonia. Am J Hum Genet. 2015;96(4):657-65.
72. Fuchs T, Saunders-Pullman R, Masuho I, Luciano MS, Raymond D, Factor S, et al. Mutations in GNAL cause primary torsion dystonia. Nat Genet. 2013;45(1):88-92.
73. Mencacci NE, Rubio-Agusti I, Zdebik A, Asmus F, Ludtmann MH, Ryten M, et al. A missense mutation in KCTD17 causes autosomal dominant myoclonus-dystonia. Am J Hum Genet. 2015;96(6):938-47.
74. Zech M, Boesch S, Maier EM, Borggraefe I, Vill K, Laccone F, et al. Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. Am J Hum Genet. 2016;99(6):1377-87.
75. Zech M, Lam DD, Francescatto L, Schormair B, Salminen AV, Jochim A, et al. Recessive mutations in the alpha3 (VI) collagen gene COL6A3 cause early-onset isolated dystonia. Am J Hum Genet. 2015;96(6):883-93.
76. Monfrini E, Cogiamanian F, Salani S, Straniero L, Fagiolari G, Garbellini M, et al. A Novel Homozygous VPS11 Variant May Cause Generalized Dystonia. Ann Neurol. 2021;89(4):834-9.
77. Steel D, Zech M, Zhao C, Barwick KES, Burke D, Demailly D, et al. Loss-of-Function Variants in HOPS Complex Genes VPS16 and VPS41 Cause Early Onset Dystonia Associated with Lysosomal Abnormalities. Ann Neurol. 2020;88(5):867-77.
78. Zech M, Kumar KR, Reining S, Reunert J, Tchan M, Riley LG, et al. Biallelic AOPEP Loss-of-Function Variants Cause Progressive Dystonia with Prominent Limb Involvement. Mov Disord. 2022;37(1):137-47.
79. Heimer G, Keratar JM, Riley LG, Balasubramaniam S, Eyal E, Pietikainen LP, et al. MECR Mutations Cause Childhood-Onset Dystonia and Optic Atrophy, a Mitochondrial Fatty Acid Synthesis Disorder. Am J Hum Genet. 2016;99(6):1229-44.
80. Sleiman S, Marshall AE, Dong X, Mhanni A, Alidou-D'Anjou I, Frosk P, et al. Compound heterozygous variants in SHQ1 are associated with a spectrum of neurological features, including early-onset dystonia. Hum Mol Genet. 2022;31(4):614-24.
81. Kuipers DJS, Mandemakers W, Lu CS, Olgiati S, Breedveld GJ, Fevga C, et al. EIF2AK2 Missense Variants Associated with Early Onset Generalized Dystonia. Ann Neurol. 2021;89(3):485-97.
82. Mochel F, Rastetter A, Ceulemans B, Platzer K, Yang S, Shinde DN, et al. Variants in the SK2 channel gene (KCNN2) lead to dominant neurodevelopmental movement disorders. Brain. 2020;143(12):3564-73.
83. Steel D, Kurian MA. Recent genetic advances in early-onset dystonia. Curr Opin Neurol 2020.
84. Papandreou A, Danti FR, Spaull R, Leuzzi V, McTague A, Kurian MA. The expanding spectrum of movement disorders in genetic epilepsies. Dev Med Child Neurol. 2020;62(2):178-91.
85. Schirinzi T, Garone G, Travaglini L, Vasco G, Galosi S, Rios L, et al. Phenomenology and clinical course of movement disorder in GNAO1 variants: Results from an analytical review. Parkinsonism Relat Disord. 2019;61:19-25.

86. Fernandez-Marmiesse A, Kusumoto H, Rekarte S, Roca I, Zhang J, Myers SJ, et al. A novel missense mutation in GRIN2A causes a nonepileptic neurodevelopmental disorder. Mov Disord. 2018;33(6):992-9.
87. Zagaglia S SD, Krithika S, Hernandez-Hernandez L, Martins Custodio H, Gorman KM, Vezyroglou A, Moller RS, King MD, Bjorg Hammer T, Spaull R, Fazeli W, Bartolomaeus T, Doummar D, Keren B, Mignot C, Bednarek N, Cross JH, Mallick AA, Sanchis-Juan A, Raymond FL, Lynch BJ, Majumdar A, Stamberger H, Weckhuysen S, Sisodiya SM, Kurian MA. RHOBTB2 mutations expand the phenotypic spectrum of alternating hemiplegia of childhood. 2020.
88. Kumar KR, Davis RL, Tchan MC, Wali GM, Mahant N, Ng K, et al. Whole genome sequencing for the genetic diagnosis of heterogenous dystonia phenotypes. Parkinsonism Relat Disord. 2019;69:111-8.
89. Wang JC, Radcliff J, Coe SJ, Mahon LW. Effects of platforms, size filter cutoffs, and targeted regions of cytogenomic microarray on detection of copy number variants and uniparental disomy in prenatal diagnosis: Results from 5026 pregnancies. Prenat Diagn. 2019;39(3):137-56.
90. Bahlo M, Bennett MF, Degorski P, Tankard RM, Delatycki MB, Lockhart PJ. Recent advances in the detection of repeat expansions with short-read next-generation sequencing. F1000Res. 2018;7.
91. Aslam S, Strickland T, Molloy EJ. Neonatal Encephalopathy: Need for Recognition of Multiple Etiologies for Optimal Management. Front Pediatr. 2019;7:142.
92. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med. 2017;19(10):1151-8.
93. Ellard S, Baple EL, Callaway A, Berry I, Forrester N, Turnbull C, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020 2020 [Available from: https://www.acgs.uk.com/quality/best-practice-guidelines/#VariantGuidelines.
94. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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. 2015;17(5):405-24.
95. Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223-8.
96. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
97. Gu S, Fang L, Xu X. Using SOAPaligner for Short Reads Alignment. Curr Protoc Bioinformatics. 2013;44:11 1-7.
98. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
99. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-i9.
100. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28(3):423-5.
101. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005;33(Database issue):D501-4.
102. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
103. Guo MH, Plummer L, Chan YM, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. Am J Hum Genet. 2018;103(4):522-34.

104. Centogene [Available from: www.centogene.com/science/centopedia/comprehensive-dystonia-panel.html.
105. Invitae [Available from: www.invitae.com/en/physician/tests/03351/.
106. Genomics England PanelApp [Internet]. 2021 [cited 30/09/2022]. Available from: https://panelapp.genomicsengland.co.uk/panels/847/.
107. Wagnon JL, Mencacci NE, Barker BS, Wengert ER, Bhatia KP, Balint B, et al. Partial loss-of-function of sodium channel SCN8A in familial isolated myoclonus. Hum Mutat. 2018;39(7):965-9.
108. Luthy K, Mei D, Fischer B, De Fusco M, Swerts J, Paesmans J, et al. TBC1D24-TLDc-related epilepsy exercise-induced dystonia: rescue by antioxidants in a disease model. Brain. 2019;142(8):2319-35.
109. Gertler TS, Thompson CH, Vanoye CG, Millichap JJ, George AL, Jr. Functional consequences of a KCNT1 variant associated with status dystonicus and early-onset infantile encephalopathy. Ann Clin Transl Neurol. 2019;6(9):1606-15.
110. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51(11):1560-5.
111. gnomAD [Available from: https://gnomad.broadinstitute.org/.
112. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886-D94.
113. Align GVGD [Available from: http://agvgd.hci.utah.edu/.
114. SIFT [Available from: https://sift.bii.a-star.edu.sg/.
115. MutationTaster Berlin Institute of Health2013 [Available from: http://www.mutationtaster.org/.
116. MutationTaster2021: Berlin Institute of Health; 2021 [Available from: https://www.genecascade.org/MutationTaster2021/#transcript.
117. PolyPhen-2 [Available from: http://genetics.bwh.harvard.edu/pph2/.
118. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185(4154):862-4.
119. Provean [Available from: http://provean.jcvi.org/index.php.
120. Online Mendelian Inheritance in Man, OMIM McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) [Available from: https://omim.org/.
121. Uniprot [Available from: www.uniprot.org/.
122. Primer3Plus [Available from: http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi.
123. Human BLAT Search [Available from: https://genome.ucsc.edu/cgi-bin/hgBlat.
124. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-30.
125. Deciphering Developmental Disorders [Available from: https://www.ddduk.org/.
126. Harari S, Humbert M. Ultra-rare disease: an European perspective. Eur Respir Rev. 2020;29(156).
127. Crooke ST. Meeting the needs of patients with ultrarare diseases. Trends Mol Med. 2022;28(2):87-96.
128. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med. 2018;20(9):1054-60.
129. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho YY, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med. 2017;19(10):1105-17.
130. Kohler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res. 2021;49(D1):D1207-D17.

131. Matthews AM, Blydt-Hansen I, Al-Jabri B, Andersen J, Tarailo-Graovac M, Price M, et al. Atypical cerebral palsy: genomics analysis enables precision medicine. Genet Med. 2019;21(7):1621-8.
132. Srivastava S, Cohen JS, Vernon H, Baranano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. Ann Neurol. 2014;76(4):473-83.
133. Kobayashi Y, Tohyama J, Kato M, Akasaka N, Magara S, Kawashima H, et al. High prevalence of genetic alterations in early-onset epileptic encephalopathies associated with infantile movement disorders. Brain Dev. 2016;38(3):285-92.
134. Dale RC, Grattan-Smith P, Nicholson M, Peters GB. Microdeletions detected using chromosome microarray in children with suspected genetic movement disorders: a single-centre study. Dev Med Child Neurol. 2012;54(7):618-23.
135. Howell KB, Kornberg AJ, Harvey AS, Ryan MM, Mackay MT, Freeman JL, et al. High resolution chromosomal microarray in undiagnosed neurological disorders. J Paediatr Child Health. 2013;49(9):716-24.
136. McMichael G, Bainbridge MN, Haan E, Corbett M, Gardner A, Thompson S, et al. Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Mol Psychiatry. 2015;20(2):176-82.
137. Segel R, Ben-Pazi H, Zeligson S, Fatal-Valevski A, Aran A, Gross-Tsur V, et al. Copy number variations in cryptogenic cerebral palsy. Neurology. 2015;84(16):1660-8.
138. Oskoui M, Coutinho F, Dykeman J, Jette N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol. 2013;55(6):509-19.
139. Montaut S, Tranchant C, Drouot N, Rudolf G, Guissart C, Tarabeux J, et al. Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders. JAMA Neurol. 2018;75(10):1234-45.
140. van Eyk CL, Corbett MA, Frank MSB, Webber DL, Newman M, Berry JG, et al. Targeted resequencing identifies genes with recurrent variation in cerebral palsy. NPJ Genom Med. 2019;4:27.
141. van Eyk CL, Webber DL, Minoche AE, Perez-Jurado LA, Corbett MA, Gardner AE, et al. Yield of clinically reportable genetic variants in unselected cerebral palsy by whole genome sequencing. NPJ Genom Med. 2021;6(1):74.
142. Zech M, Boesch S, Jochim A, Weber S, Meindl T, Schormair B, et al. Clinical exome sequencing in early-onset generalized dystonia and large-scale resequencing follow-up. Mov Disord. 2017;32(4):549-59.
143. Takezawa Y, Kikuchi A, Haginoya K, Niihori T, Numata-Uematsu Y, Inui T, et al. Genomic analysis identifies masqueraders of full-term cerebral palsy. Ann Clin Transl Neurol. 2018;5(5):538-51.
144. Ngo KJ, Rexach JE, Lee H, Petty LE, Perlman S, Valera JM, et al. A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. Hum Mutat. 2020;41(2):487-501.
145. Wirth T, Tranchant C, Drouot N, Keren B, Mignot C, Cif L, et al. Increased diagnostic yield in complex dystonia through exome sequencing. Parkinsonism Relat Disord. 2020;74:50-6.
146. Zech M, Jech R, Boesch S, Skorvanek M, Weber S, Wagner M, et al. Monogenic variants in dystonia: an exome-wide sequencing study. Lancet Neurol. 2020;19(11):908-18.
147. Moreno-De-Luca A, Millan F, Pesacreta DR, Elloumi HZ, Oetjens MT, Teigen C, et al. Molecular Diagnostic Yield of Exome Sequencing in Patients With Cerebral Palsy. JAMA. 2021;325(5):467-75.
148. May HJ, Fasheun JA, Bain JM, Baugh EH, Bier LE, Revah-Politi A, et al. Genetic testing in individuals with cerebral palsy. Dev Med Child Neurol. 2021;63(12):1448-55.
149. Mei H, Yang L, Xiao T, Wang S, Wu B, Wang H, et al. Genetic Spectrum Identified by Exome Sequencing in a Chinese Pediatric Cerebral Palsy Cohort. J Pediatr. 2022;242:206-12 e6.
150. Rosello M, Caro-Llopis A, Orellana C, Oltra S, Alemany-Albert M, Marco-Hernandez AV, et al. Hidden etiology of cerebral palsy: genetic and clinical heterogeneity and efficient diagnosis by next-generation sequencing. Pediatr Res. 2021;90(2):284-8.

151. Nejabat M, Inaloo S, Sheshdeh AT, Bahramjahan S, Sarvestani FM, Katibeh P, et al. Genetic Testing in Various Neurodevelopmental Disorders Which Manifest as Cerebral Palsy: A Case Study From Iran. Front Pediatr. 2021;9:734946.
152. Yechieli M, Gulsuner S, Ben-Pazi H, Fattal A, Aran A, Kuzminsky A, et al. Diagnostic yield of chromosomal microarray and trio whole exome sequencing in cryptogenic cerebral palsy. J Med Genet. 2022;59(8):759-67.
153. Chopra M, Gable DL, Love-Nichols J, Tsao A, Rockowitz S, Sliz P, et al. Mendelian etiologies identified with whole exome sequencing in cerebral palsy. Ann Clin Transl Neurol. 2022;9(2):193-205.
154. Thomas Q, Vitobello A, Tran Mau-Them F, Duffourd Y, Fromont A, Giroud M, et al. High efficiency and clinical relevance of exome sequencing in the daily practice of neurogenetics. J Med Genet. 2022;59(5):445-52.
155. Kim A, Kumar KR, Davis RL, Mallawaarachchi AC, Gayevskiy V, Minoche AE, et al. Increased Diagnostic Yield of Spastic Paraplegia with or Without Cerebellar Ataxia Through Whole-Genome Sequencing. Cerebellum. 2019;18(4):781-90.
156. Kim MJ, Yum MS, Seo GH, Ko TS, Lee BH. Phenotypic and Genetic Complexity in Pediatric Movement Disorders. Front Genet. 2022;13:829558.
157. Lee HF, Chi CS, Tsai CR. Diagnostic yield and treatment impact of whole-genome sequencing in paediatric neurological disorders. Dev Med Child Neurol. 2021;63(8):934-8.
158. Zarrei M, Fehlings DL, Mawjee K, Switzer L, Thiruvahindrapuram B, Walker S, et al. De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. Genet Med. 2018;20(2):172-80.
159. Corbett MA, van Eyk CL, Webber DL, Bent SJ, Newman M, Harper K, et al. Pathogenic copy number variants that affect gene expression contribute to genomic burden in cerebral palsy. NPJ Genom Med. 2018;3:33.
160. Wu MC, Chang YY, Lan MY, Chen YF, Tai CH, Lin YF, et al. A Clinical and Integrated Genetic Study of Isolated and Combined Dystonia in Taiwan. J Mol Diagn. 2022;24(3):262-73.
161. McMichael G, Girirajan S, Moreno-De-Luca A, Gecz J, Shard C, Nguyen LS, et al. Rare copy number variation in cerebral palsy. Eur J Hum Genet. 2014;22(1):40-5.
162. Cordeiro D, Bullivant G, Siriwardena K, Evans A, Kobayashi J, Cohn RD, et al. Genetic landscape of pediatric movement disorders and management implications. Neurol Genet. 2018;4(5):e265.
163. Kwong AK, Tsang MH, Fung JL, Mak CC, Chan KL, Rodenburg RJT, et al. Exome sequencing in paediatric patients with movement disorders. Orphanet J Rare Dis. 2021;16(1):32.
164. Vural A, Simsir G, Tekgul S, Kocoglu C, Akcimen F, Kartal E, et al. The Complex Genetic Landscape of Hereditary Ataxias in Turkey and Implications in Clinical Practice. Mov Disord. 2021;36(7):1676-88.
165. Odgerel Z, Sonti S, Hernandez N, Park J, Ottman R, Louis ED, et al. Whole genome sequencing and rare variant analysis in essential tremor families. PLoS One. 2019;14(8):e0220512.
166. Martinez-Rubio D, Hinarejos I, Sancho P, Gorria-Redondo N, Bernado-Fonz R, Tello C, et al. Mutations, Genes, and Phenotypes Related to Movement Disorders and Ataxias. Int J Mol Sci. 2022;23(19).
167. van Egmond ME, Lugtenberg CHA, Brouwer OF, Contarino MF, Fung VSC, Heiner-Fokkema MR, et al. A post hoc study on gene panel analysis for the diagnosis of dystonia. Mov Disord. 2017;32(4):569-75.
168. Graziola F, Garone G, Stregapede F, Bosco L, Vigevano F, Curatolo P, et al. Diagnostic Yield of a Targeted Next-Generation Sequencing Gene Panel for Pediatric-Onset Movement Disorders: A 3-Year Cohort Study. Front Genet. 2019;10:1026.
169. Reale C, Panteghini C, Carecchio M, Garavaglia B. The relevance of gene panels in movement disorders diagnosis: A lab perspective. Eur J Paediatr Neurol. 2018;22(2):285-91.
170. van de Warrenburg BP, Schouten MI, de Bot ST, Vermeer S, Meijer R, Pennings M, et al. Clinical exome sequencing for cerebellar ataxia and spastic paraplegia uncovers novel gene-disease associations and unanticipated rare disorders. Eur J Hum Genet. 2016;24(10):1460-6.

171. Eratne D, Schneider A, Lynch E, Martyn M, Velakoulis D, Fahey M, et al. The clinical utility of exome sequencing and extended bioinformatic analyses in adolescents and adults with a broad range of neurological phenotypes: an Australian perspective. J Neurol Sci. 2021;420:117260.
172. Parolin Schnekenberg R, Perkins EM, Miller JW, Davies WI, D'Adamo MC, Pessia M, et al. De novo point mutations in patients diagnosed with ataxic cerebral palsy. Brain. 2015;138(Pt 7):1817-32.
173. Honzik T, Tesarova M, Magner M, Mayr J, Jesina P, Vesela K, et al. Neonatal onset of mitochondrial disorders in 129 patients: clinical and laboratory characteristics and a new approach to diagnosis. J Inherit Metab Dis. 2012;35(5):749-59.
174. Finsterer J, Kopsa W. Basal Ganglia calcification in mitochondrial disorders. Metab Brain Dis. 2005;20(3):219-26.
175. Yahia A, Ayed IB, Hamed AA, Mohammed IN, Elseed MA, Bakhiet AM, et al. Genetic diagnosis in Sudanese and Tunisian families with syndromic intellectual disability through exome sequencing. Ann Hum Genet. 2022;86(4):181-94.
176. Bauer G, Buchert R, Haack TB, Harting I, Gutschalk A. CCDC82 frameshift mutation associated with intellectual disability, spastic paraparesis, and dysmorphic features. Clin Genet. 2022;102(1):80-1.
177. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
178. Goodman LD, Cope H, Nil Z, Ravenscroft TA, Charng WL, Lu S, et al. TNPO2 variants associate with human developmental delays, neurologic deficits, and dysmorphic features and alter TNPO2 activity in Drosophila. Am J Hum Genet. 2021;108(9):1669-91.
179. Cif L, Demailly D, Lin JP, Barwick KE, Sa M, Abela L, et al. KMT2B-related disorders: expansion of the phenotypic spectrum and long-term efficacy of deep brain stimulation. Brain. 2020;143(11):3242-61.
180. Faundes V, Newman WG, Bernardini L, Canham N, Clayton-Smith J, Dallapiccola B, et al. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am J Hum Genet. 2018;102(1):175-87.
181. Oegema R, Baillat D, Schot R, van Unen LM, Brooks A, Kia SK, et al. Human mutations in integrator complex subunits link transcriptome integrity to brain development. PLoS Genet. 2017;13(5):e1006809.
182. Krall M, Htun S, Schnur RE, Brooks AS, Baker L, de Alba Campomanes A, et al. Biallelic sequence variants in INTS1 in patients with developmental delays, cataracts, and craniofacial anomalies. Eur J Hum Genet. 2019;27(4):582-93.
183. Nemani T, Steel D, Kaliakatsos M, DeVile C, Ververi A, Scott R, et al. KIF1A-related disorders in children: A wide spectrum of central and peripheral nervous system involvement. J Peripher Nerv Syst. 2020;25(2):117-24.
184. Depienne C, Trouillard O, Gourfinkel-An I, Saint-Martin C, Bouteiller D, Graber D, et al. Mechanisms for variable expressivity of inherited SCN1A mutations causing Dravet syndrome. J Med Genet. 2010;47(6):404-10.
185. Crow YJ. Aicardi-Goutieres Syndrome. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. GeneReviews((R)). Seattle (WA)1993.
186. Brunklaus A, Brunger T, Feng T, Fons C, Lehikoinen A, Panagiotakaki E, et al. The gain of function SCN1A disorder spectrum: novel epilepsy phenotypes and therapeutic implications. Brain. 2022.
187. Straub J, Konrad EDH, Gruner J, Toutain A, Bok LA, Cho MT, et al. Missense Variants in RHOBTB2 Cause a Developmental and Epileptic Encephalopathy in Humans, and Altered Levels Cause Neurological Defects in Drosophila. Am J Hum Genet. 2018;102(1):44-57.
188. Belal H, Nakashima M, Matsumoto H, Yokochi K, Taniguchi-Ikeda M, Aoto K, et al. De novo variants in RHOBTB2, an atypical Rho GTPase gene, cause epileptic encephalopathy. Hum Mutat. 2018;39(8):1070-5.

189. International Headache Society L, United Kingdom. IHS Classification ICHD-3 2019 [Available from: https://ichd-3.org/appendix/a1-migraine/a1-6-episodic-syndromes-that-may-be-associated-with-migraine/a1-6-5-alternating-hemiplegia-of-childhood/.
190. Mikati MA, Panagiotakaki E, Arzimanoglou A. Revision of the diagnostic criteria of alternating hemiplegia of childhood. Eur J Paediatr Neurol. 2021;32:A4-A5.
191. Takeshima H, Komazaki S, Nishi M, Iino M, Kangawa K. Junctophilins: a novel family of junctional membrane complex proteins. Mol Cell. 2000;6(1):11-22.
192. Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, et al. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. Nat Genet. 2001;29(4):377-8.
193. Krench M, Cho RW, Littleton JT. A Drosophila model of Huntington disease-like 2 exhibits nuclear toxicity and distinct pathogenic mechanisms from Huntington disease. Hum Mol Genet. 2016;25(15):3164-77.
194. Wilburn B, Rudnicki DD, Zhao J, Weitz TM, Cheng Y, Gu X, et al. An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron. 2011;70(3):427-40.
195. Seixas AI, Holmes SE, Takeshima H, Pavlovich A, Sachs N, Pruitt JL, et al. Loss of junctophilin-3 contributes to Huntington disease-like 2 pathogenesis. Ann Neurol. 2012;71(2):245-57.
196. Bourinaris T, Athanasiou A, Efthymiou S, Wiethoff S, Salpietro V, Houlden H. Allelic and phenotypic heterogeneity in Junctophillin-3 related neurodevelopmental and movement disorders. Eur J Hum Genet. 2021;29(6):1027-31.
197. Steel D, Vezyroglou A, Barwick K, Smith M, Vogt J, Gibbon FM, et al. Both Heterozygous and Homozygous Loss-of-Function JPH3 Variants Are Associated with a Paroxysmal Movement Disorder. Mov Disord. 2022.
198. Gregory A, Lotia M, Jeong SY, Fox R, Zhen D, Sanford L, et al. Autosomal dominant mitochondrial membrane protein-associated neurodegeneration (MPAN). Mol Genet Genomic Med. 2019;7(7):e00736.
199. Weissbach A, Pauly MG, Herzog R, Hahn L, Halmans S, Hamami F, et al. Relationship of Genotype, Phenotype, and Treatment in Dopa-Responsive Dystonia: MDSGene Review. Mov Disord. 2022;37(2):237-52.
200. Overbeck S, Uciechowski P, Ackland ML, Ford D, Rink L. Intracellular zinc homeostasis in leukocyte subsets is regulated by different expression of zinc exporters ZnT-1 to ZnT-9. J Leukoc Biol. 2008;83(2):368-80.
201. Perez Y, Shorer Z, Liani-Leibson K, Chabosseau P, Kadir R, Volodarsky M, et al. SLC30A9 mutation affecting intracellular zinc homeostasis causes a novel cerebro-renal syndrome. Brain. 2017;140(4):928-39.
202. Kleyner R, Arif M, Marchi E, Horowitz N, Haworth A, King B, et al. Autosomal recessive SLC30A9 variants in a proband with a cerebrorenal syndrome and no parental consanguinity. Cold Spring Harb Mol Case Stud. 2022;8(2).
203. Ferreira CR, Gahl WA. Disorders of metal metabolism. Transl Sci Rare Dis. 2017;2(3-4):101-39.
204. Steel D, Danti, F.R., Abunada, M., Kamien, B., Malhotra, S., Topf, M. Kaliakatsos, M., Valentine, J., Nemeth, A., Reid, K., Jayawant, S., Mankad, K., Sudhakar, S., Ben-Pazi, H., Barwick, K.E.S., Kurian, M.A. . Clinical phenotype in individuals with Birk-Landau-Perez syndrome associated with biallelic SLC30A9 mutations. . Neurology (under review). 2022.
205. Kleyner R, Mohammad A, Marchi E, Horowitz N, Haworth A, King B, et al. Autosomal recessive SLC30A9 variants in a Proband with a Cerebro-Renal Syndrome and No Parental Consanguinity. Cold Spring Harb Mol Case Stud. 2021.
206. Jiang M, Bajpayee NS. Molecular mechanisms of go signaling. Neurosignals. 2009;17(1):23-41.
207. Nakamura K, Kodera H, Akita T, Shiina M, Kato M, Hoshino H, et al. De Novo mutations in GNAO1, encoding a Galphao subunit of heterotrimeric G proteins, cause epileptic encephalopathy. Am J Hum Genet. 2013;93(3):496-505.

208. Kulkarni N, Tang S, Bhardwaj R, Bernes S, Grebe TA. Progressive Movement Disorder in Brothers Carrying a GNAO1 Mutation Responsive to Deep Brain Stimulation. J Child Neurol. 2016;31(2):211-4.
209. Ananth AL, Robichaux-Viehoever A, Kim YM, Hanson-Kahn A, Cox R, Enns GM, et al. Clinical Course of Six Children With GNAO1 Mutations Causing a Severe and Distinctive Movement Disorder. Pediatr Neurol. 2016;59:81-4.
210. Saitsu H, Fukai R, Ben-Zeev B, Sakai Y, Mimaki M, Okamoto N, et al. Phenotypic spectrum of GNAO1 variants: epileptic encephalopathy to involuntary movements with severe developmental delay. Eur J Hum Genet. 2016;24(1):129-34.
211. Menke LA, Engelen M, Alders M, Odekerken VJ, Baas F, Cobben JM. Recurrent GNAO1 Mutations Associated With Developmental Delay and a Movement Disorder. J Child Neurol. 2016;31(14):1598-601.
212. Axeen E, Bell E, Robichaux Viehoever A, Schreiber JM, Sidiropoulos C, Goodkin HP. Results of the First GNAO1-Related Neurodevelopmental Disorders Caregiver Survey. Pediatr Neurol. 2021;121:28-32.
213. Wirth T, Garone G, Kurian MA, Piton A, Millan F, Telegrafi A, et al. Highlighting the Dystonic Phenotype Related to GNAO1. Mov Disord. 2022;37(7):1547-54.
214. Feng H, Sjogren B, Karaj B, Shaw V, Gezer A, Neubig RR. Movement disorder in GNAO1 encephalopathy associated with gain-of-function mutations. Neurology. 2017;89(8):762-70.
215. Miyamoto S, Nakashima M, Fukumura S, Kumada S, Saitsu H. An intronic GNAO1 variant leading to in-frame insertion cause movement disorder controlled by deep brain stimulation. Neurogenetics. 2022;23(2):129-35.
216. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482(7384):226-31.
217. Bryant L, Li D, Cox SG, Marchione D, Joiner EF, Wilson K, et al. Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. Sci Adv. 2020;6(49).
218. Alves C, Sherbini O, D'Arco F, Steel D, Kurian MA, Radio FC, et al. Brain Abnormalities in Patients with Germline Variants in H3F3: Novel Imaging Findings and Neurologic Symptoms Beyond Somatic Variants and Brain Tumors. AJNR Am J Neuroradiol. 2022;43(7):1048-53.
219. Gabriele M, Vulto-van Silfhout AT, Germain PL, Vitriolo A, Kumar R, Douglas E, et al. YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. Am J Hum Genet. 2017;100(6):907-25.
220. Carminho-Rodrigues MT, Steel D, Sousa SB, Brandt G, Guipponi M, Laurent S, et al. Complex movement disorder in a patient with heterozygous YY1 mutation (Gabriele-de Vries syndrome). Am J Med Genet A. 2020;182(9):2129-32.
221. Zorzi G, Keller Sarmiento IJ, Danti FR, Bustos BI, Invernizzi F, Panteghini C, et al. YY1-Related Dystonia: Clinical Aspects and Long-Term Response to Deep Brain Stimulation. Mov Disord. 2021;36(6):1461-2.
222. Malaquias MJ, Damasio J, Mendes A, Freixo JP, Magalhaes M. A Case of YY1-Related Isolated Dystonia with Severe Oromandibular Involvement. Mov Disord. 2021;36(11):2705-6.
223. Ferng A, Thulin P, Walsh E, Weissbrod PA, Friedman J. YY1: A New Gene for Childhood Onset Dystonia with Prominent Oromandibular-Laryngeal Involvement? Mov Disord. 2022;37(1):227-8.
224. Yellajoshyula D, Liang CC, Pappas SS, Penati S, Yang A, Mecano R, et al. The DYT6 Dystonia Protein THAP1 Regulates Myelination within the Oligodendrocyte Lineage. Dev Cell. 2017;42(1):52-67 e4.
225. Pevsner J, Hsu SC, Scheller RH. n-Sec1: a neural-specific syntaxin-binding protein. Proc Natl Acad Sci U S A. 1994;91(4):1445-9.

226. Saitsu H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nat Genet. 2008;40(6):782-8.
227. Stamberger H, Nikanorova M, Willemsen MH, Accorsi P, Angriman M, Baier H, et al. STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. Neurology. 2016;86(10):954-62.
228. Loussouarn A, Doummar D, Beaugendre Y, Bienvenu T, Charles P, Depienne C, et al. Tremor-like subcortical myoclonus in STXBP1 encephalopathy. Eur J Paediatr Neurol. 2021;34:62-6.
229. Xian J, Parthasarathy S, Ruggiero SM, Balagura G, Fitch E, Helbig K, et al. Assessing the landscape of STXBP1-related disorders in 534 individuals. Brain. 2022;145(5):1668-83.
230. Riedhammer KM, Stockler S, Ploski R, Wenzel M, Adis-Dutschmann B, Ahting U, et al. De novo stop-loss variants in CLDN11 cause hypomyelinating leukodystrophy. Brain. 2021;144(2):411-9.
231. Yang H, Wei H, Shen L, Kumar CS, Chen Q, Chen Y, et al. A novel stop-loss DAX1 variant affecting its protein-interaction with SF1 precedes the adrenal hypoplasia congenital with rare spontaneous precocious puberty and elevated hypothalamic-pituitary-gonadal/adrenal axis responses. Eur J Med Genet. 2021;64(5):104192.
232. Spaull R, Steel D, Barwick K, Prabhakar P, Wakeling E, Kurian MA. STXBP1 Stop-Loss Mutation Associated with Complex Early Onset Movement Disorder without Epilepsy. Mov Disord Clin Pract. 2022;9(6):837-40.
233. van der Kant R, Jonker CT, Wijdeven RH, Bakker J, Janssen L, Klumperman J, et al. Characterization of the Mammalian CORVET and HOPS Complexes and Their Modular Restructuring for Endosome Specificity. J Biol Chem. 2015;290(51):30280-90.
234. Jiang P, Nishimura T, Sakamaki Y, Itakura E, Hatta T, Natsume T, et al. The HOPS complex mediates autophagosome-lysosome fusion through interaction with syntaxin 17. Mol Biol Cell. 2014;25(8):1327-37.
235. Balderhaar HJ, Ungermann C. CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. J Cell Sci. 2013;126(Pt 6):1307-16.
236. Orr A, Song H, Rusin SF, Kettenbach AN, Wickner W. HOPS catalyzes the interdependent assembly of each vacuolar SNARE into a SNARE complex. Mol Biol Cell. 2017;28(7):975-83.
237. Xu H, Jun Y, Thompson J, Yates J, Wickner W. HOPS prevents the disassembly of trans-SNARE complexes by Sec17p/Sec18p during membrane fusion. EMBO J. 2010;29(12):1948-60.
238. Wartosch L, Gunesdogan U, Graham SC, Luzio JP. Recruitment of VPS33A to HOPS by VPS16 Is Required for Lysosome Fusion with Endosomes and Autophagosomes. Traffic. 2015;16(7):727-42.
239. Edvardson S, Gerhard F, Jalas C, Lachmann J, Golan D, Saada A, et al. Hypomyelination and developmental delay associated with VPS11 mutation in Ashkenazi-Jewish patients. J Med Genet. 2015;52(11):749-53.
240. Kondo H, Maksimova N, Otomo T, Kato H, Imai A, Asano Y, et al. Mutation in VPS33A affects metabolism of glycosaminoglycans: a new type of mucopolysaccharidosis with severe systemic symptoms. Hum Mol Genet. 2017;26(1):173-83.
241. Cai X, Chen X, Wu S, Liu W, Zhang X, Zhang D, et al. Homozygous mutation of VPS16 gene is responsible for an autosomal recessive adolescent-onset primary dystonia. Sci Rep. 2016;6:25834.
242. Kaler SG, Gallo LK, Proud VK, Percy AK, Mark Y, Segal NA, et al. Occipital horn syndrome and a mild Menkes phenotype associated with splice site mutations at the MNK locus. Nat Genet. 1994;8(2):195-202.
243. Kennerson ML, Nicholson GA, Kaler SG, Kowalski B, Mercer JF, Tang J, et al. Missense mutations in the copper transporter gene ATP7A cause X-linked distal hereditary motor neuropathy. Am J Hum Genet. 2010;86(3):343-52.
244. Tucci ML, Harrington AJ, Caldwell GA, Caldwell KA. Modeling dopamine neuron degeneration in Caenorhabditis elegans. Methods Mol Biol. 2011;793:129-48.
245. Gutierrez-Zepeda A, Luo Y. Testing the amyloid toxicity hypothesis of Alzheimer's disease in transgenic Caenorhabditis elegans model. Front Biosci. 2004;9:3333-8.

246. Ruan Q, Harrington AJ, Caldwell KA, Caldwell GA, Standaert DG. VPS41, a protein involved in lysosomal trafficking, is protective in Caenorhabditis elegans and mammalian cellular models of Parkinson's disease. Neurobiol Dis. 2010;37(2):330-8.
247. Griffin EF, Yan X, Caldwell KA, Caldwell GA. Distinct functional roles of Vps41-mediated neuroprotection in Alzheimer's and Parkinson's disease models of neurodegeneration. Hum Mol Genet. 2018;27(24):4176-93.
248. van der Welle REN JR, Burns C, Sanza P, ten Brink C, Fasano A, Chen L, Zwartkruis FJ, Zwakenberg S, Griffin EF, van der Beek J, Veenendaal T, Liv N, Blaser S, Sepulveda C, Lozano AM, Yoon G, Asensio CS, Caldwell GA, Caldwell KA, Chitayat D, Klumperman J. VPS41 recessive mutation causes ataxia and dystonia with retinal dystrophy and mental retardation by inhibiting HOPS function and mTORC1 signaling. bioarXiv. 2019.
249. Sanderson LE, Lanko K, Alsagob M, Almass R, Al-Ahmadi N, Najafi M, et al. Bi-allelic variants in HOPS complex subunit VPS41 cause cerebellar ataxia and abnormal membrane trafficking. Brain. 2021;144(3):769-80.
250. Radisky DC, Snyder WB, Emr SD, Kaplan J. Characterization of VPS41, a gene required for vacuolar trafficking and high-affinity iron transport in yeast. Proc Natl Acad Sci U S A. 1997;94(11):5662-6.
251. Pols MS, ten Brink C, Gosavi P, Oorschot V, Klumperman J. The HOPS proteins hVps41 and hVps39 are required for homotypic and heterotypic late endosome fusion. Traffic. 2013;14(2):219-32.
252. Zhan L, Chen S, Li K, Liang D, Zhu X, Liu L, et al. Autophagosome maturation mediated by Rab7 contributes to neuroprotection of hypoxic preconditioning against global cerebral ischemia in rats. Cell Death Dis. 2017;8(7):e2949.
253. Hauser F, Gokce S, Werner G, Danckwardt S, Sollfrank S, Neukirch C, et al. A non-invasive diagnostic assay for rapid detection and characterization of aberrant mRNA-splicing by nonsense mediated decay inhibition. Mol Genet Metab. 2020;130(1):27-35.
254. Bergsma AJ, In 't Groen SLM, Catalano F, Yamanaka M, Takahashi S, Okumiya T, et al. A generic assay for the identification of splicing variants that induce nonsense-mediated decay in Pompe disease. Eur J Hum Genet. 2021;29(3):422-33.
255. Pott H, Bruggemann N, Reese R, Zeuner KE, Gandor F, Gruber D, et al. Truncating VPS16 Mutations Are Rare in Early Onset Dystonia. Ann Neurol. 2021;89(3):625-6.
256. Li XY, Wang L, Guo Y, Wan XH. Mutations in the VPS16 Gene in 56 Early-Onset Dystonia Patients. Mov Disord. 2021;36(3):780-1.
257. Li LX, Jiang LT, Liu Y, Zhang XL, Pan YG, Pan LZ, et al. Mutation screening of VPS16 gene in patients with isolated dystonia. Parkinsonism Relat Disord. 2021;83:63-5.
258. Ostrozovicova M, Jech R, Steel D, Pavelekova P, Han V, Gdovinova Z, et al. A Recurrent VPS16 p.Arg187* Nonsense Variant in Early-Onset Generalized Dystonia. Mov Disord. 2021;36(8):1984-5.
259. Park J, Reilaender A, Petry-Schmelzer JN, Stobe P, Cordts I, Harmuth F, et al. Transcript-Specific Loss-of-Function Variants in VPS16 Are Enriched in Patients With Dystonia. Neurol Genet. 2022;8(1):e644.
260. Gu X, Lin J, Hou Y, Zhang L, Shang H. De Novo Missense Mutation of VPS16 in a Chinese Patient with Generalized Dystonia with Myoclonus. Mov Disord Clin Pract. 2022;9(4):551-2.
261. Sofou K, Meier K, Sanderson LE, Kaminski D, Montoliu-Gaya L, Samuelsson E, et al. Bi-allelic VPS16 variants limit HOPS/CORVET levels and cause a mucopolysaccharidosis-like disease. EMBO Mol Med. 2021;13(5):e13376.
262. Yildiz Y, Kosukcu C, Aygun D, Akcaboy M, Oztek Celebi FZ, Tasci Yildiz Y, et al. Homozygous missense VPS16 variant is associated with a novel disease, resembling mucopolysaccharidosis-plus syndrome in two siblings. Clin Genet. 2021;100(3):308-17.
263. Peng C, Ye J, Yan S, Kong S, Shen Y, Li C, et al. Ablation of vacuole protein sorting 18 (Vps18) gene leads to neurodegeneration and impaired neuronal migration by disrupting multiple vesicle transport pathways to lysosomes. J Biol Chem. 2012;287(39):32861-73.

264. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet. 2012;44(12):1365-9.
265. Klein MO, Battagello DS, Cardoso AR, Hauser DN, Bittencourt JC, Correa RG. Dopamine: Functions, Signaling, and Association with Neurological Diseases. Cell Mol Neurobiol. 2019;39(1):31-59.
266. Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. Neuron. 2012;74(5):858-73.
267. Lobo MK. Molecular profiling of striatonigral and striatopallidal medium spiny neurons past, present, and future. Int Rev Neurobiol. 2009;89:1-35.
268. Lee HJ, Weitz AJ, Bernal-Casas D, Duffy BA, Choy M, Kravitz AV, et al. Activation of Direct and Indirect Pathway Medium Spiny Neurons Drives Distinct Brain-wide Responses. Neuron. 2016;91(2):412-24.
269. Mishra A, Singh S, Shukla S. Physiological and Functional Basis of Dopamine Receptors and Their Role in Neurogenesis: Possible Implication for Parkinson's disease. J Exp Neurosci. 2018;12:1179069518779829.
270. Jackson DM, Westlind-Danielsson A. Dopamine receptors: molecular biology, biochemistry and behavioural aspects. Pharmacol Ther. 1994;64(2):291-370.
271. Willemsen MA, Verbeek MM, Kamsteeg EJ, de Rijk-van Andel JF, Aeby A, Blau N, et al. Tyrosine hydroxylase deficiency: a treatable disorder of brain catecholamine biosynthesis. Brain. 2010;133(Pt 6):1810-22.
272. Rilstone JJ, Alkhater RA, Minassian BA. Brain dopamine-serotonin vesicular transport disease and its treatment. N Engl J Med. 2013;368(6):543-50.
273. Pearson TS, Gilbert L, Opladen T, Garcia-Cazorla A, Mastrangelo M, Leuzzi V, et al. AADC deficiency from infancy to adulthood: Symptoms and developmental outcome in an international cohort of 63 patients. J Inherit Metab Dis. 2020;43(5):1121-30.
274. Diggle CP, Sukoff Rizzo SJ, Popiolek M, Hinttala R, Schulke JP, Kurian MA, et al. Biallelic Mutations in PDE10A Lead to Loss of Striatal PDE10A and a Hyperkinetic Movement Disorder with Onset in Infancy. Am J Hum Genet. 2016;98(4):735-43.
275. Mencacci NE, Kamsteeg EJ, Nakashima K, R'Bibo L, Lynch DS, Balint B, et al. De Novo Mutations in PDE10A Cause Childhood-Onset Chorea with Bilateral Striatal Lesions. Am J Hum Genet. 2016;98(4):763-71.
276. Ferrini A, Steel D, Barwick K, Kurian MA. An Update on the Phenotype, Genotype and Neurobiology of ADCY5-Related Disease. Mov Disord. 2021;36(5):1104-14.
277. Manconi M, Garcia-Borreguero D, Schormair B, Videnovic A, Berger K, Ferri R, et al. Restless legs syndrome. Nat Rev Dis Primers. 2021;7(1):80.
278. Grace AA. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. Nat Rev Neurosci. 2016;17(8):524-32.
279. Sharma A, Couture J. A review of the pathophysiology, etiology, and treatment of attention-deficit hyperactivity disorder (ADHD). Ann Pharmacother. 2014;48(2):209-25.
280. Moses TEH, Burmeister M, Greenwald MK. Heroin delay discounting and impulsivity: Modulation by DRD1 genetic variation. Addict Biol. 2020;25(3):e12777.
281. Dolzan V, Plesnicar BK, Serretti A, Mandelli L, Zalar B, Koprivsek J, et al. Polymorphisms in dopamine receptor DRD1 and DRD2 genes and psychopathological and extrapyramidal symptoms in patients on long-term antipsychotic treatment. Am J Med Genet B Neuropsychiatr Genet. 2007;144B(6):809-15.
282. Dai D, Wang Y, Wang L, Li J, Ma Q, Tao J, et al. Polymorphisms of DRD2 and DRD3 genes and Parkinson's disease: A meta-analysis. Biomed Rep. 2014;2(2):275-81.
283. Kaalund SS, Newburn EN, Ye T, Tao R, Li C, Deep-Soboslay A, et al. Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. Mol Psychiatry. 2014;19(12):1258-66.

284. van der Weijden MCM, Rodriguez-Contreras D, Delnooz CCS, Robinson BG, Condon AF, Kielhold ML, et al. A Gain-of-Function Variant in Dopamine D2 Receptor and Progressive Chorea and Dystonia Phenotype. Mov Disord. 2021;36(3):729-39.
285. Mencacci NE, Steel D, Magrinelli F, Hsu J, Keller Sarmiento IJ, Troncoso Schifferli M, et al. Childhood-Onset Chorea Caused by a Recurrent De Novo DRD2 Variant. Mov Disord. 2021;36(6):1472-3.
286. Yin J, Chen KM, Clark MJ, Hijazi M, Kumari P, Bai XC, et al. Structure of a D2 dopamine receptor-G-protein complex in a lipid membrane. Nature. 2020;584(7819):125-9.
287. Hyland K, Surtees RA, Heales SJ, Bowron A, Howells DW, Smith I. Cerebrospinal fluid concentrations of pterins and metabolites of serotonin and dopamine in a pediatric reference population. Pediatr Res. 1993;34(1):10-4.
288. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017;45(D1):D840-D5.
289. Wegler M, Jia X, Alders M, Bouman A, Chen J, Duan X, et al. De novo variants in the PABP domain of PABPC1 lead to developmental delay. Genet Med. 2022;24(8):1761-73.
290. The Genotype Tissue Expression Project [Internet]. GTEx Consortium, National Institute of Health. 2015 [cited 09/08/2022]. Available from: https://gtexportal.org/home/aboutGTEx.
291. Costa Mdo C, Paulson HL. Toward understanding Machado-Joseph disease. Prog Neurobiol. 2012;97(2):239-57.
292. Oo JA, Irmer B, Gunther S, Warwick T, Palfi K, Izquierdo Ponce J, et al. ZNF354C is a transcriptional repressor that inhibits endothelial angiogenic sprouting. Sci Rep. 2020;10(1):19079.
293. Kumar NN, Velic A, Soliz J, Shi Y, Li K, Wang S, et al. PHYSIOLOGY. Regulation of breathing by CO(2) requires the proton-activated receptor GPR4 in retrotrapezoid nucleus neurons. Science. 2015;348(6240):1255-60.
294. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84(4):524-33.
295. Barow E, Schneider SA, Bhatia KP, Ganos C. Oculogyric crises: Etiology, pathophysiology and therapeutic approaches. Parkinsonism Relat Disord. 2017;36:3-9.
296. Xu M, Moratalla R, Gold LH, Hiroi N, Koob GF, Graybiel AM, et al. Dopamine D1 receptor mutant mice are deficient in striatal expression of dynorphin and in dopamine-mediated behavioral responses. Cell. 1994;79(4):729-42.
297. Drago J, Gerfen CR, Lachowicz JE, Steiner H, Hollon TR, Love PE, et al. Altered striatal function in a mutant mouse lacking D1A dopamine receptors. Proc Natl Acad Sci U S A. 1994;91(26):12564-8.
298. Smith DR, Striplin CD, Geller AM, Mailman RB, Drago J, Lawler CP, et al. Behavioural assessment of mice lacking D1A dopamine receptors. Neuroscience. 1998;86(1):135-46.
299. Tran AH, Tamura R, Uwano T, Kobayashi T, Katsuki M, Ono T. Dopamine D1 receptors involved in locomotor activity and accumbens neural responses to prediction of reward associated with place. Proc Natl Acad Sci U S A. 2005;102(6):2117-22.
300. Nakamura T, Sato A, Kitsukawa T, Momiyama T, Yamamori T, Sasaoka T. Distinct motor impairments of dopamine D1 and D2 receptor knockout mice revealed by three types of motor behavior. Front Integr Neurosci. 2014;8:56.
301. Durieux PF, Schiffmann SN, de Kerchove d'Exaerde A. Differential regulation of motor control and response to dopaminergic drugs by D1R and D2R neurons in distinct dorsal striatum subregions. EMBO J. 2012;31(3):640-53.
302. Kim YC, Lee HG, Han KA. D1 dopamine receptor dDA1 is required in the mushroom body neurons for aversive and appetitive learning in Drosophila. J Neurosci. 2007;27(29):7640-7.
303. Silva B, Hidalgo S, Campusano JM. Dop1R1, a type 1 dopaminergic receptor expressed in Mushroom Bodies, modulates Drosophila larval locomotion. PLoS One. 2020;15(2):e0229671.

304. Kottler B, Faville R, Bridi JC, Hirth F. Inverse Control of Turning Behavior by Dopamine D1 Receptor Signaling in Columnar and Ring Neurons of the Central Complex in Drosophila. Curr Biol. 2019;29(4):567-77 e6.
305. Hall A, Provins L, Valade A. Novel Strategies To Activate the Dopamine D1 Receptor: Recent Advances in Orthosteric Agonism and Positive Allosteric Modulation. J Med Chem. 2019;62(1):128-40.
306. Xiao P, Yan W, Gou L, Zhong YN, Kong L, Wu C, et al. Ligand recognition and allosteric regulation of DRD1-Gs signaling complexes. Cell. 2021;184(4):943-56 e18.
307. Casado-Anguera V, Cortes A, Casado V, Moreno E. Targeting the receptor-based interactome of the dopamine D1 receptor: looking for heteromer-selective drugs. Expert Opin Drug Discov. 2019;14(12):1297-312.
308. Meng F, Guo Z, Hu Y, Mai W, Zhang Z, Zhang B, et al. CD73-derived adenosine controls inflammation and neurodegeneration by modulating dopamine signalling. Brain. 2019;142(3):700-18.
309. Hasbi A, Madras BK, Bergman J, Kohut S, Lin Z, Withey SL, et al. Delta-Tetrahydrocannabinol Increases Dopamine D1-D2 Receptor Heteromer and Elicits Phenotypic Reprogramming in Adult Primate Striatal Neurons. iScience. 2020;23(1):100794.
310. Liu Z, Le Q, Lv Y, Chen X, Cui J, Zhou Y, et al. A distinct D1-MSN subpopulation down-regulates dopamine to promote negative emotional state. Cell Res. 2022;32(2):139-56.
311. Tassone A, Sciamanna G, Bonsi P, Martella G, Pisani A. Experimental models of dystonia. Int Rev Neurobiol. 2011;98:551-72.
312. Keam SJ. Eladocagene Exuparvovec: First Approval. Drugs. 2022;82(13):1427-32.
313. Tai CH, Lee NC, Chien YH, Byrne BJ, Muramatsu SI, Tseng SH, et al. Long-term efficacy and safety of eladocagene exuparvovec in patients with AADC deficiency. Mol Ther. 2022;30(2):509-18.
314. Pearson TS, Gupta N, San Sebastian W, Imamura-Ching J, Viehoever A, Grijalvo-Perez A, et al. Gene therapy for aromatic L-amino acid decarboxylase deficiency by MR-guided direct delivery of AAV2-AADC to midbrain dopaminergic neurons. Nat Commun. 2021;12(1):4251.
315. UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480-D9.
316. Tinsley RB, Bye CR, Parish CL, Tziotis-Vais A, George S, Culvenor JG, et al. Dopamine D2 receptor knockout mice develop features of Parkinson disease. Ann Neurol. 2009;66(4):472-84.
317. Reid KM, Steel, D., Nair, S., Bhate, S., Biassoni, L., Sudhakar, S., Heys, M., Burke, E., Kamsteeg, E-J., "Genomics England Research Consortium", Hameed, B., Zech, M., Mencacci, N., Barwick, K., Topf, M., Kurian, M.A. Loss of function variants in DRD1 in infantile parkinsonism-dystonia (unpublished manuscript). 2023.
318. Rajan-Babu IS, Peng JJ, Chiu R, Study I, Study C, Li C, et al. Genome-wide sequencing as a first-tier screening test for short tandem repeat expansions. Genome Med. 2021;13(1):126.
319. Ding Q, Somerville C, Manshaei R, Trost B, Reuter MS, Kalbfleisch K, et al. SCIP: software for efficient clinical interpretation of copy number variants detected by whole-genome sequencing. Hum Genet. 2022.
320. Mitsuhashi S, Matsumoto N. Long-read sequencing for rare human genetic diseases. J Hum Genet. 2020;65(1):11-9.
321. Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. 2017;27(11):1895-903.
322. The UniProt C. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158-D69.
323. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. Nucleic Acids Res. 2021;49(D1):D981-D7.
324. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1 30 1-1 3.

325. Shirley MD, Tang H, Gallione CJ, Baugher JD, Frelin LP, Cohen B, et al. Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. N Engl J Med. 2013;368(21):1971-9.
326. Jutel A. Sociology of diagnosis: a preliminary review. Sociol Health Illn. 2009;31(2):278-99.
327. Rotstein M, Doran J, Yang H, Ullner PM, Engelstad K, De Vivo DC. Glut1 deficiency and alternating hemiplegia of childhood. Neurology. 2009;73(23):2042-4.
328. Cordani R, Pisciotta L, Mancardi MM, Stagnaro M, Prato G, Giacomini T, et al. Alternating Hemiplegia of Childhood in a Child Harboring a Novel TBC1D24 Mutation: Case Report and Literature Review. Neuropediatrics. 2022;53(1):69-74.
329. de Vries B, Stam AH, Beker F, van den Maagdenberg AM, Vanmolkot KR, Laan L, et al. CACNA1A mutation linking hemiplegic migraine and alternating hemiplegia of childhood. Cephalalgia. 2008;28(8):887-91.
330. Bax MC. Terminology and Classification of Cerebral Palsy. Dev Med Child Neurol. 1964;6:295-7.
331. McHale DP, Jackson AP, Campbell, Levene MI, Corry P, Woods CG, et al. A gene for ataxic cerebral palsy maps to chromosome 9p12-q12. Eur J Hum Genet. 2000;8(4):267-72.
332. Blumel J, Evans EB, Eggers GW. Hereditary cerebral palsy; a preliminary report. J Pediatr. 1957;50(4):454-8.
333. Hutton JL, Cooke T, Pharoah PO. Life expectancy in children with cerebral palsy. BMJ. 1994;309(6952):431-5.
334. Nygaard TG, Waran SP, Levine RA, Naini AB, Chutorian AM. Dopa-responsive dystonia simulating cerebral palsy. Pediatr Neurol. 1994;11(3):236-40.
335. Gucuyener K, Seyrantepe V, Topaloglu H, Ozguc. Mitochondrial deletion in a boy with growth hormone deficiency mimicking cerebral palsy. J Inherit Metab Dis. 1998;21(2):173-4.
336. Aravamuthan BR, Shusterman M, Green Snyder L, Lemmon ME, Bain JM, Gross P, et al. Diagnostic preferences include discussion of etiology for adults with cerebral palsy and their caregivers. Dev Med Child Neurol. 2022;64(6):723-33.
337. Darwin C. Letter to J.D. Hooker 1857 [Available from: https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-2130.xml.
338. Burki T. UK explores whole-genome sequencing for newborn babies. Lancet. 2022;400(10348):260-1.
339. Wofford S, Noblin S, Davis JM, Farach LS, Hashmi SS, Mancias P, et al. Genetic Testing Practices of Genetic Counselors, Geneticists, and Pediatric Neurologists With Regard to Childhood-Onset Neurogenetic Conditions. J Child Neurol. 2019;34(4):177-83.
340. SWAN-UK. 2021 [Available from: https://www.undiagnosed.org.uk/swanuk-news/introducing-swan-europe/.
341. GeneticAllianceUK. Genetic Alliance UK 2021 [Available from: https://geneticalliance.org.uk/.
342. EURODIS. RareConnect.org 2009 [Available from: https://www.rareconnect.org/en/.
343. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988-D95.
344. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.
345. NIHR. NIHR Bioresource 2022 [Available from: https://bioresource.nihr.ac.uk/.
346. RARE-X. RARE-X. 2022.
347. Ghadessi M, Tang R, Zhou J, Liu R, Wang C, Toyoizumi K, et al. A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). Orphanet J Rare Dis. 2020;15(1):69.
348. Thompson R, Robertson A, Lochmuller H. Natural History, Trial Readiness and Gene Discovery: Advances in Patient Registries for Neuromuscular Disease. Adv Exp Med Biol. 2017;1031:97-124.
349. Budinger D, Barral S, Soo AKS, Kurian MA. The role of manganese dysregulation in neurological disease: emerging evidence. Lancet Neurol. 2021;20(11):956-68.

350. Meneret A, Gras D, McGovern E, Roze E. Caffeine and the Dyskinesia Related to Mutations in the ADCY5 Gene. Ann Intern Med. 2019;171(6):439.
351. ClinicalTrials.gov. CoA-Z in Pantothenate Kinase-associated Neurodegeneration (PKAN) 2022 [Available from: https://clinicaltrials.gov/ct2/show/NCT04182763?term=coa-z&draw=2&rank=1.
352. Pineda M, Walterfang M, Patterson MC. Miglustat in Niemann-Pick disease type C patients: a review. Orphanet J Rare Dis. 2018;13(1):140.
353. Sehara Y, Fujimoto KI, Ikeguchi K, Katakai Y, Ono F, Takino N, et al. Persistent Expression of Dopamine-Synthesizing Enzymes 15 Years After Gene Transfer in a Primate Model of Parkinson's Disease. Hum Gene Ther Clin Dev. 2017;28(2):74-9.
354. Kurian MA, Zhen J, Cheng SY, Li Y, Mordekar SR, Jardine P, et al. Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. J Clin Invest. 2009;119(6):1595-603.
355. Ng J, Barral S, De La Fuente Barrigon C, Lignani G, Erdem FA, Wallings R, et al. Gene therapy restores dopamine transporter expression and ameliorates pathology in iPSC and mouse models of infantile parkinsonism. Sci Transl Med. 2021;13(594).
356. Cota-Coronado JA, Sandoval-Avila S, Gaytan-Davila YP, Diaz NF, Vega-Ruiz B, Padilla-Camberos E, et al. New transgenic models of Parkinson's disease using genome editing technology. Neurologia (Engl Ed). 2020;35(7):486-99.
357. Rahman MU, Bilal M, Shah JA, Kaushik A, Teissedre PL, Kujawska M. CRISPR-Cas9-Based Technology and Its Relevance to Gene Editing in Parkinson's Disease. Pharmaceutics. 2022;14(6).
358. Raguram A, Banskota S, Liu DR. Therapeutic in vivo delivery of gene editing agents. Cell. 2022;185(15):2806-27.
359. Gillmore JD, Gane E, Taubel J, Kao J, Fontana M, Maitland ML, et al. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. N Engl J Med. 2021;385(6):493-502.
360. Alshaer W, Zureigat H, Al Karaki A, Al-Kadash A, Gharaibeh L, Hatmal MM, et al. siRNA: Mechanism of action, challenges, and therapeutic approaches. Eur J Pharmacol. 2021;905:174178.
361. Tan CP, Sinigaglia L, Gomez V, Nicholls J, Habib NA. RNA Activation-A Novel Approach to Therapeutically Upregulate Gene Transcription. Molecules. 2021;26(21).
362. Goh SY, Chao YX, Dheen ST, Tan EK, Tay SS. Role of MicroRNAs in Parkinson's Disease. Int J Mol Sci. 2019;20(22).
363. Chapman SD, Farina L, Kronforst K, Dizon M. MicroRNA Profile Differences in Neonates at Risk for Cerebral Palsy. Phys Med Rehabil Int. 2018;5(3).
364. Kotowska-Zimmer A, Pewinska M, Olejniczak M. Artificial miRNAs as therapeutic tools: Challenges and opportunities. Wiley Interdiscip Rev RNA. 2021;12(4):e1640.
365. Rinaldi C, Wood MJA. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. Nat Rev Neurol. 2018;14(1):9-21.
366. Merola A, Kobayashi N, Romagnolo A, Wright BA, Artusi CA, Imbalzano G, et al. Gene Therapy in Movement Disorders: A Systematic Review of Ongoing and Completed Clinical Trials. Front Neurol. 2021;12:648532.
367. Brunet de Courssou JB, Durr A, Adams D, Corvol JC, Mariani LL. Antisense therapies in neurological diseases. Brain. 2022;145(3):816-31.
368. Tabrizi SJ, Estevez-Fraga C, van Roon-Mom WMC, Flower MD, Scahill RI, Wild EJ, et al. Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. Lancet Neurol. 2022;21(7):645-58.
369. Lin K, Zhong X, Li L, Ying M, Yang T, Zhang Z, et al. AAV9-Retro mediates efficient transduction with axon terminal absorption and blood-brain barrier transportation. Mol Brain. 2020;13(1):138.
370. Zhang X, Chai Z, Lee Dobbins A, Itano MS, Askew C, Miao Z, et al. Customized blood-brain barrier shuttle peptide to increase AAV9 vector crossing the BBB and augment transduction in the brain. Biomaterials. 2022;281:121340.
371. Hordeaux J, Jeffrey BA, Jian J, Choudhury GR, Michalson K, Mitchell TW, et al. Efficacy and Safety of a Krabbe Disease Gene Therapy. Hum Gene Ther. 2022;33(9-10):499-517.

372. Crooke ST. Addressing the Needs of Patients with Ultra-Rare Mutations One Patient at a Time: The n-Lorem Approach. Nucleic Acid Ther. 2022;32(2):95-100.
373. Garcia VOS, Ivy C, Fu J. Syntopic frogs reveal different patterns of interaction with the landscape: A comparative landscape genetic study of Pelophylax nigromaculatus and Fejervarya limnocharis from central China. Ecol Evol. 2017;7(22):9294-306.
374. Calderon LD, Silva ADE, Ciancaglini P, Stabeli RG. Antimicrobial peptides from Phyllomedusa frogs: from biomolecular diversity to potential nanotechnologic medical applications. Amino Acids. 2011;40(1):29-49.
375. Kriegman S, Blackiston D, Levin M, Bongard J. Kinematic self-replication in reconfigurable organisms. P Natl Acad Sci USA. 2021;118(49).
376. Hammond SA, Warren RL, Vandervalk BP, Kucuk E, Khan H, Gibb EA, et al. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nature Communications. 2017;8.

Appendix 1: Proforma used for collection of phenotypic data from overseas referrers

STUDY OF GENETIC MOVEMENT DISORDERS: CLINICAL INFORMATION

- *Please ensure the patient or parents have given consent to share clinical information.*
- *The more detail you can provide, the more accurate our genetic analysis will be.*

Patient's name:

Date of birth:

Sex:

1. Family History

a) Are the parents related, and/or from the same village?

b) Does any relative have any relevant health condition (such as a movement disorder, epilepsy, learning disability…)

2. Medical Background

a) Please describe any complications during pregnancy or at birth.

b) Has the patient ever had a major illness (such as meningitis or sepsis)

c) Has the patient ever had epileptic seizures (including febrile convulsions)?

d) Does the patient have any dysmorphic feature or an unusual appearance?

e) Does the patient suffer from any other medical conditions?

f) What medication does the patient take?

3. Development

a) Does the patient have developmental delay and/or intellectual disability? If so, how severe?

b) Did the patient ever have developmental regression? If so, at what age?

c) How old was the patient when they first:

Sat unsupported

Walked

Started talking

4. Movement disorder

a) What were the first symptoms, and what age did they start?

b) What is the nature of the movement disorder (e.g. dystonia, chorea, spasticity…)?

c) Which parts of the body are affected?

d) Do symptoms fluctuate? If so, does anything in particular make them better or worse?

e) Are the symptoms getting worse over time?

f) How severe are the symptoms? For example, do they affect the patient's ability to walk, to get dressed, to swallow?

g) Has any treatment been tried? Did it help?

5. Examination Findings

Please describe the neurological examination of this patient.

6. Test Results

What was the result (if known) of:

a) MRI scan or other imaging

b) Biochemical testing

c) Neurotransmitter testing

d) Genetic tests

f) Nerve conduction studies

g) Any other test results

Thank you very much

Appendix 2: Template report

LONDON'S GLOBAL UNIVERSITY

UCL

Patient's Name: **Date of Report:**
Patient's Date of Birth:
Brief Phenotype:

Report to: XXXXXX
Address: XXXXX
CC to: XXXXX

RESEARCH WHOLE-GENOME SEQUENCING REPORT

(This analysis was performed on a research basis and not in an accredited diagnostic laboratory)

Following next-generation sequencing of the whole genome, analysis of biallelic (that is, compound heterozygous or homozygous) variants and *de novo* heterozygous variants was undertaken on a gene-agnostic basis, i.e. without reference to specific gene panels. In addition, analysis of inherited heterozygous variants was undertaken using our Version 3 Megapanel (full details available on request): only variants considered relevant to the patient phenotype are reported below.

RESULT: No clearly pathogenic variant was identified OR
Variant(s) were identified as follows (gene, c., p., genotype) which could be consistent with a diagnosis of.....

The following variants of uncertain significance (ACMG Class 3) were identified:

Gene and variant	Notes

This table does not include variants believed unlikely to be contributory to the patient's phenotype, or variants in genes where no clear disease phenotype has yet been established.

(Include details of parental testing if relevant.)

(If negative) We cannot exclude the chance that a significant genetic variant exists but we have been unable to identify it. Even with a negative WGS report, a single-gene disorder remains a possibility. Please note that our analysis does not exclude copy-number variants, structural variants, and other variants for which standard *in silico* prediction techniques are ineffective, such as deeply-intronic splicing variants.

[Due to their uncertain significance], these variants [have/have not] been checked by Sanger sequencing.

(If positive) Please note that we are **not** an accredited diagnostic laboratory. If you consider this result clinically relevant, then we strongly recommend that you request confirmation from your local genetic diagnostic service. We would be grateful if you could discuss this result with the patient and/or family and arrange genetic counselling as appropriate. We are happy to be contacted with any queries.

Further analysis may be undertaken on a research basis. A further report will be issued **only** if a clinically significant positive result is identified.

Kurian Neurogenetic Research Group, Molecular Neurosciences Section, Developmental Neurosciences Programme, Zayed Centre for Research into Rare Disease in Children, 20c Guilford Street, London, WC1N 1DZ

Dr Dora Steel
Clinical Research Fellow

Professor Manju Kurian
NIHR Research Professor and
UCL Professor of Neurogenetics

Appendix 3: Variant shortlists for cases discussed in Chapters 3 and 4

DY131: WGS; triome; panel-free analysis for biallelic and Megapanel for heterozygous variants

Gene (inheritance; MAF)	Variant	CADD	SIFT	PolyPhen-2	Mutation Taster	Align GVGD	Other scores	Notes
RHOBTB2 **De novo** **Novel**	c.722C>A p.Ser241Tyr	26	Deleterious (0)	HD: Prob Dam (1.000) HV: Prob Dam (0.99)	Disease causing (1)	C15	.	*DEE64*
SPTBN5 **Homozygous** MAF 0.08% Hom 0.0024%	c.8079+5G>A p.?	<10	.	.	.	.	MaxEnt: -100.0% NNSPLICE: -52.8% SSF: -17.2%	*Widely expressed cytoskeletal protein*
MGA **Homozygous** MAF 0.18% Hom 0.0043%	c.5450T>C p.Leu1817Pro	24.7	Deleterious (0.01)	HD: Prob Dam (0.95) HV: Poss Dam (0.642)	Disease causing (1)	C0	.	*Transcription factor, expressed brain and elsewhere*

PE5: WGS; singleton analysis; panel-free analysis for biallelic and Megapanel for heterozygous variants

Gene (inheritance; MAF)	Variant	CADD	SIFT	Mutation Taster (v2013)	PolyPhen-2	Align GVGD	Other scores	Notes
JPH3 **Hom** Novel	c.1310del; p.Arg437Leufs*34	.	.	Disease causing (1)	.	.	Provean: Deleterious (-238.702)	*Junctional complex (plasma membrane-ER). Brain specific. AD CAG expansion: Huntington-like disease. See also PE19 (heterozygous)*
NDUFAF5 **Hom** MAF 0.0037%	c.524A>G; p.His175Arg	21.8	Tolerated (0.17)	Disease causing (1)	HD: Poss Dam (0.804) HV: Benign (0.425)	C0	.	*A complex 1 assembly factor: AR mitochondrial complex 1 deficiency.*
SRA1 **Hom** MAF 0.00064%	c.22C>T; plGln8*	33	.	Disease causing (1)	.	.	Provean: Deleterious (-443.328)	*Steroid receptor RNA activator. Widely expressed but mostly liver and muscle.*
SOX6 **Hom** Novel	c.721G>A; p.Aps241Asn	25.4	Deleterious (0.02)	Disease causing (1)	HD: Prob Dam (0.998) HV: Prob Dam (0.991)	C0	.	*Tolchin-Le Caignec syndrome (AD): ID and bony anomalies. Transcription factor, widely expressed. NB AD pathogenic variants are known only in HMG domain, or LoF)*
COPRS **Hom** MAF 0.0021%	c.482A>T; p.His161Leu	25.8	Deleterious (0)	Disease causing (0.944)	HD: Prob Dam (0.999) HV: Prob Dam (0.997)	C65	.	*Histone-binding protein, widely expressed.*
RALGAPA2 **Hom** MAF 0.0068%	c.5270G>A	25.9	Deleterious (0.02)	Disease causing (1)	HD: Prob Dam (1.000) HV: Prob Dam (1.000)	C0	.	*GTPase activator; widely expressed.*

PE19: WGS; analysis of proband and affected mother; panel-free analysis of shared variants and variants biallelic in proband

Gene (Maternal only; MAF)	Variant	CADD	SIFT	Mutation Taster (v.2013)	PolyPhen-2	Align GVGD	Provean	Splicing predictors	Notes
JPH3 Novel	c.1014C>G; p.Tyr338*	56	.	Disease causing (1)	.	.	Provean: Deleterious (-506)	.	*Huntington's-like disease – but from a triplet expansion. Haploinsufficient mice have a motor phenotype though.* ***pLI 1, O/E 0.04***
PPP2R5A MAF 0.00037%	c.1098+3_1100del; p.Val367Cysfs*10	35	.	Disease causing (1)	.	.	Provean: Deleterious (-344.686)	MaxEnt: -100.0% NNSPLICE: -100.0% SSF: -100.0%	*Modulates activity of a protein phosphatase. Ubiquitous.* ***pLI 0.82, O/E 0.18***
KCNJ18 Known dbSNP	c.631C>T; p.Leu211Phe	22.8	Tolerated (0.27)	.	HD: Prob Dam (0.959) HV: Poss Dam (0.747)	C0	.	.	*Susceptibility to thyrotoxic period paralysis (AD) – this gene is a frequent flier*
CTNND2 MAF 0.0012%	c.1361G>A; p.Arg454His	25.9	Tolerated (0.13)	Disease causing (1)	HD: Prob Dam (0.987) HV: Poss Dam (0.474)	C0	.	.	*Involved in forming and maintaining dendritic spines. Highly expressed in foetal brain.* ***pLI 1, MC 0.7***

CD24: WGS; analysis of proband alone (subsequent targeted segregation testing in affected sibling); Megapanel analysis of biallelic and heterozygous variants

Gene (MAF; zygosity)	Variant	CADD	SIFT	PolyPhen-2	Mutation Taster (v2013)	Align GVGD	Other scores	Notes
SLC30A9 Novel **Homozygous**	c.1253G>T; p.Gly418Val	34	Deleterious (0)	HD: Prob Dam (0.99) HV: Prob Dam (0.95)	Disease causing (1)	C15	MaxEnt: -28% NNSPLICE: -52% SSF: -7.7%	*Birk-Landau-Perez syndrome* ***CD24S also homozygous***
IARS2 Novel **Homozygous**	c.329G>A; p.Cys110Tyr	26.6	Tolerated (0.45)	HD: Prob Dam (0.99) HV: Prob Dam (0.93)	Disease causing (1)	C0	.	*CAGSS, Leigh and West syndromes* ***CD24S was heterozygous for variant***
UNC13B MAF 0.002% **Comp het**	c.2537G>A; p.Arg846His	33	Deleterious (0)	HD: Prob Dam (1.00) HV: Prob Dam (0.94)	Disease causing (1)	C25	.	*Involved in vesicle maturation and NT release at glutamatergic synapses. No known disease association. 2nd variant tolerated in homozygosity.*
UNC13B MAF 0.86% Hom 0.011% **Comp het**	c.3288C>G; p.Phe1096Leu	25.4	Deleterious (0)	HD: Prob Dam (0.96) HV: Poss Dam (0.85)	Disease causing (1)	C15	.	
TREX1 MAF 0.0035% Heterozygous	c.350G>A p.Arg117His	26.4	Deleterious (0)	HD: Prob Dam (1.000) HV: Prob Dam (0.998)	.	C25	.	*Aicardi-Goutieres (including AD cases). Note frequency; poor fit.*

DY197: WGS; triome; panel-free analysis of biallelic and *de novo* variants (reanalysis of VCF files from Congenica)

Gene (MAF; inheritance)	Variant	CADD	SIFT	Mutation Taster (v2013)	PolyPhen-2	Align GVGD	Other scores	Notes
H3-3B Novel **De novo**	c.365C>G; p.Pro122Arg	32	Deleterious	Disease causing (1)	HD: Prob Dam (0.997) HV: Prob Dam (0.915)	C65	MC (o/e): 0.1 pLI 0.81	*Histone variant. Ubiquitous. Somatic mutations cause cancer. Neurodevelopmental disorder.*
UBC Novel **De novo**	c.2038C>T; p.Arg680Cys	27.7	Deleterious (0)	Disease causing (1)	HD: Prob Dam (1.000) HV: Prob Dam (0.995)	C0	MC (o/e): 0.46 pLI 0.14	*Ubiquitin C. Ubiquitous. NB haploinsufficiency in mice is tolerated. GM: series with ID only*
ERBIN MAF 0.002% **Paternal**	c.3989G>A; p.Arg1330Gln	24.3	Tolerated	Disease causing (0.9997)	HD: Prob Dam (1.000) HV: Prob Dam (0.932)	.	.	*Adaptor for ERBB2 receptor. Ubiquitous but highly expressed in brain. Involved in multiple signal transductions. Possibly implicated in anxiety modulation in amygdala (in mice). GM: nothing so far.*
ERBIN Novel **Maternal**	c.1897A>G; p.Lys633Glu	28.1	Deleterious	Disease causing (1)	HD: Prob Dam (0.998) HV: Prob Dam (0.993)	.	.	

DY47: WGS; singleton; Megapanel analysis of biallelic and *de novo* variants

Gene (MAF)	Variant	CADD	SIFT	PolyPhen-2	Mutation Taster (v2013)	Align GVGD	Other scores	Notes
ABCD1 gnomAD 0%	c.1922G>T; p.Gly641Val	27.4	Deleterious (0.02)	HD: Prob Dam (0.998) HV: Prob Dam (0.973)	Disease causing (1)	C0	NA	*X-ADL* *Low read depth: Sanger found false positive*
CHRNB2 MAF 0.0004%	c.534C>A; p.Tyr178*	35	NA	NA	Disease causing (1)	NA	Provean: Deleterious (-855.028)	*AD nocturnal frontal lobe epilepsy – not a phenotypic fit*
H3F3A Novel	c.73G>A; p.Ala25Thr	24.7	Deleterious (0)	HD: Benign (0.152) HV: Benign (0.025)	Disease causing (1)	C55	MC 0.03	*Variant histone: ubiquitous; involved in glioma*
RAB35 Novel	c.313G>A; p.Glu105Lys	26.3	Deleterious (0)	HD: Prob Dam (0.978) HV: Prob Dam (0.915)	Disease causing (1)	C55	MC 0.41	*Regulates vesicular trafficking, expressed widely including brain* *Listed on GeneMatcher: nil relevant*

CD39: WES; singleton; panel-free analysis of biallelic and Megapanel analysis of heterozygous variants

Gene (MAF)	Variant	CADD	SIFT	PolyPhen-2	Mutation Taster (v2013)	Align GVGD	Other scores	Notes
YY1 Novel	c.907T>C; p.Cys303Arg	31	Deleterious (0.01)	HD: Prob Dam (1.00) HV: Prob Dam (0.99)	Disease causing (1)	C0	(Trivial splicing effects)	*Gabriele-De Vries syndrome (AD)*
PTPRZ1 Known dbSNP	c.5252A>G; p.Lys1751Arg	26.2	Deleterious (0)	HD: Poss Dam (0.764) HV: Poss Dam (0.773)	Disease causing (1)	C25	LoFC 0.19 pLI 0.99	*Regulates oligodendrocyte development*
PCDHGB MAF 0.0032%	c.2011dup; p.Leu671 Profs*6	28.3	.	.	Disease Causing (1)		LoFC 0.27, pLI 0.01 Deleterious (-437.909)	*Cell adhesion molecule, possibly involved in establishing axonal connections.*
MYH14 MAF 0.0012% (ClinVar VUS)	c.4849G>A p.Glu1617Lys	29.5	Deleterious (0)	HD: Prob Dam (1.000) HV: Prob Dam (0.999)	Disease causing (1)	C55	.	*Neuropathy, myopathy, hoarseness and hearing loss (AD)*

TR5: WGS; singleton; Megapanel analysis of biallelic and heterozygous variants

Gene (MAF)	Variant	CADD	SIFT	PolyPhen-2	Mutation Taster (v2013)	Align GVGD	Other scores	Notes
STXBP1 Novel	c.1783T>C; p.*595Glnext*67	22.1	.	.	.	.	Stop loss: 67 extra residues	*AD EIEE gene with broader ND spectrum*
KCNJ4 Novel	c.809C>T; p.Pro270Leu	29.3	Deleterious (0)	HD Prob Dam (1.000) HV Prob Dam (1.000)	Disease-causing (1)	C65	MC 0.39 (LoFC 0; pLI 0.96)	*Potassium channel in heart and brain; no known phenotype*